On the interplay between data and decisions in discrete location problems

Zur Erlangung des akademischen Grades eines
Doktors der Ingenieurwissenschaften
(Dr.-Ing.)
von der KIT-Fakultät für Wirtschaftswissenschaften
des Karlsruher Instituts für Technologie (KIT)
genehmigte

Dissertation

von

Hannah Bakker

Tag der mündlichen Prüfung:	15.03.2024
Referent:	Prof. Dr. Stefan Nickel
Korreferent:	Prof. Dr. M. Grazia Speranza

Acknowledgements

I am deeply grateful to all those who have supported and guided me throughout my journey to completing this thesis. It would not have been possible without the continuous support of many individuals. In particular, I would like to thank:

Prof. Dr. Stefan Nickel, for his continuous support throughout supervising this thesis, for sparking my interest in OR many years ago, and for giving me the opportunity to freely pursue my research ideas.

Prof. Dr. M. Grazia Speranza, for refereeing this work and being open to and interested in my ideas.

Prof. Dr. Oliver Stein, for his insightful questions during my examination that allowed me to evaluate my own work from yet a new perspective.

Prof. Dr. Clemens Puppe, for his friendly conduct as the chairman of the examination committee.

My colleagues, for ensuring that I always enjoyed coming to work. Particular thanks go to Fabian Dunke for providing me with valuable feedback on the present manuscript, never tiring of discussing new ideas, and encouraging me to start my PhD journey years ago.

My friends, for being patient when I needed time to focus and being present to celebrate small and large successes.

Barbara, Ernst, and Brigitte, for teaching me curiosity and the value of learning from an early age, and their continuous support in everything I do.

Jonas, for his moral support, eye for detail, and always cheering me up.

Finally, I thank Christoph for his continuous support, patience, and encouragement.

Abstract

The capacitated facility location problem (CFLP) is a core problem in location science. Several exact and heuristic solution approaches and various modeling extensions integrating time, uncertainty, or further supply chain network design decisions exist. However, the average-case-based analysis underlying the majority of works in this field leaves several questions unaddressed: Why does the runtime of state-of-the-art MIP solvers differ drastically between seemingly similar problem instances? Why does an explicit consideration of the temporal development of the parameters in a multi-period model lead to a significant improvement of the objective value in some instances and is negligible in others? When does moving to a more complex model result in different decisions and when are the same decisions merely evaluated differently? None of these questions are easy to address, yet one thing is clear: when the same model and algorithm are used, the reason for the observable differences must lie in the problem data. The present work offers new insights into the relationship between data and the optimal decisions of CFLP instances.

A new methodological approach that characterizes instances based on their decisions in well-performing, that is optimal and near-optimal solutions is presented. It is shown that the combination of spatial patterns, the capacity-to-demand, and fixed-costs-to-profit ratio result in the facilities operating in the optimal solution to individual instances interdepending to varying degrees. A newly proposed measure based on the decisions in well-performing solutions of an instance allows the quantification of the level of this interdependence. It gives rise to a classification of CFLP instances as either primarily composed of independent or primarily composed of interdependent facilities. This distinction allows the anticipation and explanation of the performance of different exact and heuristic solution algorithms.

It is demonstrated that facilities serving the same subset of customers interdepend more strongly, which induces an implied separation of the sets of candidates and customers into distinct regions. A pattern-based approach to retrieve these regions already from integer-infeasible solutions is introduced. It is shown that knowledge of the interdependence relationships and, more precisely, the implied regions opens up new opportunities for algorithm development. Computational experiments show that integrating it into the branch-variable selection step of CPLEX's branch-and-cut solver leads to a significant runtime reduction.

Moving to a more complex model increases the flexibility of a decision-maker but also comes at the cost of added complexity. Thus, it is important to understand the conditions

under which a model extension results in an added value. Model extensions of the CFLP in the two most commonly found directions for location problems are analyzed: time and uncertainty. For the first, a multi-period CFLP with phase-in constraints is examined; for the latter, a two-stage stochastic program and an adjustable robust program of the CFLP are compared. A central question regards the quantification of the added value. For the multi-period CFLP, the value of the multi-period solution is used; for the uncertainty-aware formulations, novel measures to assess and compare the added value are presented. For the multi-period model, it is shown that not the degree to which parameters change over the planning horizon but the degree to which the model extensions allow the reduction of relevant cost components of the objective value is decisive. Consequently, the mere fact that decisions are taken in a temporal context does not necessarily make time an important aspect to include in the model. For uncertainty-aware models, the characterization of instances allows the anticipation of the level of flexibility with which location decisions can be adapted to hedge against uncertainty. When the implied patterns are rigid in the sense that few candidates clearly dominate, the added value of moving to a robust or stochastic approach diminishes as the more complex model proposes the same location decisions. Both results encourage further research on data-driven modeling.

This work presents novel methods to analyze solutions in the decision space. Insights are validated by experiments on several sets of benchmark instances from literature. Deriving and subsequently analyzing multi-sets of optimal and near-optimal solutions to these instances is possible only with the help of state-of-the-art potent MIP solvers and efficient data analytics tools. As such, this thesis demonstrates that recent advances in computing power not only allow addressing larger and more complex problems but also open up new perspectives on problems that have been studied for decades – like the CFLP.

Contents

List of Figures

List of Tables

Glossary

average dependence density measure for the level of interdependence between facilities operating in an optimal solution (see Def. 2.13)

breakpoint a point in time during the planning horizon at which decisions can be taken (see Def. 6.1)

core facilities subset of facilities that operate in all or most optimal and near-optimal solutions to a particular problem instance

decision pattern recurring structure that a set of solutions to a particular problem instance share

dependend a facility depends on another facility if the removal of the latter changes the opening status of the former from open to closed in the optimal solution (see Def. 2.9)

discretization ordered set of breakpoints that induce a separation into individual periods of a planning horizon (see Def. 6.2)

facility-customer space set of facilities and customers arranged in a hypothetical plane arranged according to their relative positions implied by the unit transport cost matrix

implicit warehousing warehousing occurring throughout a single period in a multi-period model and that is evaluated differently depending on the discretization of the planning horizon

independent a facility is independent if it does not depend on any other facility (see Def. 2.11)

interdepend two facilities interdepend on one another if one depends on the other and vice versa (see Def. 2.10)

kernel persistence measure for the relative importance of a subset of facilities in a set of solutions (see Def. 2.4)

overlap coefficient measure for the similarity between two finite sets (see Def. 2.7)

perturbation level maximum absolute deviation of any parameter from its nominal or expected value

search strategy strategy to determine which node to explore next in a Branch-and-Bound procedure

service regions subsets of customers served jointly from subsets of interdependent facilities

spatial patterns estimated spatial distribution of facilities and customers in a hypothetical two-dimensional plane (random, clustered, or evenly distributed)

variable selection strategy strategy to determine which variable to branch on given a fractional solution in a node in a Branch-and-Bound procedure

worst-case gap relative arithmetic difference between the objective value of a robust counterpart and the actual worst-case realization of the objective value (see Def. 7.2)

Acronyms

2SS-CFLP Two-stage Stochastic CFLP

2SS-CFLP-DE Deterministic Equivalent Two-stage Stochastic CFLP

ARC-CFLP Adjustable Robust Counterpart to the CFLP

AVUS Average Value of an Uncertainty-aware Solution

B&B Branch-and-Bound

B&C Branch-and-Cut

CFLP Capacitated Facility Location Problem

EVP Expected Value Problem

EVPI Expected Value of Perfect Information

HC Heuristic Concentration

HCI Heuristic Concentration Integer

KS Kernel Search

LP Linear Programming

LR Lagrangean Relaxation

MDS Multidimensional Scaling

MP-CFLP Multi-period Capacitated Facility Location Problem (with phase-in constraints)

OR Operations Research

SS-CFLP Single-Source CFLP

SVD Singular Value Decomposition

VMPS Value of the Multi-period Solution

VSS Value of the Stochastic Solution

WCVUS Worst-Case Value of an Uncertainty-aware Solution

1. Introduction

1.1. Motivation

The capacitated facility location problem (Capacitated Facility Location Problem (CFLP)) is a core problem in location science [Fernández and Landete, 2019]. It considers the problem of serving a set of customers with demands from a set of capacitated facilities with the objective of minimizing costs. Operating a facility incurs a fixed cost, and transporting goods to customers incurs variable costs depending on the transport volume. The model decides which facilities to operate and how to allocate customers.

The CFLP has various applications and plays a critical role in many areas [Drezner and Hamacher, 2004] including supply chain management [Melo et al., 2006], humanitarian relief logistics [Balcik and Beamon, 2008], or emergency systems [Jia et al., 2007]. While the standard formulation finds the cost-minimal policy that serves all customer demand, several recent works deploy a net profit-maximizing formulation that allows for partial demand satisfaction (see, e.g., Ardestani-Jaafari and Delage [2018]).

Let I be the set of candidate facilities and J the set of customers. The binary decision variable y_i represents the location decisions and is equal to 1 if a facility is opened at the corresponding candidate location i and 0 otherwise. The non-negative allocation decisions x_{ij} denote the fraction of the demand of customer j served from a facility at candidate location i. The objective (1.1) maximizes the profit. The allocation of customers to facilities incurs variable transportation costs c_{ij} and generates a variable profit r_j per unit of the customer demand D_j. The establishment of a facility has a fixed cost F_i. Customers cannot be served more than their original demand (1.2), and the facility capacities Q_i must not be exceeded (1.3).

CFLP

$$\max -\sum\nolimits_i F_i y_i + \sum\nolimits_i \sum\nolimits_i (r_j - c_{ij})\, D_j x_{ij} \tag{1.1}$$

$$\text{s.t.} \sum\nolimits_i x_{ij} \leq 1 \qquad \forall j \in J \tag{1.2}$$

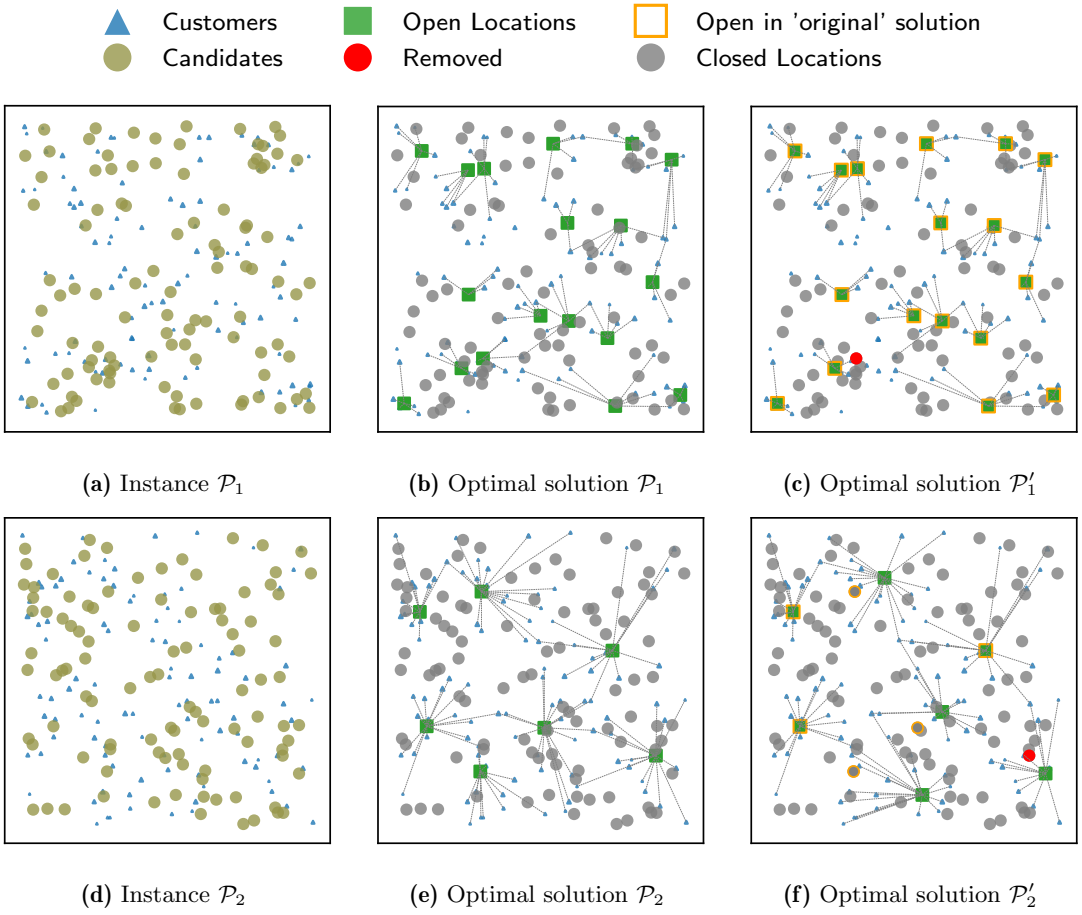
$$\sum\nolimits_j D_j x_{ij} \leq Q_i y_i \qquad \forall i \in I \tag{1.3}$$

$$y_i \in \{0,1\} \qquad \forall i \in I \tag{1.4}$$

$$x_{ij} \geq 0 \qquad \forall i \in I,\ j \in J. \tag{1.5}$$

Example A Figure 1.1a and Figure 1.1d depict two problem instances, $\mathcal{P}_1$ and $\mathcal{P}_2$, of the CFLP obtained with the randomized data generation procedure introduced by Klose and Drexl [2005]. Both instances contain 100 customers and candidate locations, respectively. They primarily differ in the capacity-demand ratio – an input parameter of the generation procedure that determines the ratio between the sum of all capacity and the sum of all demand in the network. $\mathcal{P}_1$ has a ratio of 3, meaning that the sum of all capacities is

three times the sum of all demands. $\mathcal{P}_2$ has a ratio of 10. Consequently, the capacities of individual facilities in $\mathcal{P}_2$ are – on average – significantly larger than those in $\mathcal{P}_1$.



(a) Instance $\mathcal{P}_1$ **(b)** Optimal solution $\mathcal{P}_1$ **(c)** Optimal solution $\mathcal{P}_1'$

(d) Instance $\mathcal{P}_2$ **(e)** Optimal solution $\mathcal{P}_2$ **(f)** Optimal solution $\mathcal{P}_2'$

Figure 1.1.: Visualization of instances and optimal solutions ($\mathcal{P}_1$-$\mathcal{P}_2$, Ex. A)

Figure 1.1b and Figure 1.1e depict the (unique) optimal solutions to the CFLP for both instances. We can see that for $\mathcal{P}_1$, 18 facilities are open in the optimal solution, while for $\mathcal{P}_2$, only seven facilities operate. The optimal objective value for $\mathcal{P}_1$ is 2373.1 while for $\mathcal{P}_2$ it is 19803.9.

Suppose that one of the candidate facilities operating in the optimal solution of each instance is removed from the set of candidates. The removed facilities are depicted in red in Figure 1.1c and Figure 1.1f, respectively. The two figures, furthermore, depict the optimal solution to the problem obtained when resolving the problem with the reduced set of candidates, instances $\mathcal{P}_1'$ and $\mathcal{P}_2'$. The optimal objective values are 2107.3 and 19687.9, respectively.

While for $\mathcal{P}_1$, the removal of one particular candidate facility resulted in a reduction of the optimal objective value of 11.2%, for $\mathcal{P}_2$, the removal of one particular candidate only resulted in the reduction of the optimal objective value of 0.6%. This is remarkable, given that in both instances we removed the candidate facility that incurred the most significant reduction of the objective value. Furthermore, while for $\mathcal{P}_1$, the removal of the candidate did not affect the optimality status of any other facility, the index set of the operating facilities in $\mathcal{P}_2$ changed significantly with the removal of that candidate. Three facilities

operating in the optimal solution of the original problem are no longer operating, while four facilities closed in that solution are newly operating.

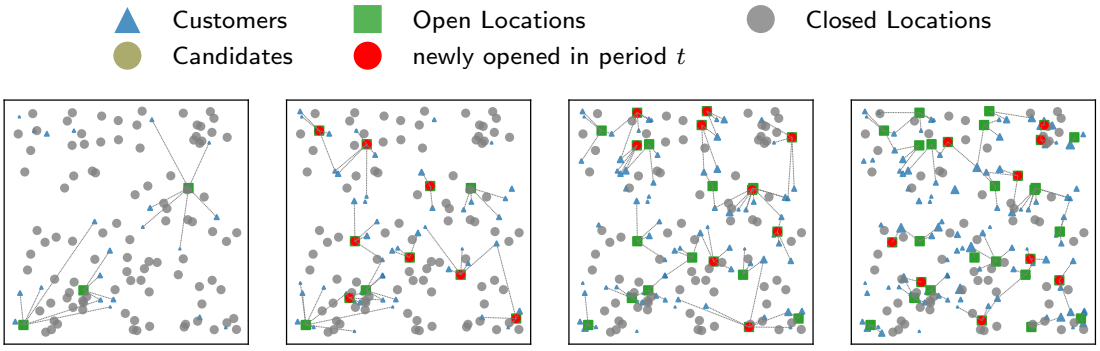
When solving both instances with CPLEX Version 12.10 on an Intel(R) Core(TM) i7-7700 processor with 3.60 GHz and 64 GB RAM, the solve time of $\mathcal{P}_1$ was 0.2 seconds while for $\mathcal{P}_2$ it was 7.0 seconds. Both instances are of the same size, with 100 candidates and customers, respectively. This difference increases when the experiment is repeated with similar instances produced with the same generator with 200 candidate facilities and customers and capacity-demand ratios of 3 and 10, respectively. These instances will be referred to as $\mathcal{P}_3$ and $\mathcal{P}_4$. The instance with a ratio of 3 took 1.6 seconds to solve, while the instance with a ratio of 10 took 55.0 seconds. ▲

Example A illustrates significant structural differences between optimal solutions of two facility location instances, which, at first sight, seem similar. While individual facilities are of extreme relevance for achieving the optimal objective value in $\mathcal{P}_1$, contributions of individual facilities are much less significant in the solution to $\mathcal{P}_2$. Furthermore, while in the optimal solution to $\mathcal{P}_1$, the optimality status of individual candidates is entirely unaffected by the removal of another facility, in the optimal solution to $\mathcal{P}_2$, a subset of the candidates operating in the optimal solution seems to be optimal only in combination.

Being aware of these structural differences can be of utmost relevance for a decision-maker. A recent literature review on location data analytics by Ferro-Diez et al. [2020] states that all reviewed facility location use cases concern long-term decisions due to the high costs and expectations associated with establishing new facilities. Thus, being aware of the importance of individual facilities for the overall performance of a solution and relationships between the individual decisions a solution is composed of is of great interest. Additionally, Example A suggests that these structural differences are related to the complexity of the solution process, making a more thorough understanding of the solution attractive from a theoretical perspective.

Location decisions oftentimes involve high investments and are taken at a strategic level. Multi-period models that explicitly consider time have received considerable attention [Melo et al., 2006, Cortinhal et al., 2015]. A set of periods induces a discretization of the planning horizon, which imposes a prespecified number of moments for implementing decisions [Nickel and Saldanha da Gama, 2019]. Individual periods must be linked to one another, or else the problem will fall into separate subproblems for each period. For situations in which suppliers face increasing demands, a common linking constraint is a phase-in constraint. It ensures that once a facility has been established, it operates until the end of the planning horizon.

Example B Assume that $\mathcal{P}_1$ and $\mathcal{P}_2$ actually aggregate parameters over a planning horizon comprising four periods, and $\mathcal{P}_1^4$ and $\mathcal{P}_2^4$ denote the corresponding multi-period instances. Fixed costs and capacities are time-invariant such that $q_{it} = Q_i/4$ and $f_{it} = F_i/4$ for all facilities $i \in I$ and periods $t \in T$. Unit costs and profits are time-invariant as well. Each customer's demand increases constantly in each period by a total of $\delta = 100\%$ over the planning horizon. Thus, $\sum_t d_{jt} = D_j$ and $d_{j4} = (1 + \delta)d_{j1}$.



(a) $\mathcal{P}_1^4$, $t = 1$ **(b)** $\mathcal{P}_1^4$, $t = 2$ **(c)** $\mathcal{P}_1^4$, $t = 3$ **(d)** $\mathcal{P}_1^4$, $t = 4$

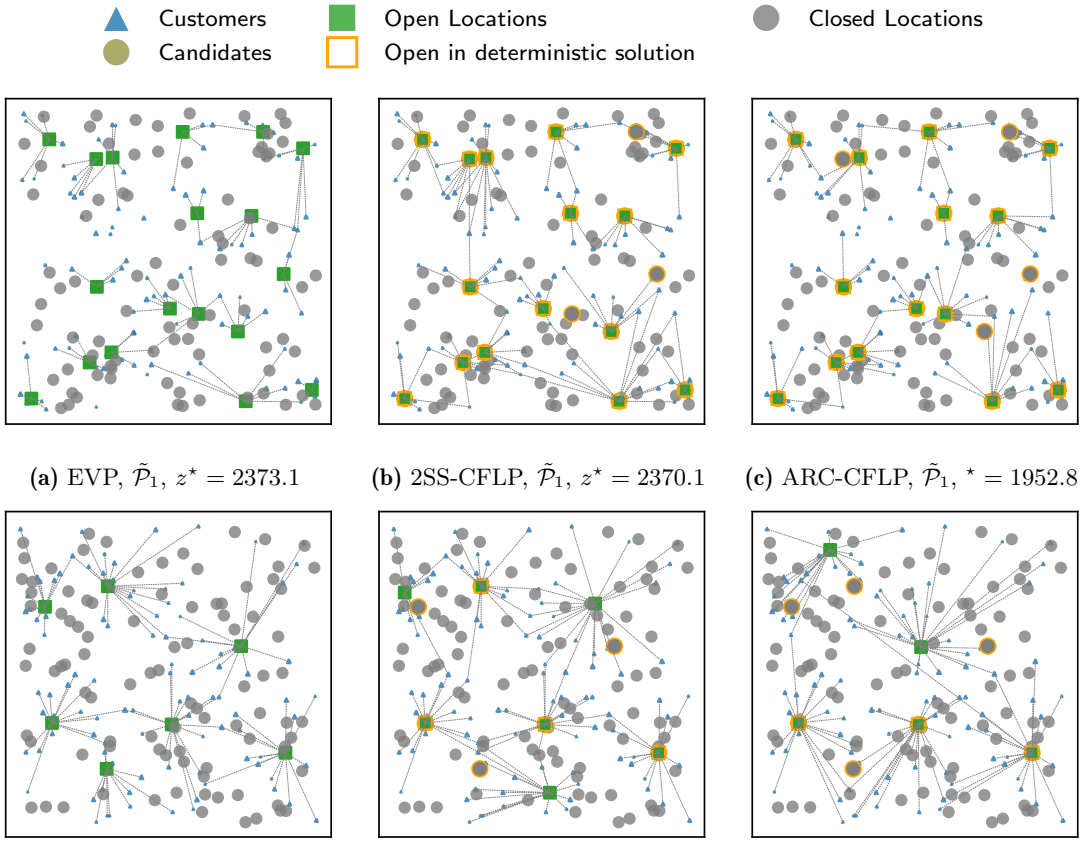
(e) $\mathcal{P}_2^4$, $t = 1$ **(f)** $\mathcal{P}_2^4$, $t = 2$ **(g)** $\mathcal{P}_2^4$, $t = 3$ **(h)** $\mathcal{P}_2^4$, $t = 4$

Figure 1.2.: Visualization of optimal multi-period solutions ($\mathcal{P}_1^4$-$\mathcal{P}_2^4$, Ex. B)

Figure 1.2 visualizes the optimal solutions to $\mathcal{P}_1^4$ and $\mathcal{P}_2^4$. It illustrates the successive opening of facilities over the course of the planning horizon. Recall that over the course of the planning horizon, the demand of customers is exactly the same as the demand in the static problem instances from Example A. In $\mathcal{P}_1^4$, the set of facilities operating in the final period is a superset of those facilities operating in the optimal solution to the static model $\mathcal{P}_1$. This is intuitive, as the demand capacity ratio is higher in those final periods than on average across all periods. However, in $\mathcal{P}_2^4$, the set of facilities differs significantly from that operating in the optimal static solution. Furthermore, there is a noticeable difference between the two solutions regarding the development of customer allocations over time. In $\mathcal{P}_1^4$, facilities that are newly opened in each period serve customers that have not been served before. They open up new customer regions. In contrast, in $\mathcal{P}_2^4$, customers are successively redistributed to the newly operating facility in each period. As in both instances the demand of each customer increases at a constant rate, this discrepancy is remarkable. In the optimal solution to $\mathcal{P}_1^4$, new facilities are opened in a new region as soon as the demand suffices to make them profitable. Meanwhile, in $\mathcal{P}_2^4$, the entire customer demand is covered from the beginning, and with successively increasing demands, facilities are opened to close in on customers. ▲

Together with the longevity of location decisions comes the problem that by the time decisions have to be taken, relevant problem parameters are, at least to some degree, uncertain. A common source of uncertainty is customer demand, which usually needs to be estimated and is consequently bound to estimation errors. Several ways to include uncertainty in the model formulation exist. Two of the most prominent modeling paradigms are stochastic programming and robust optimization.

Example C Assume that the customer demand in instances $\mathcal{P}_1$ and $\mathcal{P}_2$ is subject to uncertainty, yielding two uncertainty-affected problem instances $\tilde{\mathcal{P}}_1$ and $\tilde{\mathcal{P}}_2$. In each of the two instances, the demand of every customer deviates by at most 20% from the estimated value. Further, assume that each customer follows a symmetric, known probability distribution on the interval $[0.8D_j, 1.2D_j]$ with D_j equaling the expected value.



(a) EVP, $\tilde{\mathcal{P}}_1$, $z^\star = 2373.1$ **(b)** 2SS-CFLP, $\tilde{\mathcal{P}}_1$, $z^\star = 2370.1$ **(c)** ARC-CFLP, $\tilde{\mathcal{P}}_1$, $^\star = 1952.8$

(d) EVP, $\tilde{\mathcal{P}}_2$, $z^\star = 19803.8$ **(e)** 2SS-CFLP, $\tilde{\mathcal{P}}_2$, $z^\star = 18953.8$ **(f)** ARC-CFLP, $\tilde{\mathcal{P}}_2$, $z^\star = 15511.4$

Figure 1.3.: Visualization of optimal uncertainty-aware solutions ($\tilde{\mathcal{P}}_1$-$\tilde{\mathcal{P}}_2$, Ex. C)

Figure 1.3 displays the optimal solutions to the CFLP in which all demand parameters have been set to their expected values (EVP), a two-stage stochastic program (see Subsection 7.2.2 for the mathematical programming formulation to 2SS-CFLP), and an adjustable robust counterpart (see Subsection 7.2.3 for ARC-CFLP). The stochastic program and the adjustable robust counterpart evaluate different objective functions. The stochastic program maximizes the expected profit, while the robust counterpart maximizes the worst-case profit using a max-min formulation. In consequence, a direct comparison of the objective values yields little insight. Nevertheless, compared to the expected value problem, the deviations of the objective value lie in similar ranges for both instances. The stochastic programming objective value is 0.1% ($\tilde{\mathcal{P}}_1$) and 4.3% ($\tilde{\mathcal{P}}_2$) lower than that of the expected value problem. At the same time, the robust counterpart yields a worst-case decrease of 17.7% ($\tilde{\mathcal{P}}_1$) and 21.7% ($\tilde{\mathcal{P}}_2$), respectively.

However, looking at the effect different modeling approaches have on the subset of facilities operating in the respective optimal solutions, the discrepancy between instances $\tilde{\mathcal{P}}_1$ and

$\tilde{\mathcal{P}}_2$ subjected to the same level of uncertainty is significant. For $\tilde{\mathcal{P}}_1$, an explicit model of uncertainty implies operating fewer facilities, more precisely, a subset of those facilities operating in the expected value problem. While 18 facilities operate in the latter, only 15 operate in the optimal solution to the stochastic program, and only 14 operate in the optimal solution to the adjustable robust counterpart. Meanwhile, for $\tilde{\mathcal{P}}_2$, seven facilities operate in the optimal solution to the expected value problem and the stochastic program, but only five operate in the solution to the adjustable robust counterpart. However, the set of operating facilities changes significantly between the three corresponding optimal solutions. The optimal solution to the stochastic program operates three, and the optimal solution to the adjustable robust counterpart operates two facilities not operating in the optimal solution to the expected value problem. ▲

Once again, a similar situation – a random perturbation of the demand of each customer within a defined magnitude – impacts the optimal solutions to two relatively similar problem instances differently. In Example C, the sharp discrepancy of these effects is not observable from objective values but only becomes visible when comparing solutions in the decision space.

The previous examples share that the same external impulse to seemingly similar problem instances leads to extremely different outcomes, particularly in terms of the optimality of individual location decisions. They demonstrate that the interplay between the problem's input data and the optimal decisions is insufficiently explored and point to the following fundamental question:

How are properties of the optimal solutions of facility location problems connected to properties of the input data, the performance of specific algorithms, and the potential value of model extensions?

Establishing a link between the problem's input data and the solutions to mathematical programs is a non-trivial task. While it can be of immense value, particularly in the identification of valuable modeling extensions, algorithm choice, or simply the generation of "challenging" instances to test new procedures, it is a rarely explored research direction. The parameter vectors describing a particular instance and the decision vectors denoting an optimal solution are high-dimensional data vectors, which, without further restrictions, can be extremely heterogeneous. With no conceptual framework that systematically links data analytics to the different aspects of mathematical programming, there are endless opportunities for exploratory research. Therefore, ensuring the generalizability of one's findings, given the infinite set of possible data vectors, as well as the lack of a concise objective or problem formulation, makes the establishment of any kind of connection between data, decisions, algorithm performance, and model choice very challenging.

We focus on the profit-maximizing CFLP as a core problem of location science. It is the basis for several supply chain network design models. As pointed out in Fadda et al. [2021], large parts of location literature developed in recent years focused on particular tailored models. Still, insights into the above research question are still missing, even for simple problem formulations.

1.2. Research questions and outline

Figure 1.4 provides an outline of this thesis. The overarching research question can be separated into three components: the relationship between decisions and the problem's input data, the decisions and the performance of solution algorithms, and the decisions and the added value of model extensions. Why do we focus on the relationships to the problem's decisions and not the problem's input data? The latter is readily available and does not require solving the mathematical program first. However, without further restrictions, the infinite number of possible input data combinations cannot be meaningfully linked to any part of the decision-making. Meanwhile, decisions combine the information from the problem's input data with the information from the specific restrictions of the mathematical program, making them a much more powerful indicator of the characteristic properties of the underlying decision problem. Still, in order not to perform a pure ex-post analysis, when we refer to decisions in the following, we do not restrict ourselves to the optimal values of the decision variables in the optimal solution to a particular problem. Instead, we are interested in characteristic properties "good", that is, optimal and near-optimal solutions to a particular problem instance share.

How are properties of the optimal solutions of facility location problems connected to properties of the input data, the performance of specific algorithms, and the potential value of model extensions?		
Characteristic decision patterns and link to input data	RQ1: What detectable decision patterns characterize well-performing solutions to a facility location problem?	CFLP
	RQ2: How can persistent service regions be identified in arbitrary sets of solutions?	
	RQ3: Can service regions be anticipated from the problem's input data?	
Link to solution algorithms	RQ4: How do service regions affect the performance of exact and heuristic solution procedures for the CFLP?	
	RQ5: Can the acknowledgment of service regions improve the performance of exact solution algorithms?	
Link to model extensions	RQ6: When is it worth explicitly considering time in the CFLP?	Multi-period CFLP
	RQ7: When is it worth explicitly considering uncertainty in the CFLP?	Stochastic & Robust CFLP

Figure 1.4.: Outline of the thesis

The first part of the thesis addresses the first component of the overarching research question: identify characteristic decision patterns optimal and near-optimal solutions to a particular problem instance of the CFLP share and connect them to characteristic properties of the problem's input data. We approach this objective by answering the specific research questions RQ1-RQ3 in Chapter 2 through Chapter 4. Once we have established an idea of what patterns characterize these solutions, we move on to the second component

of our overarching research question and examine their link to the performance of solution algorithms. More precisely, we address RQ4 and RQ5 in Chapter 5. Up to this point, we focus solely on the CFLP as a basic formulation of the underlying decision problem. Then, in the third component of our research question, we ask how these characteristic decision patterns are linked to the potential value of model extensions. In particular, we ask ourselves: When the problem's input data exhibits properties that favor a particular set of decisions, how likely is it that more comprehensive model formulations will alter these decisions rather than merely evaluate them differently? Or, when all optimal and near-optimal solutions to a problem instance exhibit a strong decision pattern, how likely is it that a more comprehensive model will alter this pattern? We move away from the CFLP to address this question. In Chapter 6 and RQ6, we analyze the effect of explicitly considering time in the problem formulation and look at the multi-period facility location problem with phase-in constraints. In Chapter 7 and RQ7, we consider the impact of explicitly modeling demand uncertainty. Here, we use a two-stage stochastic programming and an adjustable robust counterpart formulation of the CFLP.

In the following, we provide a short outline of the individual chapters.

Chapter 2 To identify characteristic decision patterns in facility location problems, we analyze sets of optimal and near-optimal solutions from various data sets found in the literature (RQ1). As a result, we develop concise and measurable indicators for these decision patterns, which can be used to compare the decisions of different solutions to a facility location instance. These measures allow us to demonstrate that some but not all instances exhibit a subset of favorable core facilities that persist throughout many or all well-performing solutions of a particular problem instance. Instead, we identify a persistent division of the facility-customer space into distinct service regions as a decision pattern that can be observed throughout these sets of solutions. This means that, in different solutions, customers may be served from different facilities but are persistently served from facilities within their service region. The service regions underlying different instances vary in size. When they are small in the sense that a single facility can serve its customers, this facility appears as a persistent core facility. Many works exploit the existence of subsets of such core facilities. However, when the service regions are large in the sense that their customers are served by more than one facility, then these facilities strongly interdepend, exhibiting the combinatorial nature of the mixed-binary problem formulation of the CFLP.

Chapter 3 Service regions cannot be derived from a single solution but require a set of solutions to capture the interdependence relationships between individual facilities. In Chapter 2, we observe service regions with different coherence levels in a specific set of solutions. In Chapter 3, we present an application of spectral biclustering to derive service regions with a defined target level of coherence from the aggregated allocation matrices of sets of solutions on the same facility-customer space (RQ2). This technique allows deriving service regions as the characterizing decision pattern in sets of solutions to a given instance that are not further specified, i.e.,

sets of infeasible solutions or sets of solutions to instance variations with perturbed parameters.

Chapter 4 Deriving service regions from sets of well-performing or even optimal solutions restricts insights to an ex-post analysis. In Chapter 4, we relate the level of interdependence between the facilities operating in an optimal solution, an indicator for the size of the underlying service regions, to characteristic properties of the problem's input data. In particular, we evaluate the relationship with the spatial pattern of the distribution of candidates and customers. The experiments show that overlapping effects of many different parameters characterizing a given instance make it impossible to anticipate the level of interdependence or distinct service regions purely from the data. We then demonstrate that service regions that characterize well-performing solutions can be obtained from sets of easy-to-obtain solutions, more precisely, integer-infeasible solutions.

The implicit division into service regions leads to some facilities being optimal "by themselves" and others being optimal only in combination, serving a larger subset of customers optimally. In Chapter 5, we connect these newborn insights into differences in the underlying decision patterns to the performance of exact and heuristic solution procedures.

Chapter 5 We examine the degree to which larger service regions whose customers are served optimally by a subset rather than individual facilities affect the performance of both heuristic and exact solution procedures (RQ4). In particular, we perform an exploratory analysis on instances exhibiting large and instances exhibiting small customer service regions. We solve both with the ADD heuristic, a variant of a greedy procedure, and the kernel search heuristic, one of the best-performing heuristics to date. We show that both heuristics perform exceptionally well when service regions are small, and each region is served optimally by just a single facility. In this case, the greedy procedure often finds the optimal solution. However, for larger service regions, the individually best facility is not necessarily part of the best subset serving a particular region, leading both the greedy and kernel search to good but decidedly non-optimal results. Ultimately, both procedures lack subroutines that explicitly check for interdependencies among subsets of facilities. We conclude that kernel search will likely benefit from ensuring that all facilities potentially serving a particular service region are jointly considered at least once.

We then evaluate the effect of large service regions on the performance of CPLEX's MIP solver as an example of a state-of-the-art exact procedure. We show that with an increasing size of the service regions, the size of the search tree of the branching procedure increases. In particular, the ignorance of the implied spatial relationship between facilities serving the same service regions leads the procedure to jump between facilities from different regions, leading to inefficient branch variable selection.

We take this information as a starting point for tackling RQ5 and use the regional clustering algorithm presented in Chapter 3 to detect service regions at the beginning

of the branching procedure based on potentially infeasible solutions early in the search tree. We then use this information to alter the branch variable selection strategy so that facilities in the same region are branched on soon after another. A computational experiment confirms the effectiveness of our procedure to speed up the solution process and reduce the size of the search tree.

We then approach the third and last component of our overarching research question by addressing the degree to which characteristic decision patterns, in particular the size of the implied service regions, affect the added value of modeling extensions. This, however, requires a concise definition of this added value.

Chapter 6 Location decisions are often long-lasting, and multi-period models have received considerable attention. The price to pay for the increased flexibility in decision-making is increased complexity. When facilities are capacitated, multi-period models impose additional restrictions regarding when to use resources. The value of the multi-period solution was introduced to quantify the benefit resulting from shifting to a multi-period model by setting its objective value in relation to that of a static counterpart. As the latter is not uniquely defined, different concepts of this value are reported in the literature. We show that different static counterparts require different interpretations and evaluate different aspects of the aforementioned trade-off. We discuss which characteristics of the problem data result in decision-makers benefiting from the added flexibility of multi-period modeling. Thereby, we show that rather than the degree to which the problem parameters change over time, the degree to which the multi-period model allows reducing relevant cost components determines the value of a multi-period approach.

Chapter 7 We compare the performance of a two-stage stochastic programming formulation and an adjustable robust counterpart formulation with budgeted uncertainty. We evaluate the average and worst-case value of the uncertainty-aware solution to compare the solutions obtained with a stochastic program to those obtained with an adjustable robust counterpart. We present a formal description of this unified evaluation approach. This allows us to address RQ7 and evaluate the interplay between characteristic decision patterns and the added value of explicitly considering uncertainty in the mathematical program. The results indicate that the added value of both the stochastic program and the robust counterpart decreases compared to a deterministic expected value solution with increasing separability of the facility-customer space into smaller service regions. Thereby, the latter can be derived with the help of the procedure introduced in Chapter 3 from the set of deterministic optimal solutions for the individual scenarios.

1.3. Data and computational set-up

The objective of this work is to gain a better understanding of the relationship between data, decisions, models, and solution algorithms in the context of facility location problems. A critical factor in this regard is that we validate our findings on a large database of

instances of interest. Thus, throughout this thesis, we perform extensive computational experiments on datasets from literature. We introduce these datasets and outline the computational set-up for our experimental validations in the following.

Data

Real-world data for capacitated facility location problems and their extensions is hard to obtain as abstract cost parameters and demand estimates are difficult to estimate, and sensitive strategic data is unlikely to be disclosed. Most works use self-generated test instances or re-use instances from literature. While some authors make their data available, no comprehensive online database exists.

The objective of this thesis is to link structural patterns in the optimal solutions to facility location problems to different aspects of mathematical programming. To test the validity of our findings, we survey existing datasets from literature to obtain a large base of test instances for our experiments. While the range of problem instances and their features is infinite, working with a large set of test instances from location literature allows us to draw general conclusions for problems of interest.

One can distinguish between instance generators and ready-to-use data sets. The former provide an algorithmic description of how to randomly generate problem instances that exhibit specified properties. The latter contain full-length parameter sets defining individual problem instances. These instances are often made available on the authors' personal websites or repositories of their affiliated universities. We present an incomplete overview of often-used data sets in Table 1.1. It shows that most problem instances are not based on real-world data but are artificially generated. To show that the instances worked with throughout this thesis are still relevant in literature today, we add some recent publications that use them for computational experiments.

Name	Type	Presented in	Recently used in	Used
Avella	G	Avella et al. [2009]	Guastaroba and Speranza [2012] Fischetti et al. [2016] Matos et al. [2020]	
Barcelo	G	Barcelo et al. [1991]	Contreras and Díaz [2007] Albareda-Sambola et al. [2011a] Gadegaard et al. [2018]	✓
Holmberg	G	Holmberg et al. [1999]	Di Francesco et al. [2021] Filippi et al. [2021] Saif and Delage [2021]	✓
Klose	G	Klose and Görtz [2007]	Görtz and Klose [2012] Fischetti et al. [2016] Gadegaard et al. [2018] Corberán et al. [2020]	✓
ORLIB	RW/G	Beasley [1990]	Fischetti et al. [2017] Pagès-Bernaus et al. [2019] Caserta and Voß [2020]	✓

G: generated, RW: real-world based

Table 1.1.: Overview of test instances for the CFLP

The following briefly introduces the data sets we use throughout this work. Together

with a brief introduction, we present a visualization of the spatial patterns underlying the instances. For this purpose, we follow the idea of Guazzelli and Cunha [2018] and perform Multidimensional Scaling (MDS) on the transport cost matrix to obtain the respective coordinates of facilities and customers in the plane. MDS is a dimensionality reduction technique from machine learning. It analyses the similarity or dissimilarity of data in the original high-dimensional space and then attempts to model it in a lower-dimensional space. We implement MDS using the *sklearn.manifold* package, taking the unit transport costs as input features. As there are usually more customers than facilities, we consider each customer as an observation and their distances to the respective facilities as features. Distances among facilities are usually not provided in the transport cost matrix, so they must be estimated. Thus, for each facility, we approximate the distance to another facility as the average distance to the four closest customers of both facilities. We reduce the resulting extended transport cost matrix with dimensions $((|J|+|I|)\times|J|)$ to a $(|J|+|I|)\times 2$-dimensional feature matrix. The two columns serve as a basis for the coordinates in a plane. We scale these two columns to fit into the first quadrant of a coordinate system to obtain the final coordinates.

Visual representations yield insights into the spatial patterns underlying different instances. They also show the proximity relationship between individual facilities that are not directly accessible when looking at the transport cost matrix. A central challenge in this work is translating observations from visual representations of instances and solutions into quantifiable measures that can be evaluated without visual assistance.

ORLIB instances

Beasley [1990] first presented the set of instances from the OR-library. It consists of 40 problem instances, *cap41* - *cap134*, that are based on factory locations in the United States first presented in Kuehn and Hamburger [1963], as well as three sets of larger instances that the authors generated themselves *capax* - *capcx*.

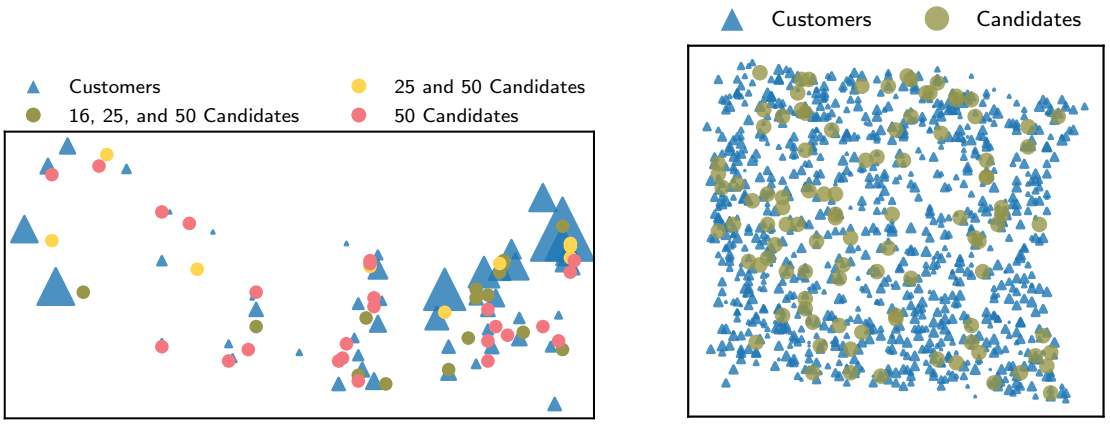
	$\|I\|$	Q_i	F_i
cap41-44		5000	7500/12500/17500/25000
cap61-64	16	15000	7500/12500/17500/25000
cap71-74		58268	7500/12500/17500/25000
cap81-84		5000	7500/12500/17500/25000
cap91-94	25	15000	7500/12500/17500/25000
cap101-104		58268	7500/12500/17500/25000
cap111-114		5000	7500/12500/17500/25000
cap121-124	50	15000	7500/12500/17500/25000
cap131-134		58268	7500/12500/17500/25000

Table 1.2.: Overview of parameter variations in the ORLIB instances *cap4x-cap13x*

We consider 36 of the 40 US-based instances and further refer to them as the ORLIB. These 36 instances systematically differ in the number of candidate locations $|I|$, fixed costs F_i, and capacities Q_i. There are 12 instances of 16, 25, and 50 candidate sites, respectively. Within each of these 12 instances are four instances in which facilities have a fixed capacity of 5000, 15000, and 58268 (equaling the total network demand), respectively.

Then, within these four instances, the fixed costs are varied such that the ratio fixed costs of operating one facility account for approximately 1%, 2%, 3%, and 4% of the variable transportation costs, respectively. Table 1.2 provides an overview.

Figure 1.5a shows a visual representation obtained via MDS. One can recognize that the instances are derived from a US-based problem. One can see the large metropolises on the west coast and the densely populated area around New York City on the east coast. Meanwhile, the population in the mid-western regions is sparse.



(a) *cap41-cap134* **(b)** *capax-capcx*

Figure 1.5.: Visualization of the ORLIB and ORLIB-L instances

The self-generated data, further referred to as ORLIB-L, is based on randomly located customers and facilities on a unit square. The unit transportation costs are derived from the scaled Euclidean distances, whereby the scaling factor is drawn from $U[1, 1.25]$, with $U[a, b]$ the uniform distribution on the interval $[a, b]$. Customer demands were drawn from $U[1, 100]$. The facility capacities are varied, and the fixed costs were chosen such that a desired number of sites should be open in the optimal solution. Table 1.3 lists both. Figure 1.5b shows an exemplary visual representation obtained via MDS.

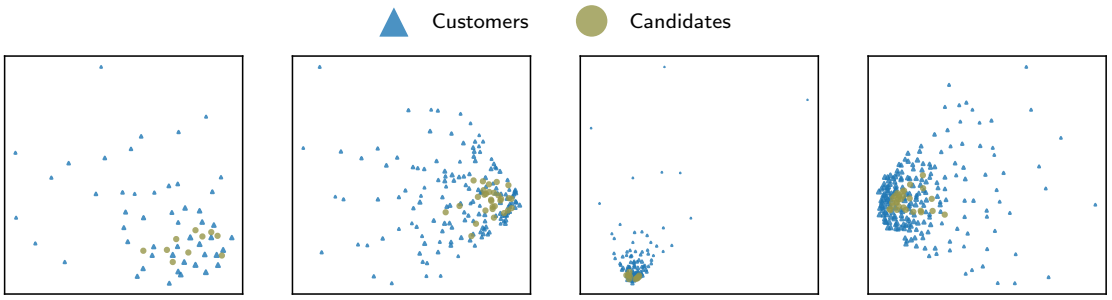
	$\|I\|$	$\|J\|$	F_i	desired number of sites
capa	100	1000	8000/10000/12000/14000	5
capb	100	1000	5000/6000/7000/8000	10
capc	100	1000	5000/5750/6500/7250	15

Table 1.3.: Overview of parameter variations in ORLIB-L instances *capax-capcx*

The ORLIB instances have been widely used (refer to Table 1.1). Several authors modified or extended the data to qualify for extensions of the CFLP, e.g., to include capacity acquisition costs Verter and Dincer [1995], or stochastic transportation costs Pagès-Bernaus et al. [2019].

Holmberg et al. [1999] instances

Holmberg et al. [1999] provide four randomly generated sets of test problems. Their properties are summarized in Table 1.4. For problems *p1-p24* and *p56-p71*, locations



(a) *p1* **(b)** *p25* **(c)** *p41* **(d)** *p56*

Figure 1.6.: Visualization of some HOL-1999 instances

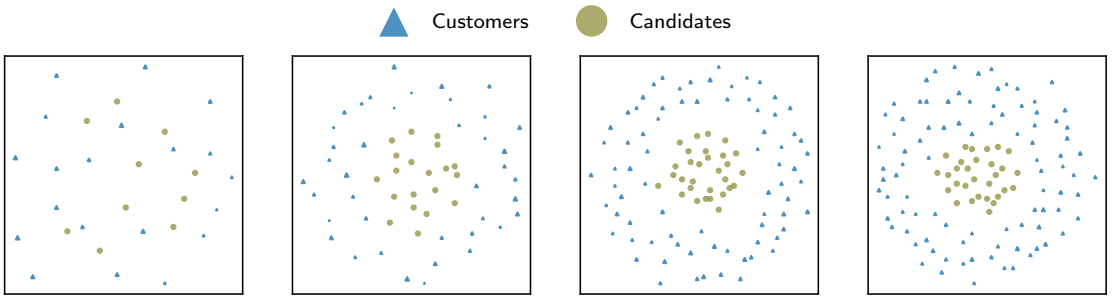
ID	$\|I\|$	$\|J\|$	D_j	Q_i	F_i
p1-p12	10	50	$U[10, 50]$	$U[100, 500]$	$U[300, 700]$
p13-p24	20	50	$U[30, 80]$	$U[100, 500]$	$U[300, 700]$
p25-p40	30	150	$U[10, 50]$	$U[200, 600]$	$U[300, 700]$
p41-p55	10-30	70-100	$U[10, 50]$	$U[100, 500]$	$U[300, 600]$
p56-p71	30	200	$U[10, 50]$	$U[500, 800]$	$U[500, 1500]$

Table 1.4.: Overview of parameter variations of HOL-1999 instances

were uniformly randomly chosen on a unit square. Subsequently, a p-median heuristic was applied to determine the subset of facility locations. Unit transportation costs are based on Euclidean distances. Problems *p25-p40* are generated similarly, except that unit transport costs are now based on transport costs obtained from a vehicle routing problem. Problems *p41-p55* are modified instances from a vehicle routing problem with varying patterns of customer locations. Demands, capacities, and fixed costs were drawn from different uniform distributions. Figure 1.6 shows exemplary visual representations of the five sets obtained via MDS. One can see that a very wide distribution of customers can characterize all instances while candidates are placed relatively close to each other. This is surprising given the random placements of candidates and customers on the unit square and must be a result of the procedure to estimate transport costs. We will refer to this dataset by HOL-1999.

Barcelo instances

This set contains 57 instances subdivided into seven sets, $C_1 - C_7$, according to the number of customers and candidate facilities. Their sizes range between 10 and 30 facilities and 20 to 90 customers. Delmaire et al. [1999] provide a detailed description of sets $C_1 - C_4$, whereas a brief description of the generation procedure of sets $C_5 - C_7$ can be found in Barcelo et al. [1991]. The authors state that the instances were generated randomly, whereby their objective was to form "very tight to very loose situations, from the point of view of the relationships among costs, capacities, and demands." Figure 1.7 visually represents the instances. One can see that a circle best captures the spatial relationship between customers and candidates. This implies little heterogeneity between individual candidate locations regarding their transport costs. We refer to this dataset as BAR-1991.



(a) $p5$ **(b)** $p23$ **(c)** $p47$ **(d)** $p58$

Figure 1.7.: Visualization of some BAR-1991 instances

Data generator by Andreas Klose

Data generators for static CFLP instances are, e.g., provided by Andreas Klose (`https://github.com/AndreasKlose/CFLP-Generator`) and (his co-author) Sune Gadegaard (`https://github.com/SuneGadegaard/SSCFLPgenerator`). Both generators implement a generation procedure for facility location problems proposed by Cornuejols et al. [1991]. It consists of the following steps:

1. The locations of facilities and customers are generated by uniformly randomly locating them on a unit square. Transportation costs per unit are then computed by multiplying the resulting Euclidean distances by 10.

2. The customer demands are drawn from $U[5, 35]$, with $U[a, b]$ denoting a uniform distribution on the interval $[a, b]$.

3. The facilities' capacities are drawn from $U[10, 160]$ and subsequently scaled according to a predefined "tightness" ratio, the ratio between the total capacity and the total demand, such that

$$tightness := \frac{\sum_i Q_i}{\sum_j D_j} \Leftrightarrow Q_i^{new} = Q_i \cdot \left(\frac{r \sum_j D_j}{\sum_i Q_i} \right), \tag{1.6}$$

   with Q_i the capacity of the i-th facility and D_j the demand of the j-th customer.

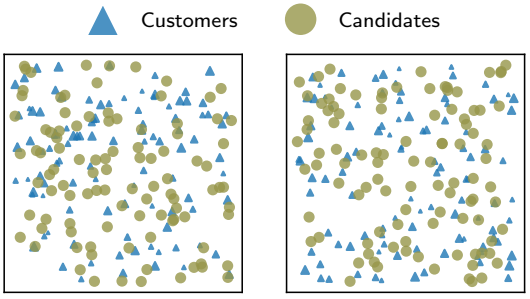
4. Fixed costs, F_i, of a facility i are generated according to the following formula, which takes into account the facilities' capacity to reflect economies of scale

$$F_i = U[0, 90] + U[100, 110]\sqrt{Q_i}. \tag{1.7}$$

The generator from Andreas Klose was, e.g., used to generate test problems used in Klose and Görtz [2007] and Görtz and Klose [2012]. Besides the procedure from Cornuejols et al. [1991], Sune Gadegaard further implements a generation procedure according to Stidsen et al. [2014]. The procedure primarily differs in the ranges for the uniform distributions and the definition of the fixed costs. In the latter procedure, the fixed costs for facility i equal $(\min_j c_{ij} + 1) \cdot 10$.

We generated instances according to the parameter settings of the generator used in Klose

and Görtz [2007]. These instances vary in the number of customers and candidate locations and the capacity-demand ratio. The latter is called tightness and equals the total sum of capacity divided by the total sum of demands. We generated five instances with 50, 100, and 200 candidates and customers each for tightness levels of 3 and 10, respectively. Thus, we consider a total of 30 instances. Figure 1.8 shows an exemplary visual representation obtained via MDS. We refer to these instances as KLO-2007.



(a) $|I| = 100$, $|J| = 100$, $tightness = 3$ **(b)** $|I| = 100$, $|J| = 100$, $tightness = 10$

Figure 1.8.: Visualization of some KLO-2007 instances

We point out that all available test problems are defined for cost-minimizing and not profit-maximizing formulations. Therefore, we generate values for the unit reward r_j. To maintain the structure of the resulting solutions compared to the cost-minimizing problems, we choose r_j so that it is profitable to serve all demands and open the same number of facilities as in the cost-minimizing problem. We identify such a value of r_j by successively increasing it while resolving the model and comparing the number of facilities operating in the optimal solution to that in the cost-minimizing problem.

Computational set-up

Computational experiments are performed on an Intel(R) Core(TM) i7-7700 processor with 3.60 GHz and 64 GB RAM. Mathematical programs were solved using CPLEX version 12.10 and implemented using the `docplex` interface in `python` version 3.7. Unless specified otherwise, the target MIP gap was set to 0.01%.

With the described target MIP gap, the optimal solutions to 178 of the 206 considered instances have a unique optimal solution. For 25 instances, another solution within 0.01% of the optimal solution could be found. More than one solution within 0.01% of the optimal solution was identified for three instances.

2. Characteristic decision patterns: from core facilities to service regions

A prerequisite for establishing a link between a problem's input data and its well-performing, or even optimal, decisions is the identification of measurable characteristics that summarize their essential properties. Otherwise, one faces two high-dimensional data vectors, and any relationship that might be established with state-of-the-art data science techniques is difficult to explain and validate. Example A shows that individual facilities in the optimal solutions to different problem instances of the CFLP exhibit different characteristics regarding their stability and contribution to the optimal objective value. The optimal solutions to $\mathcal{P}_1$ and $\mathcal{P}_2$ are obtained using the same model and solved with the same off-the-shelf MIP solver. Consequently, differences in the problem's input data must explain the observed differences in the optimal solutions. Yet, at this point, even a formal description of the observed behavior regarding the stability of individual location decisions is missing. The upcoming chapter addresses the following research question:

> ***RQ1:*** What detectable decision patterns characterize well-performing solutions to a facility location problem?

RQ1 lays the foundation for answering our superordinate research question by identifying properties that can be detected and consequently used to classify and compare the decisions of different solutions to arbitrary facility location instances. Decision patterns that persist throughout well-performing – that is, optimal and near-optimal – solutions yield insights into the characterizing properties of a problem's solution in the decision space. To identify these patterns, we perform different exploratory experiments comparing multi-sets of well-performing solutions to a particular instance. In a detailed analysis of the instances $\mathcal{P}_1$ to $\mathcal{P}_4$ from Example A, we observe characteristic patterns and translate them into concise measures. Thus, "detectable" in this context means that the existence and relevance of found patterns can be evaluated and quantified on arbitrary instances.

There are several challenges to overcome. Firstly, even though the systematic relationship between the performance of certain location decisions and the spatial characteristics of a problem instance has long been recognized, the structured analysis of the individual decisions of a solution is a largely overlooked issue in Operations Research (OR) literature in general, and literature on facility location problems in particular [Mai et al., 2018]. Most OR literature focuses on analyzing prescriptive solutions in the objective value space, i.e., the value of the objective function. At least in the theoretical discussion, the prescribed decisions to achieve this value remain a side concern. This is challenging when gathering related work, as consistent terminology is lacking. Secondly, shifting the focus from objective value to decision space is non-trivial. While the objective value is usually a real-valued number, decision vectors are high-dimensional. This makes them much more difficult to describe and analyze. Furthermore, the objective value is minimized or maximized, making it easy to compare and evaluate solutions in this regard. High-dimensional decision vectors lack an optimization sense, making an ordinal comparison challenging. When de-

tecting persistent decision patterns in sets of well-performing solutions, it is intuitive to compare graphical representations of individual solutions visually. The challenge lies in translating these observations into concise measures that can be evaluated on arbitrary instances without visual assistance.

We review related work in the upcoming Section 2.2, at the end of which we highlight our contributions to the current debate and provide a short chapter outline.

2.1. Related work on the analysis of patterns in data and decisions

Most works on location problems focus on the mathematical model, the solution procedures, and the resulting objective value. If at all, properties of the problem instance are discussed as part of the data generation procedure for the computational experiments. The structure of solutions in the decision space is rarely examined. This poses a challenge when gathering related work, as neither a conceptual framework nor consistent terminology exists.

Table 2.1 presents an overview of the topics in literature that we think are connected to the current work. Firstly, we look at sensitivity analysis as an ex-post analysis tool that examines the relationship between minor perturbations of the data and optimal decisions (Subsection 2.1.1). Secondly, we look at heuristics (Subsection 2.1.2). First, we survey problem-specific heuristics that identify "good" or "bad" candidate locations from the problem data before solving the problem once (Subsection 2.1.2.1). We then review heuristics that exploit the fact that favorable candidate facilities often operate in many or all near-optimal solutions and take on high values in the linear programming relaxation. This implied dominance of certain candidate locations is used, e.g., in heuristic concentration and kernel search (Subsection 2.1.2.2). Thirdly, we review works that try to analyze the optimal solutions to facility location problems ex-post, that is, after solving the mathematical program, to better understand the problem at hand (Subsection 2.1.3). These works examine commonalities in sets of solutions to a particular problem instance. Lastly, we look at recent approaches coupling machine learning algorithms with facility location problems to learn the relationships between the input data and the optimal decisions (Subsection 2.1.4). We provide a short overview of the main findings in these areas and point out the differences to the research questions discussed in the present work. We conclude by pointing out our contributions to the current state of literature.

2.1.1. Sensitivity analysis

Sensitivity analysis is the systematic study of how (minor) changes in the data affect the optimal solution to a problem. It is a post-optimality analysis tool performed after an optimal solution is available. In linear programming theory, comprehensive techniques for sensitivity and parametric analysis exist. They directly follow from the problem structure and duality theory [Hillier et al., 1997]. As location problems usually comprise binary location decisions, linear programming sensitivity analysis cannot be applied. While systematic, theoretical approaches for sensitivity analysis for integer and integer linear programs

Topic	Relation to present work
Sensitivity analysis	Relate perturbations in the data to changes in the optimal decisions.
Heuristics: Data-based selection rules	Link data and decisions by identifying general characteristics of "good" (location) decisions.
Heuristics: Kernel-based methods	Exploit persistent decision patterns in the form of facilities that operate in many optimal and near-optimal solutions.
Operating frequency of facilities	Identify persistent facilities across (multi-)sets of solutions by their operating frequency.
Learning optimal location decisions	Learn relationship between perturbations in the data and changes in the optimal solution.

Table 2.1.: Relation between topics addressed in literature review and present work

have been developed [Geoffrion and Nauss, 1977, Dawande and Hooker, 2000], to the best of our knowledge, they have yet to be applied to location problems.

Instead, in the context of location problems, two different approaches to sensitivity analysis can be found. Firstly, several works perform in-depth, theoretical sensitivity analysis in very restrictive problem settings. For example, Drezner [1985] analyze the Euclidean, single-facility mini-sum problem, also known as the Weber problem. The authors examine the sensitivity of the optimal site of the new facility to changes in the locations and weights of the customers. Labbé et al. [1991] examine a discrete 1-median problem. The authors evaluate trade-offs between cost and perturbations of the demand volumes.

The superordinate question in these works is: How much can the data change before optimal decisions change? The present work looks at this problem from a different angle. What characteristic decisions persist when the data changes to a given degree? The focus is on the type of change induced in the decision vector.

In more complex problem settings, sensitivity analysis is generally conducted on an experimental basis. The effect of different variations of individual parameters is evaluated by repeatedly solving the problem under different parameter constellations. The changes in the resulting solutions are evaluated with respect to the objective value or problem-specific performance measures [Kılcı et al., 2015, Hu et al., 2019]. Very few works take a closer look at the effects of these perturbations on individual decisions. The analysis is often reduced to evaluating the number of operating facilities and less concerned with which individual facilities operate [Pelegrín et al., 2012]. Few authors examine the degree to which changes in parameters result in changes in the individual optimal location decisions. Alumur et al. [2012b] perform such an analysis to show that their solutions are robust towards parameter changes, and thus, the optimal locations do not change. Jones et al. [2022] examine the location of robotic devices at health facilities and list the optimal locations under different parameter settings. They evaluate the frequency of individual location decisions across different solutions, which is the most comprehensive analysis of individual location decisions we could find. Thereby, without explicitly stating it, the findings of both works support the notion that in problem instances to discrete location problems, certain facilities dominate in the sense that they operate in many or even all optimal and near-optimal solutions.

One of the main findings in the upcoming chapter is that such dominant subsets of core facilities exist, but not necessarily for every instance. Nonetheless, persistent decision patterns aside from dominant location decisions can be identified.

2.1.2. Heuristics

Heuristics often exploit characteristic patterns in the problem's input data to identify optimal decisions or the problem's solution during their search process. We review two types. The first deploys some data-based selection rule to rank candidate locations and then uses this rank to identify either the candidates themselves in a greedy manner or promising solution neighborhoods in a local search. Thereby, these works link characteristics in the data to favorable decisions. We review these heuristics in Subsection 2.1.2.1. The second type exploits that many optimal solutions of facility location problems have a relatively stable integer part, namely the previously mentioned subset of core facilities. The search process is designed to find these stable facilities. Kernel search and heuristic concentration benefit from stable integer decisions and work particularly well on discrete location problems (Subsection 2.1.2.2).

2.1.2.1. Data-based facility selection rules

To the best of our knowledge, Park [1989] were the first authors to acknowledge the systematic relationship between the performance of certain location decisions and the spatial characteristics of an instance. The authors present seven simple rules for selecting candidate locations for a service delivery system based on spatial properties derived from the problem data. Each rule operates by assigning a score to each candidate location. Candidates with the highest or lowest scores are selected sequentially. While more complex rules require an algorithmic subroutine, others assign scores purely based on the problem data. The simplest rules consider only data regarding that specific candidate facility, e.g., the size of that facility. Other rules, like the maximum geographic isolation of demands and the minimum regret of eliminating candidates, include data from all candidates in the system. They either add candidates to an empty set or drop candidates from the complete set of all candidates in a greedy manner according to these scores. The authors evaluate their rules in a computational study based on real-world data from Iowa in the United States. They hypothesize that some rules perform well because the geographic pattern of demands in a spatial system has specific properties that the rule implicitly assumes to exist. According to the authors, simple rules are likely to perform well in an area with a well-developed structure. Thereby, they imply the generalizability of their study to other instances of interest. According to the authors, theoretical reasons exist for expecting real-world environments to exhibit specific characteristics. One of them is the iterative adjustment of providers and consumers to be close to one another, which results in a spatial structure similar to a p-median solution. Also, once services are located in certain places, the urban multiplier effect results in populations of such places being proportional to the number of services in place. This has been validated in empirical studies of central places [Berry and Barnum, 1962, Mulligan, 1980].

Peker et al. [2016] determine general characteristics of optimal hub locations by examining optimal solutions to data sets for hub location problems from the literature. They derive rules to identify subsets of candidate hubs that meet these criteria. From subsequent experiments, the authors conclude that simple measures that combine the magnitude of demand with measures representing the relative spatial dispersion of the candidates are effective in finding optimal or near-optimal hub locations. They conclude that by identifying favorable candidates or regions for hubs, it may be possible to speed up the solution process by focusing further analysis on promising regions.

Mu and Tong [2018] propose including spatial knowledge on the problem instance into heuristic solution procedures of p-median problems. In an iterative search, the authors use data-driven selection rules to identify candidates for swapping facilities in the current solution. The score of a candidate facility is based on the demand density in their neighborhood. The demand density is the sum of demand within a given radius of that facility. Eskandarpour et al. [2017] present a set of rules to identify both "good" and "bad" candidates based on which they determine promising swaps in a large-scale variable neighborhood search for a multi-layer, multi-modal supply chain network design problem. They evaluate a large set of rules by comparing the objective value attained with this rule with the objective value attained with a random choice of candidates. Interestingly, no operator outperforms another, and each brings specific contributions. They recommend the simultaneous use of several operators.

The above works look for general characteristics of well-performing facilities across instances. As problem instances may be arbitrarily heterogeneous, some works argue why the subset of instances they consider represents all instances of interest. In contrast, we are interested in general means to identify the characteristics of optimal and near-optimal solutions to a particular instance. Thus, given an optimal solution to a problem instance, we look for the meta-information on the relevance, stability of, and the relationship between individual location and allocation decisions.

2.1.2.2. Core-based methods

Heuristic Concentration (HC) and Kernel Search (KS) have been applied very successfully to facility location problems. These meta-heuristics rely on considering only subsets of the integer decisions in an iterative search, thereby benefiting from situations in which the integer part of the solution is relatively stable.

Rosing and ReVelle [1997] first presented HC for a p-median problem. In the first stage, solutions to the problem obtained with some previously known base heuristic, e.g., the interchange heuristic, are used to create a concentration set of potential candidate locations for a new, smaller optimization problem, which can then be solved to optimality by standard optimization techniques as branch-and-bound. Multiple starts of the base heuristic in the first stage are used to produce a list of all candidates that operated in the m best unique solutions of q individual runs. The resulting list is called the concentration set. The latter is assumed to have a high probability of containing the facilities that comprise the still smaller set of optimal facilities. The authors claim that "... it is generally true

that solutions whose functional values differ little also are derived from largely identical solution sets". They provide an example in which the best solutions obtained from the 1-opt exchange heuristic mostly comprise the optimal nodes with a few exceptional nodes. They observe that these exceptional nodes are usually close to each other and form a small and localized trap that a 1-opt heuristic cannot identify. Since these traps occur in different network regions for every 1-opt solution, the union of the nodes of the best solutions tends to capture the optimal solution. Rosing et al. [1999] propose a variant of HC in which the reduced problem in the second stage is not solved to optimality but with another heuristic, called Gamma heuristic.

Marianov et al. [2009] propose a modified version of HC called Heuristic Concentration Integer (HCI). HCI uses a 2-opt heuristic in the second stage. The main difference to the Gamma heuristic presented by Rosing et al. [1999] is the ability of HCI to pick the same element repeatedly. Another difference lies in the definition of the concentration sets. They now contain each location that was selected at least once in any solution of the first stage. HC and HCI have been applied to several combinatorial optimization problems, particularly various location problems. ReVelle et al. [2008] apply HC to solve the maximal covering location problem. Jayaraman et al. [2003] and Geng and Wang [2014] apply HC to a model for reverse logistics with collection sites. Loree and Aros-Vera [2018] extend HCI to solve the problem of placing points of distribution in post-disaster logistics. Other recent applications include home healthcare routing and scheduling Grenouilleau et al. [2019] and vehicle routing with stochastic demands and duration constraints Mendoza et al. [2016]. Sanci and Daskin [2019] combine Sample Average Approximation with HC in the context of humanitarian relief logistics. To the best of our knowledge, they are the first to extend HC to problems involving uncertainty. The authors emphasize that while it might be the first thought to solve the deterministic equivalent MIP for each scenario independently and derive concentration sets from the union of the selected nodes, this approach did not yield robust solutions, meaning that the set of nodes in the individual solutions can differ significantly. Thus, the authors randomly split the scenarios into subgroups, solve the stochastic program for each of these subgroups, and derive the concentration set from the resulting solutions.

KS is a matheuristic that was first proposed by Angelelli et al. [2010]. KS identifies a subset of variables with a high probability of taking a non-zero value in the optimal solution and refers to it as a kernel. The remaining decision variables are partitioned into buckets that are successively concatenated with the kernel and then mutually searched. The kernel is continuously updated during this process, and a sequence of restricted mixed-integer linear programs is solved. The initial kernel is derived from the Linear Programming (LP) relaxation of the problem. KS iteratively explores promising portions of the solution space. The heuristic has been applied to the cost-minimizing CFLP by Guastaroba and Speranza [2012] and to the single-source CFLP in Guastaroba and Speranza [2014]. Other applications of KS to location models can be found in Zhang et al. [2019]. Compared to p-median problems, the CFLP bears the additional difficulty that each location decision is linked to many allocation decisions. The kernel for the set of allocation variables has to

be kept consistent with the set of variables for the location variables.

The performance of the above heuristics likely improves when the sets of facilities operating in the near-optimal and optimal solutions largely overlap. Thus, the fact that they have been identified as well-performing solution methods for discrete facility location problems further suggests that these problems exhibit stable subsets of core facilities. One of the main conclusions from our upcoming analysis is that while this notion persists throughout several works in literature, it is valid for some but not all instances.

2.1.3. Operating frequency of individual facilities

Few works examine multi-sets of solutions to either similar instances or similar models. These analyses focus on detecting subsets of core facilities that operate in most or even all of the solutions in the set. Fadda et al. [2021] compare solutions to a service facility location problem obtained from different coverage location models. Besides evaluating the solutions regarding several key performance indicators, the authors determine the frequency with which individual facilities operate across the solutions to different models. They define a core as a subset of candidate facilities that operate in the optimal solution of at least a defined number of the considered models. Based on this core, they evaluate the relative importance of core locations for a particular instance. They state that the higher the relevance of these core locations, the less important is the model choice. The authors then relate the existence of core locations to the spatial properties they determine during their data generation procedure. In particular, they find that the occurrence of core locations is higher for instances where the spatial distribution of candidates and customers is not derived from a uniform distribution. They conclude that core solutions result from non-random spatial properties of an instance.

To the best of our knowledge, the only thorough analysis of solutions to the CFLP is performed by Guazzelli and Cunha [2018], who examine the structure of k best solutions to a particular instance. The authors base their study on two sets of problem instances from the literature. The first set contains the 36 instances from the OR-library [Beasley, 1988] whose spatial distribution resembles the population distribution of the United States (see ORLIB in Section 1.3). The second set composes 20 benchmark instances from Yang et al. [2012] whose spatial distribution resembles randomly spread points on a unit square. The authors compute k best alternative solutions for each of these instances, whereby k ranges between 20 and 200 depending on the problem size. They limit all considerations to the binary location decisions, meaning that two solutions are only considered to be different if the opening status of at least one facility differs among them. The authors then analyze the resulting objective values, the set of operating facilities, and several performance indicators. The results of this analysis are mostly intuitive. The number of operating facilities decreases with increasing fixed costs and capacities in the optimal solutions to the individual instances. When considering the k best solutions, the objective value degraded more slowly for instances with more candidates. Furthermore, k best solutions are often relatively stable and primarily differ only in a single facility. For the instances from the ORLIB data set, the ratio of simple changes, merely adding or removing

one facility or replacing one facility with another regarding the optimal solution, is 47.2%. This ratio decreases to 16.1% for the second set of benchmark instances. The authors suppose this is due to the randomized geographical distribution, which results in less clustering.

Both works above support the existence of core facilities. They even state that the relevance of such cores increases with more "structured" spatial patterns underlying the problem data, i.e., locations obtained from real-world settings, compared to instances based on a uniform distribution of candidates and customers on a unit square. Yet, a central shortcoming at this point is the lack of concise measures to detect and quantify the relevance of these core facilities in the well-performing solutions to a particular instance. We provide such a measure, which allows us to evaluate the degree to which these observations can be generalized.

2.1.4. Learning optimal location decisions

Recent works integrate statistical machine learning into operations research. Thereby, the problem data serves as an input for statistical learning algorithms. Mai et al. [2018] propose a heuristic based on the expectation-maximization algorithm to a capacitated p-median problem using the geographic locations and demands of customers as inputs. Amongst others, the authors evaluate the effectiveness of their algorithm on self-generated instances with different spatial patterns, in particular, random patterns, regular patterns, and patterns exhibiting a natural clustering. They observe that their algorithm performs particularly well compared to other approaches in instances with clustered spatial distribution of candidates and customers.

Lodi et al. [2020] present an approach to couple the solution of a single-source CFLP with methods from machine learning. For a given problem, the authors assume that data exists for many variations of this problem (historical or simulated) along with their optimal solutions. They then examine how one can use this data to predict the optimal solution to a new unseen variation of the reference problem. More specifically, the authors aim to predict the number of facilities that are expected to change, thus the amount of re-planning needed, given a perturbed instance by different learning algorithms such as tree regressors, neural network classifiers, logistic regression, classification, and naive Bayes classification. The resulting information is then included in the mathematical program as an additional constraint. The authors conclude that their approach can identify whether a perturbed instance derived from a reference one will share all or a part of its optimal facilities with the solution of the reference instance.

Both works suggest that spatial patterns in the problem's input data lead to relatively stable subsets of core facilities. However, as Mai et al. [2018] observe, in the absence of spatial regularities, the dominance of these subsets of facilities diminishes, and the performance of algorithms that implicitly expect them decreases. We look closer at the types of characteristic patterns that can be expected. In Chapter 4, we explicitly link these patterns to the performance of individual algorithms and demonstrate which decision patterns are particularly difficult to identify.

The review of literature shows that the existence and relevance of structural patterns, in particular the spatial characteristics underlying the problem's input data, is widely acknowledged and explicitly or implicitly exploited in several well-performing heuristics. However, a systematic approach to address RQ1 is missing. Instead, existing works imply that many instances exhibit a subset of favorable core locations that characterize optimal and near-optimal solutions. However, already Example A indicates that the persistence of individual facilities throughout well-performing solutions differs massively amongst facilities and instances.

In Section 2.2, we present a concise measure to quantify the size and relevance of persistent subsets of core facilities in a particular set of solutions. We then perform extensive experiments on k best alternative solutions to a large set of instances from literature as presented in Section 1.3. Results indicate that persistent subsets of core facilities exist in some but not all instances. In Section 2.3, we extensively analyze a particular set of solutions to detect persistent patterns. We identify an implicit division of the facility customer space into service regions of different coherence levels as a characterizing pattern that well-performing solutions to arbitrary instances share. We end with a summary of our main findings and an outlook in Section 2.4.

2.2. Persistent subsets of core facilities

This work aims to understand the interplay between characteristics in the problem instance and well-performing, or even optimal, decisions. A preliminary for the analysis of this relationship is the identification of the persistent, favorable decision patterns that near-optimal and optimal solutions to a particular problem instance share. Existing works in literature suggest that facility location problems exhibit a persistent subset of core facilities that operate in many or all optimal and near-optimal solutions to a particular instance. Several well-performing heuristics implicitly or explicitly exploit this property. Works that focus on identifying stable decision patterns in multi-sets of solutions to particular instances support the idea of their existence. The level of persistence of these core facilities has been related to the spatial patterns underlying a problem instance with the hypothesis that random spatial patterns lead to less persistent cores. However, up to this point, no quantitative analysis on the existence of such core locations exists. In particular, there is a lack of a concise measure to quantify the relevance of such core facilities in the well-performing solutions to a particular instance. Consequently, the relationship between the persistence of core facilities and the underlying spatial patterns of the data is not confirmed and must yet be validated.

In the upcoming section, we extend the work of Guazzelli and Cunha [2018] on exploring persistent decision patterns in k best alternative solutions to a CFLP. In Subsection 2.2.1, we develop a concise measure to quantify the relevance in terms of size and persistence of subsets of core facilities in a particular multi-set of solutions. Thereby, we draw on ideas first presented by Fadda et al. [2021]. We apply this measure to the 5 best alternative solutions to $\mathcal{P}_1$-$\mathcal{P}_4$ from Example A, all of whom have been generated on a random spatial pattern. The results indicate significant differences in the relevance of core facilities,

putting the existence of this decision pattern in general instances and its relationship to random spatial patterns in question. In Subsection 2.2.2, we perform an experiment on the instances from the data sets described in Section 1.3 to validate that, firstly, not all instances exhibit persistent subsets of favorable core locations, and, secondly, their existence is not related to the randomness of the underlying spatial pattern.

For a consistent notation, let the tuple $s := (z, y, x)$ denote a solution to a CFLP, whereby z is the objective value associated with the decisions $x \in R_{\geq 0}^{|I|\times|J|}$, the matrix of allocation decisions, and $y \in B^{|I|}$, the vector of binary location decisions. Let $s^\star := (z^\star, y^\star, x^\star)$ denote the optimal solution and $\mathcal{S}$ an arbitrary set of solutions to a particular instance.

2.2.1. Measuring the relevance of core facilities

We extend the work of Fadda et al. [2021] to derive a concise measure that allows quantifying and comparing the relevance of subsets of core facilities in an arbitrary (multi-)set of solutions to a given instance. The measure considers not only the persistence of a particular subset of operating facilities but also the size of this subset compared to the number of facilities operating in the optimal solution and the number of candidates. Fadda et al. [2021] define a kernel as the subset of candidate locations operating in a given proportion of the considered solutions. The focus is solely on the location decisions. We formally take over this idea as follows.

Definition 2.1. *The **t-kernel** of a set of solutions $\mathcal{S}$, $N_c(t, \mathcal{S})$, is the set of candidate facilities that are open in at least $t \cdot 100\%$, $t \in [0, 1]$, of the solutions in $\mathcal{S}$, such that*

$$N_c(t, \mathcal{S}) := \left\{ i \in I \;\middle|\; \frac{\sum_{s\in\mathcal{S}} y_i^s}{|\mathcal{S}|} \geq t \right\}, \tag{2.1}$$

with y_i^s denoting the value of the decision variable indicating the opening status of facility $i \in I$ in solution $s \in \mathcal{S}$.

Notice that when $t = \tau/|\mathcal{S}|$, with $\tau \in \mathbb{N}_0$, then $N_c(t, \mathcal{S})$ denotes the set of candidate locations that are open in exactly τ solutions.

Fadda et al. [2021] measure the relevance of a t-kernel in a given set of solutions by dividing the cardinality of $N_c(t, \mathcal{S})$ by the cardinality of the set of candidate locations. However, this makes it difficult to compare the relative importance of t-kernels between different instances. The fraction decreases with the number of candidate locations, even if these are irrelevant, i.e., if they are not part of any solution in $\mathcal{S}$. Therefore, we set the size of a t-kernel in relation to the set of relevant candidate locations in $\mathcal{S}$.

Definition 2.2. *The **set of relevant candidates** in a set of solutions $\mathcal{S}$, $I^R(\mathcal{S}) \subseteq I$, is the subset of candidate locations that operate in at least one of the solutions in $\mathcal{S}$, such that*

$$I^R(\mathcal{S}) := \{i \in I \mid \exists s \in \mathcal{S} \; s.t. \; y_i^s = 1\} \iff I^R(\mathcal{S}) = N_c\left(\frac{1}{|S|}, \mathcal{S}\right). \tag{2.2}$$

To quantify the relevance of a t-kernel in a given set of solutions, we set the cardinality of $N_c(t, \mathcal{S})$ in relation to the cardinality of $I^R(\mathcal{S})$.

Definition 2.3. *The* ***t-kernel relevance*** *in a set of solutions $\mathcal{S}$, $KER(t, \mathcal{S})$, is the cardinality of the t-kernel, $N_c(t, \mathcal{S})$, divided by the cardinality of the set of relevant candidates, $I^R(\mathcal{S})$, such that*

$$KER(t, \mathcal{S}) := \frac{|N_c(t, \mathcal{S})|}{|I^R(\mathcal{S})|}. \tag{2.3}$$

$KER(t, \mathcal{S})$ is restricted to $[0, 1]$ and by definition $KER(1/|\mathcal{S}|, \mathcal{S}) = 1$. Notice that $KER(1, S)$ equals the Jaccard index, a well-known measure to evaluate the similarity of two sets. It compares the cardinality of their intersection with the cardinality of their union, such that for two finite sets A and B,

$$Jacc(A, B) = \frac{A \cap B}{A \cup B}. \tag{2.4}$$

$KER(t, \mathcal{S})$ can be considered a refinement of that index.

Due to its dependence on the parameter t, the t-kernel relevance does not make a concise definition of the importance of a subset of facilities in a particular set of solutions. Therefore, we present the kernel persistence as a measure that combines $KER(t, \mathcal{S})$ for all possible values of t.

Definition 2.4. *The* ***kernel persistence*** *in a set of solutions $\mathcal{S}$, $KER(\mathcal{S})$, is the sum of $KER(\tau/|\mathcal{S}|, \mathcal{S})$ for $\tau \in \{1, \ldots, |\mathcal{S}|\}$ divided by the cardinality of $\mathcal{S}$ such that*

$$KER(\mathcal{S}) := \frac{\sum_{\tau=1}^{|\mathcal{S}|} KER(\frac{\tau}{|\mathcal{S}|}, \mathcal{S})}{|\mathcal{S}|}. \tag{2.5}$$

Dividing the sum by the size of $\mathcal{S}$ standardizes $KER(\mathcal{S})$ to $(0, 1]$. If $KER(\mathcal{S}) = |\mathcal{S}|^{-1}$, then the solutions in $\mathcal{S}$ do not share any facility. If $KER(\mathcal{S}) = 1$, all solutions operate with the same set of facilities. Thus, $KER(\mathcal{S})$ measures the stability of the set of operating candidate locations across the different solutions in $\mathcal{S}$.

Another way to approach the notion of persistence of individual facilities is to set the average number of candidate facilities operating in a solution in $\mathcal{S}$ in relation to the size

of $I^R(\mathcal{S})$. In fact, Eq. (2.6) shows that $KER(\mathcal{S})$ does exactly this:

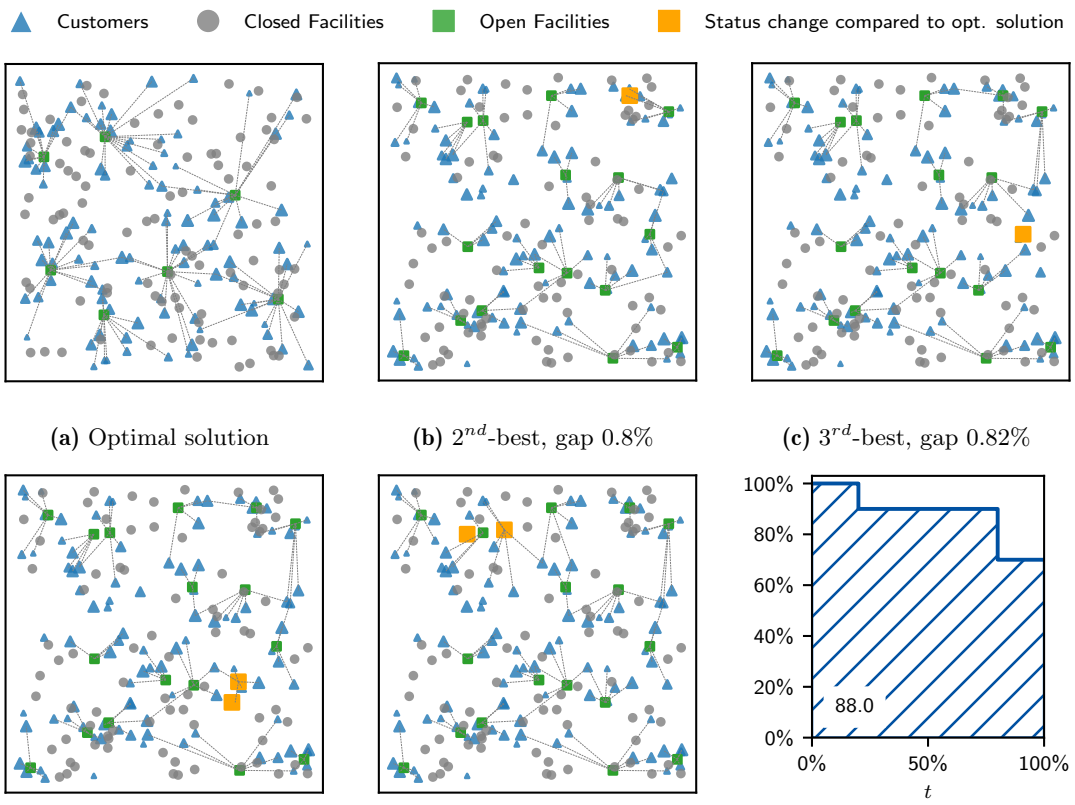
$$\begin{aligned}
KER(\mathcal{S}) =& |\mathcal{S}|^{-1} \cdot \sum\nolimits_{\tau=1}^{|\mathcal{S}|} KER(\frac{\tau}{|\mathcal{S}|}, \mathcal{S}) \\
=& \left(|I^R(\mathcal{S})| \cdot |\mathcal{S}|\right)^{-1} \cdot \sum\nolimits_{\tau=1}^{|\mathcal{S}|} |N_c(t, \mathcal{S})| \\
=& \left(|I^R(\mathcal{S})| \cdot |\mathcal{S}|\right)^{-1} \cdot \sum\nolimits_{\tau=1}^{|\mathcal{S}|} \left| \left\{ i \in I \middle| \frac{\sum_{s\in\mathcal{S}} y_i^s}{|\mathcal{S}|} \geq \frac{\tau}{|\mathcal{S}|} \right\} \right| \\
=& \left(|I^R(\mathcal{S})| \cdot |\mathcal{S}|\right)^{-1} \cdot \sum\nolimits_{\tau=1}^{|\mathcal{S}|} \sum\nolimits_i \mathbb{1} \left(\frac{\sum_{s\in\mathcal{S}} y_i^s}{|\mathcal{S}|} \geq \frac{\tau}{|\mathcal{S}|} \right) \\
=& \left(|I^R(\mathcal{S})| \cdot |\mathcal{S}|\right)^{-1} \cdot \sum\nolimits_{\tau=1}^{|\mathcal{S}|} \sum\nolimits_i \mathbb{1} \left(\sum_{s\in\mathcal{S}} y_i^s \geq \tau \right) \\
=& \left(|I^R(\mathcal{S})| \cdot |\mathcal{S}|\right)^{-1} \cdot \sum\nolimits_i \sum\nolimits_{s\in\mathcal{S}} y_i^s \\
=& |I^R(\mathcal{S})|^{-1} \cdot \frac{\sum_{i,s} y_i^{s^\star}}{|S|}.
\end{aligned} \tag{2.6}$$

Example A 2.1 (Kernel persistence) Consider instances $\mathcal{P}_1$-$\mathcal{P}_4$ from Example A. We generate sets of solutions composed of the k best alternative solutions, further denoted by $\mathcal{S}^k_{k-best}$. Following the approach of Guazzelli and Cunha [2018], we distinguish solutions in this context only based on their location decisions. We iteratively determine the k best alternative solutions by solving a sequence of CFLPs. In each iteration, we exclude the previously found solutions. Thus, let O_k be the set of candidate locations operating in the k-th solution and N_k the set of closed candidate locations in this solution. To eliminate all previously found solutions from the search space, we add the following constraint

$$\sum\nolimits_{i\in O_n} y_i - \sum\nolimits_{i\in N_n} y_i \leq |O_n| - 1 \qquad \forall n \in \{1, \ldots, k-1\}. \tag{2.7}$$

Figure 2.1 displays the optimal and 4 near-optimal solutions to $\mathcal{P}_1$. The gap denotes the relative deterioration of the objective value compared to the optimal solution. Compared to the optimal solution, facilities are either omitted or substituted in the near-optimal solutions. Their demand is partially served by facilities also operating in the optimal solution (see, e.g., Figure 2.1b and Figure 2.1c), or a single facility is replaced by other facilities (Figure 2.1d and Figure 2.1e). The kernel persistence in $\mathcal{S}^5_{k-best}$ is 0.88. 18 facilities are operating in the optimal solution, and 2 additional facilities operate in the solutions depicted in Figure 2.1d and Figure 2.1e, respectively. Thus, the cardinality of the set of relevant facilities $I^R(\mathcal{S}^5_{k-best})$ is 20. Consequently, 14 out of 20 facilities (70%) operate in all 5 ($t = 100\%$) solutions. 18 (90%) operate in at least 4 out of 5 solutions, and 2 facilities (10%) operate in only a single solution.

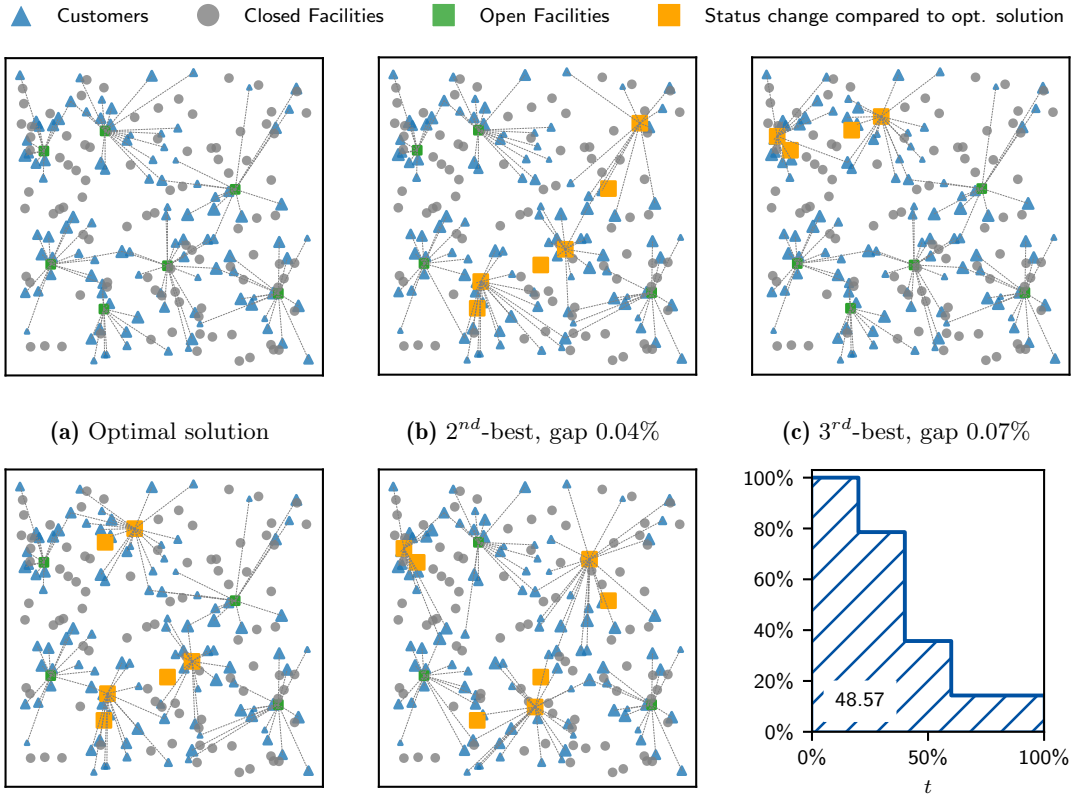
This changes when looking at instance $\mathcal{P}_2$ as depicted in Figure 2.2. The set of facilities operating in the different solutions differs significantly, and the kernel persistence is only 0.49. However, compared to $\mathcal{P}_1$, fewer facilities operate in any optimal or near-optimal solution to $\mathcal{P}_2$. Given that both instances have the same number of candidate locations, this makes a more frequent operation of individual facilities in the 5 best solutions to $\mathcal{P}_1$



(a) Optimal solution **(b)** 2^{nd}-best, gap 0.8% **(c)** 3^{rd}-best, gap 0.82%

(d) 4^{th}-best, gap 1.08% **(e)** 5^{th}-best, 1.38% **(f)** $KER(t, \mathcal{S}^5_{k-best})$

Figure 2.1.: 5 best alternative solutions and development of $KER(t,\ \mathcal{S}^5_{k-best})$ ($\mathcal{P}_1$, Ex. A)



(a) Optimal solution **(b)** 2^{nd}-best, gap 0.04% **(c)** 3^{rd}-best, gap 0.07%

(d) 4^{th}-best, gap 0.13% **(e)** 5^{th}-best, gap 0.21% **(f)** $KER(t, \mathcal{S}^5_{k-best})$

Figure 2.2.: 5 best alternative solutions and development of $KER(t,\ \mathcal{S}^5_{k-best})$ ($\mathcal{P}_2$, Ex. A)

more probable. ▲

Example A 2.1 illustrates a huge difference in the kernel persistence between individual instances. At the same time, it highlights that to interpret $KER(\mathcal{S})$ meaningfully, one must set the number of facilities operating, on average, in a solution in $\mathcal{S}$ in relation to the number of candidates. Let n denote the average number of facilities operating in any solution. When the ratio between n and $|I|$ approaches 1 – thus, a considerable proportion of the facilities operates in any given solution – then $KER(\mathcal{S})$ will also be relatively high. However, this does not necessarily imply that certain facilities are overly persistent in terms of a strategic advantage but might result from a lack of other available candidates.

For example, consider a situation in which out of $|I| = 10$ identical candidates, 9 operate in every solution in $\mathcal{S}$, e.g. because any combination of 9 facilities produces an optimal solution. Suppose $\mathcal{S}$ consists of 2 distinct optimal solutions. In that case, this implies that 8 candidates are operating in both solutions and 2 candidates are operating in at least one solution, which produces a value for $KER(\mathcal{S})$ of 90%. On a scale from 0 to 100%, this is relatively high. However, it has no expressiveness regarding some candidates being more favorable than others. One would expect a similar result if 9 out 10 candidates were chosen randomly for each of the two solutions. Therefore, we need a set-specific reference point to interpret $KER(\mathcal{S})$ meaningfully.

Let $\tilde{\mathcal{S}}$ denote a random counterpart to $\mathcal{S}$. Assume that a solution $\tilde{s}$ results from a Laplace experiment in which n out of $|I|$ facilities are drawn randomly, with each candidate having the same probability of being drawn. Then, assume that $\tilde{\mathcal{S}}$ is the result of a second Laplace experiment which draws $|\tilde{\mathcal{S}}|$ solutions from all possible outcomes of the former experiment. As such, the cardinality of the set $I^R(\mathcal{S})$ is a discrete random variable with finite moments. We can obtain its expectation $\mathbb{E}\left(|I^R(\mathcal{S})|\right)$ and with that the expected value of $KER(\tilde{\mathcal{S}})$. The latter can serve as a reference value to $KER(\mathcal{S})$ as it represents a situation in which no strategic advantage of any candidate exists. To determine the expected value of the cardinality of $I^R(\mathcal{S})$, we can use the formula for the expectation of discrete random variables,

$$\mathbb{E}\left(|I^R(\tilde{\mathcal{S}})|\right) = \sum\nolimits_{r=1}^{|I|} r \cdot P\left(|I^R(\tilde{\mathcal{S}})| = r\right). \tag{2.8}$$

A detailed derivation of the probability of $P\left(|I^R(\tilde{\mathcal{S}})| = r\right)$ can be found in Appendix A, This leads to the following expression for the expectation of $KER(\tilde{\mathcal{S}})$

$$\mathbb{E}\left(KER(\tilde{\mathcal{S}})\right) = \frac{\left\lceil \sum_i \sum_{\tilde{s}\in\tilde{\mathcal{S}}} y_i^{\tilde{s}} \right\rceil}{\left\lceil \mathbb{E}\left(|I^R(\tilde{\mathcal{S}})|\right) \right\rceil}. \tag{2.9}$$

Notice that the assumption of a random Laplace experiment equals assuming complete symmetry between the individual candidate facilities.

Example A 2.2 (Expected vs. observed kernel persistence) Consider again instances $\mathcal{P}_1$-$\mathcal{P}_4$ from Example A and their k best solutions for $k \in \{5, 10, 20\}$. Table 2.2 displays the

observed kernel persistence values and the corresponding expected values for the random counterpart. Due to the higher per facility capacities, it is to be expected that, on average, fewer facilities operate in the k best solutions to $\mathcal{P}_2$ and $\mathcal{P}_4$ compared to $\mathcal{P}_1$ and $\mathcal{P}_3$, resulting in lower values for the expected kernel persistence in the random counterparts. However, one can see that the differences in the observed values significantly exceed the differences in the corresponding expectations. This indicates that the increased persistence results from these facilities having a strategic advantage over other candidates rather than from a simple lack of alternative candidates.

	$KER\left(\mathcal{S}^{5}_{k-best}\right)$	$\mathbb{E}\left(.\right)$	$KER\left(\mathcal{S}^{10}_{k-best}\right)$	$\mathbb{E}\left(.\right)$	$KER\left(\mathcal{S}^{20}_{k-best}\right)$	$\mathbb{E}\left(.\right)$
$\mathcal{P}_1$	0.88	0.29	0.83	0.21	0.73	0.18
$\mathcal{P}_2$	0.49	0.23	0.36	0.14	0.29	0.09
$\mathcal{P}_3$	0.95	0.29	0.94	0.21	0.93	0.19
$\mathcal{P}_4$	0.57	0.23	0.43	0.13	0.35	0.08

Table 2.2.: Persistent kernels in k best alternative solutions ($\mathcal{P}_1$-$\mathcal{P}_4$, Ex. A)

▲

Example A 2.2 demonstrates that there are instances with a remarkable persistence of individual facilities across all optimal and near-optimal solutions. These form a persistent core of favorable facilities. However, while Guazzelli and Cunha [2018] and Fadda et al. [2021] conclude that, in general, location-allocation problems exhibit such a core, Example A 2.1 and Example A 2.2 demonstrate that this is not true for all instances. In particular, no significantly persistent core could be identified for instances $\mathcal{P}_2$ and $\mathcal{P}_4$. Thereby, these two instances do not exhibit highly artificial properties, e.g., symmetries, but were generated with the same randomized generation procedure as instances $\mathcal{P}_1$ and $\mathcal{P}_3$ for which these cores can be observed. These instances only differ in terms of the capacity of individual facilities. We evaluate the prevalence of subsets of core facilities in more instances from literature in the following.

2.2.2. Experimental validation: persistent subsets of core facilities

We examine the kernel persistence in sets of k best alternative solutions to instances from the data sets presented in Section 1.3 for $k \in \{5, 10, 20\}$. Figure 2.3a displays the distribution of kernel persistence values, Figure 2.3b shows the distribution of expected values of the corresponding random counterparts, and Figure 2.3c shows the distribution of their respective differences.

With increasing k, the kernel persistence values decrease throughout all instances (Figure 2.3a). This is expected, given that the k best solutions are generated by forcing the set of optimal facilities to differ between alternative solutions. However, in 4 out of 5 considered data sets from literature, the kernel persistence is remarkably high. For $k = 5$ and data sets BAR-1991, KLO-2007, and ORLIB, the median exceeds 0.8. This explains the impression that persistent subsets of core facilities play a significant role in many, or even all, solutions to discrete location problems.

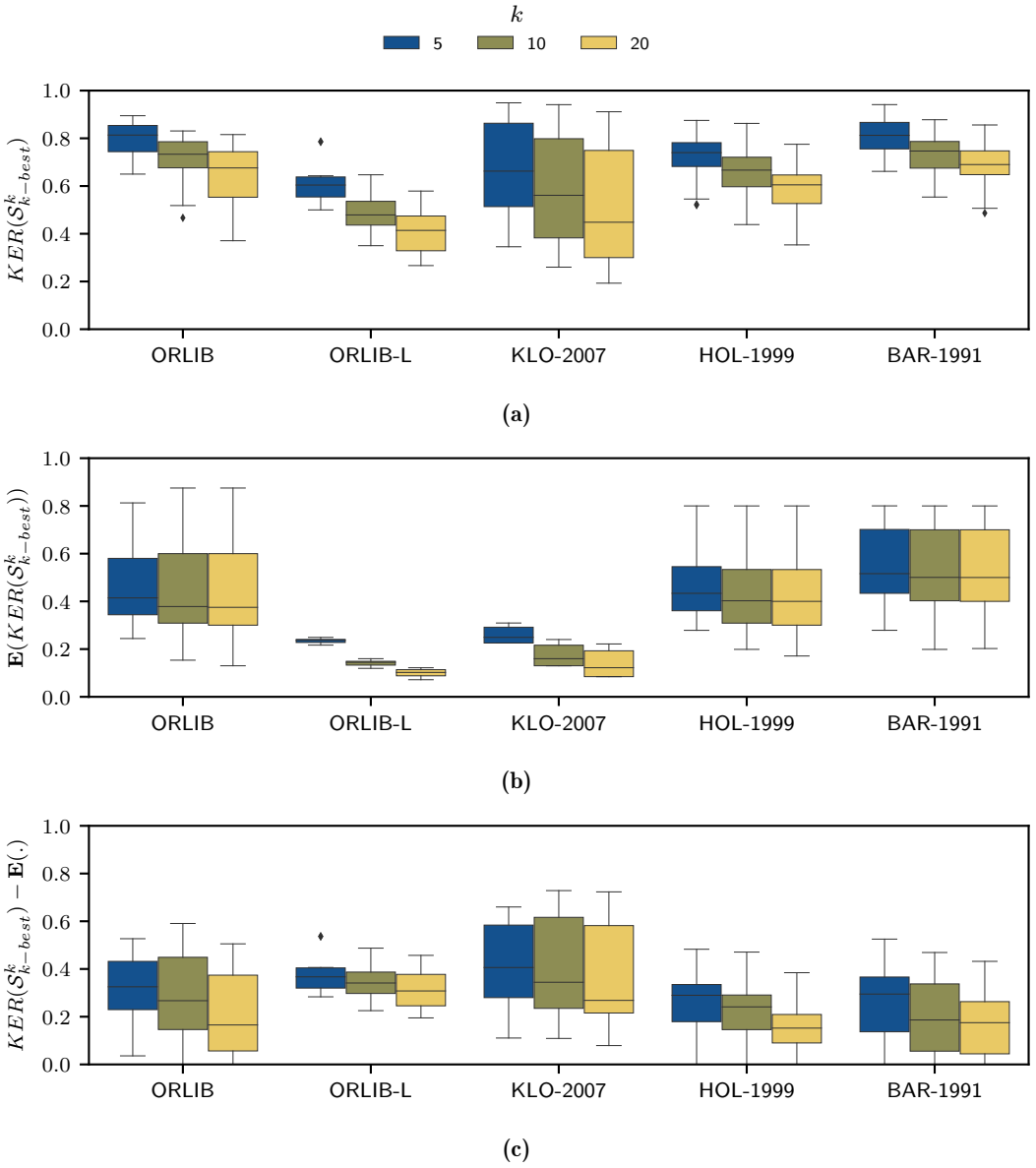


Figure 2.3.: Distribution of $KER(\mathcal{S}^k_{k-best})$, $\mathbb{E}(KER(\mathcal{S}^k_{k-best}))$, and their difference for $k \in \{5, 10, 20\}$

However, Figure 2.3b illustrates that, in particular, for instances from the ORLIB and BAR-1991 data sets, the high kernel persistence values are a direct consequence of the relatively small number of candidates compared to the number of facilities operating in the optimal solution. The only data set for which parts of the instances significantly exceed their expected kernel persistence values are instances from KLO-2007. Meanwhile, the differences between individual instances from this data set are significant. For sets of 10 best alternative solution, the smallest observed difference between the observed and expected kernel persistence of an instance from KLO-2007 is 0.11, while the largest difference for an instance from this set is 0.73. Since for all instances in this data set candidates and customers are randomly distributed, this contradicts the hypothesis that the persistence of core facilities decreases with a random spatial pattern.

In conclusion, some but not all considered CFLP instances exhibit subsets of favorable core facilities that persist throughout optimal or near-optimal solutions. Instances that exhibit such core facilities often have a limited number of alternative candidates to serve all demands profitably. A relationship between the persistence of such facilities and the randomness of the spatial distribution of candidates and customers cannot be observed.

Up to this point, the analysis of persistent decision patterns in well-performing solutions was restricted to location decisions and primarily centered around counting the frequency with which individual facilities operate. In the following, we will perform an exploratory analysis to examine the role of individual facilities in an optimal solution to the CFLP and evaluate which other stable patterns can be identified across optimal and near-optimal solutions.

2.3. Persistent service regions

In a particular problem instance, some decisions are more favorable than others. They respond to some patterns in the data particularly well, leading them to be part of many or even all optimal and near-optimal solutions. Up to this point, there was an implied notion that for discrete location problems, favorable decisions constitute themselves as subsets of facilities. The stronger the structures in the spatial pattern of the distribution of candidates and customers, the more persistent are certain favorable facilities. The results of the previous section have shown that persistent subsets of core facilities can also be observed in instances in which candidates and customers are uniformly distributed, and thus, the spatial pattern is purely random. Furthermore, instances exist that do not exhibit a particularly persistent subset of favorable core locations. The question arises whether there are other favorable decision patterns that respond particularly well to the requirements of these instances. Understanding these patterns provides contextual information for a decision-maker and allows for specific tailoring of solution algorithms.

To the best of our knowledge, no structured approach to identifying such patterns exists at this point. An optimal solution to the CFLP consists of a binary vector of location decisions $y^\star$ and a matrix of continuous allocation decisions $x^\star$. Without further information, differences in the importance of individual facilities or interdependence between certain

location decisions cannot be determined. In what follows, we perform an experiment on a particular set of solutions to a given instance that allows us to evaluate the role and interdependence of individual location decisions. Let

$$I^\star := \{i \in I \mid y_i^\star = 1\} \tag{2.10}$$

denote the set of facilities operating in the optimal solution. For each $i \in I^\star$, we resolve the model with the additional constraint that facility i must not operate ($y_i = 0$). The effect is the same as removing i from the set of candidates or setting the capacity of facility i to 0. We denote the resulting solution as $s_{I\setminus i^\star}$ and derive the following set of solutions for further analysis

$$\mathcal{S}_{I\setminus i} := \left\{ s^\star_{I\setminus i} = \left(z^\star_{I\setminus i}, y^\star_{I\setminus i}, x^\star_{I\setminus i} \right),\ i \in I^\star \right\}. \tag{2.11}$$

Simply put, we explore the role of individual optimal facilities and the interplay among them by analyzing what happens when they are not there.

In Subsection 2.3.1, we use $\mathcal{S}_{I\setminus i}$ to derive measures that characterize the role of individual facilities in an optimal solution. We relate these measures to the persistence of individual facilities in well-performing solutions. Based on the facilities operating in the optimal solutions to $\mathcal{P}_1$-$\mathcal{P}_4$ from Example A, we then distinguish between compositions of facilities in optimal solutions. Furthermore, the experiments on Example A demonstrate that facilities in the optimal solutions to different instances exhibit varying degrees of interdependence. While, per se, the fact that certain facilities are optimal only in combination is an expression of the combinatorial part of the mixed-binary CFLP, the varying level of interdependence between the optimal solutions to different instances has not been examined at this point. In Subsection 2.3.2, we derive concise measures to describe the interdependence of facilities operating in a particular optimal solution. In particular, we quantify the level of interdependence of the facilities in an optimal solution and the separability of the set of optimal facilities into subsets of stronger and weaker coherence. We show that the level of interdependence of individual decisions is closely related to the existence of persistent core facilities. In Subsection 2.3.3, we relate the interdependence of optimal facilities to their joint service of larger subsets of customers. We derive the concept of persistent service regions as a persistent decision pattern underlying well-performing solutions to a particular instance. In Subsection 2.3.4, we validate the observations from the previous subsections on the instances from Section 1.3.

2.3.1. Different roles of individual facilities

In the following subsection, we use $\mathcal{S}_{I\setminus i}$ to derive measures that characterize the role of individual facilities in the optimal solution in the objective value and the decision space. We relate these measures to the persistence of individual facilities in well-performing solutions and assess whether different types of optimal facilities can be characterized. We first extend the concept of kernel persistence to the persistence of individual facilities (Subsection 2.3.1.1). Then, we assess the relationship between the persistence of individual

facilities and their contribution to the optimal objective value (Subsection 2.3.1.2). The latter assumes that upon removing a facility from the optimal solution, the opening status of other optimal facilities can be modified. We present a measure that allows quantifying the extent of this modification, the level of change induced by removing a facility in the decision space (Subsection 2.3.1.3). We then present a measure that assesses the necessity of modifying the decision vector to maintain an optimal strategy and, again, relate it to the persistence level (Subsection 2.3.1.3). We perform our analysis on the optimal solutions to instances $\mathcal{P}_1$-$\mathcal{P}_4$ from Example A. We validate our main conclusions on instances from data sets described in Section 1.3 in Subsection 2.3.4.

2.3.1.1. Persistence of individual facilities

We extend the concept of measuring the persistence of subsets of facilities to measuring the persistence of individual facilities in a particular set of solutions $\mathcal{S}$.

Definition 2.5. *The **persistence** of a facility i in a set of solutions $\mathcal{S}$, $C(i, \mathcal{S})$, equals the proportion of the solutions in $\mathcal{S}$ in which facility i operates such that*

$$C(i, \mathcal{S}) := \frac{\sum_{\tau=1}^{|\mathcal{S}|} \mathbb{1}_{i \in N_c(\frac{\tau}{|\mathcal{S}|}, \mathcal{S})}}{|\mathcal{S}|} = \frac{\sum_s y_i^s}{|\mathcal{S}|}. \tag{2.12}$$

Another interpretation of $C(i, \mathcal{S})$ is that it equals the highest level t such that facility i is an element of the associated t-kernel. Candidates that are not part of any solution have a persistence of $C(i, \mathcal{S}) = 0$. Candidate locations that are part of every solution have a persistence of $C(i, \mathcal{S}) = 1$.

2.3.1.2. Pseudo-reduced costs

Example A demonstrates that a single optimal facility might be crucial for obtaining a certain objective value. For $\mathcal{P}_1$, re-optimizing the problem after removing a single candidate reduced the optimal objective value by more than 11%. To quantify this effect for arbitrary facilities, we follow the idea of Guazzelli and Cunha [2018] and compute the pseudo-reduced costs of a candidate location i.

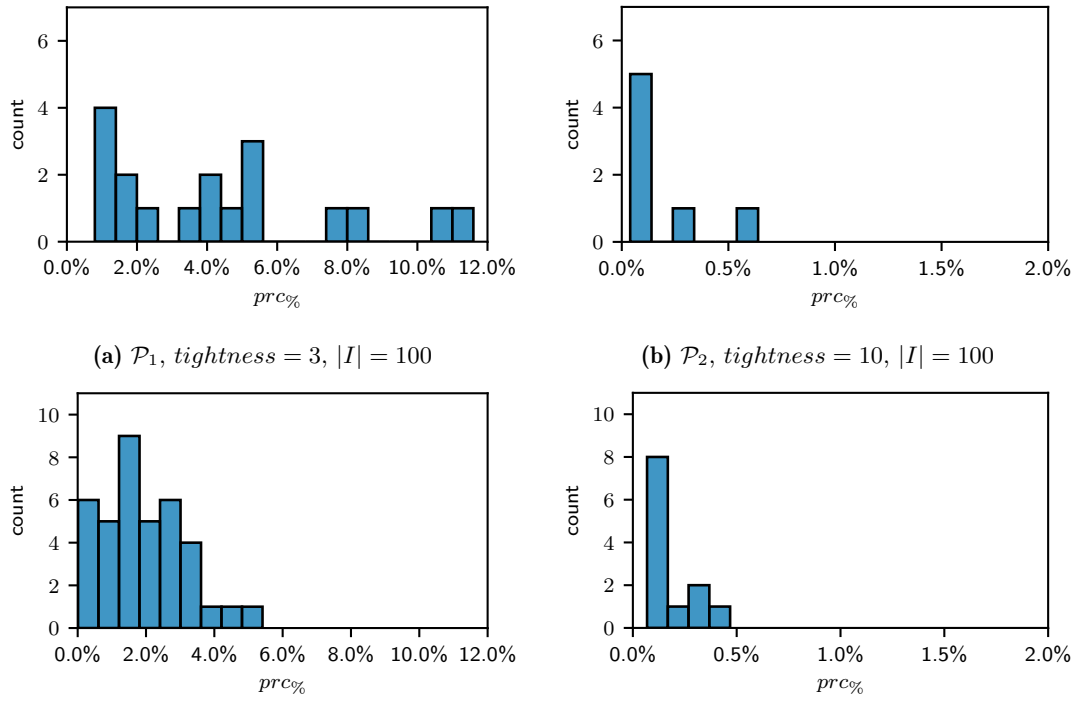
Definition 2.6. *The **pseudo-reduced costs** of a candidate facility i, $prc_{\%}(i)$, denote the relative arithmetic difference between the original optimal objective value, $z^\star$, and the optimal objective value, $z^\star_{I \setminus i}$, to the modified problem instance in which facility i has been removed from the set of candidate locations such that*

$$prc_{\%}(i) := \frac{z^\star - z^\star_{I \setminus i}}{|z^\star|}. \tag{2.13}$$

In other words, $prc_{\%}(i)$ denotes the relative loss from removing facility i from the set of candidates. Scaling this loss to the original objective value allows for comparing the pseudo-reduced costs of individual facilities across different instances. The pseudo-reduced costs of facilities not operating in the optimal solution are 0.

In the following Example A 2.3, we analyze the distribution of the pseudo-reduced costs of the optimal facilities in instances exhibiting different levels of kernel persistence across their well-performing solutions, in particular, across sets of k best alternative solutions. We then relate pseudo-reduced costs of individual facilities to their persistence in this set of solutions.

Example A 2.3 (Pseudo-reduced costs and persistence) Figure 2.4 depicts the distribution of the $prc_{\%}$-values of the optimal facilities in instances $\mathcal{P}_1$-$\mathcal{P}_4$. The $prc_{\%}$-values across all optimal facilities of a particular solution differ significantly between the instances. The following observations are noteworthy.



(a) $\mathcal{P}_1$, $tightness = 3$, $|I| = 100$

(b) $\mathcal{P}_2$, $tightness = 10$, $|I| = 100$

(c) $\mathcal{P}_3$, $tightness = 3$, $|I| = 200$

(d) $\mathcal{P}_4$, $tightness = 10$, $|I| = 200$

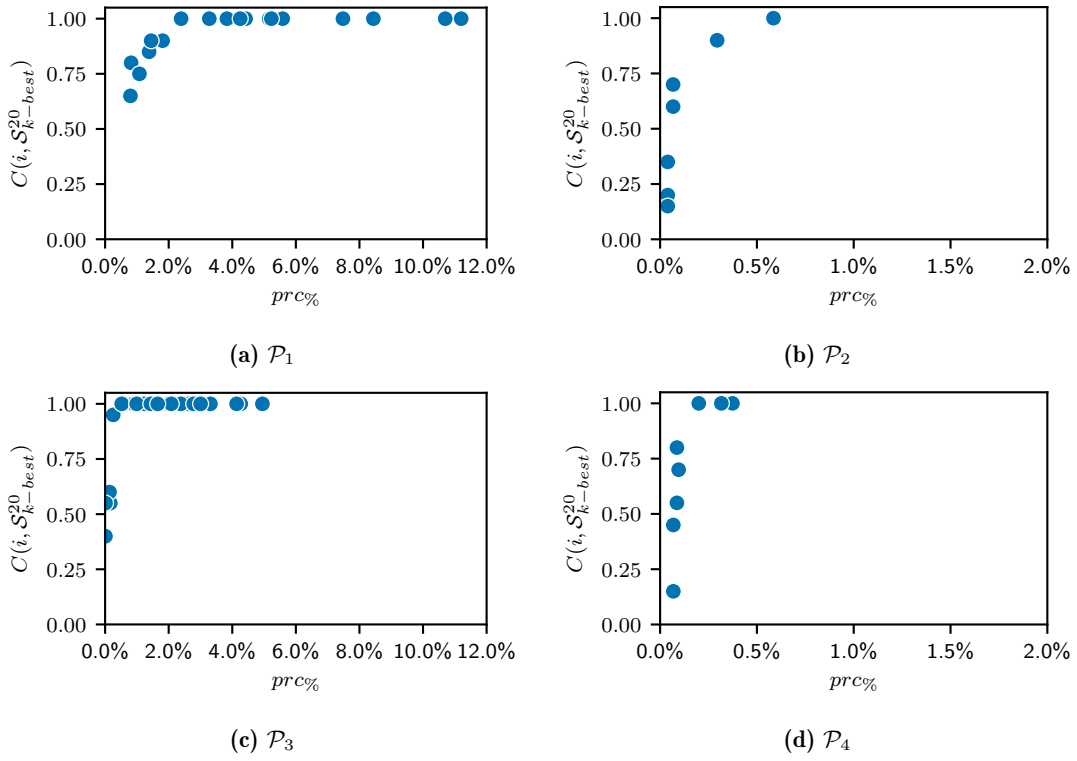
Figure 2.4.: Distribution of $prc_{\%}$-values of the optimal facilities ($\mathcal{P}_1$-$\mathcal{P}_4$, Ex. A)

Between instances with the same tightness but a different number of candidate facilities, the $prc_{\%}$-values decrease with an increasing number of candidates. For example, comparing the distribution of $prc_{\%}$-values between instance $\mathcal{P}_1$ with 100 candidates and $\mathcal{P}_2$ with 200 candidates, one can see that the maximum $prc_{\%}$-value between all facilities decreases from 11.2% to 4.9%. At the same time, the average value across all optimal facilities decreases from 4.4% to 1.9%.

The difference in the $prc_{\%}$-values of instances with the same number of candidates but different tightness ratios is even more significant. In $\mathcal{P}_1$, the $prc_{\%}$-values exceed 5% for 7 out of 18 optimal facilities. In contrast, the maximum $prc_{\%}$-value for any facility in $\mathcal{P}_2$ is 0.6%. In $\mathcal{P}_3$, the $prc_{\%}$-values exceed 2.0% for 17 out of 38 optimal facilities. Meanwhile, the maximum $prc_{\%}$-value observable for facilities in $\mathcal{P}_4$ is 0.4%.

Furthermore, there is a significant difference in the distribution of the $prc_{\%}$-values within the individual instances. Even in instances $\mathcal{P}_1$ and $\mathcal{P}_3$, in which maximum and average

$prc_{\%}$-values are relatively high, several optimal facilities can be removed without significant loss indicated by a $prc_{\%}$-value close to 0.



(a) $\mathcal{P}_1$ **(b)** $\mathcal{P}_2$

(c) $\mathcal{P}_3$ **(d)** $\mathcal{P}_4$

Figure 2.5.: $prc_{\%}$-values versus persistence in 20 best alternative solutions ($\mathcal{P}_1$-$\mathcal{P}_4$, Ex. A)

Figure 2.5 depicts the relationship between the $prc_{\%}$-value of individual facilities and their persistence in the set of 20 best alternative solutions. While in all instances, the persistence of individual facilities increases with increasing pseudo-reduced costs, we observe extremely high persistence values for facilities with very low $prc_{\%}$-values. Particularly in instance $\mathcal{P}_3$, we observe persistence levels of 1.0 for facilities with a $prc_{\%}$-value of only 0.5%. This contradicts the idea that a profound contribution to optimal objective value explains the high persistence of individual facilities across well-performing solutions. ▲

2.3.1.3. Overlap coefficient

We derive the set of solutions $\mathcal{S}_{I\setminus i}$ and, hence, also the optimal objective values $z^{\star}_{I\setminus i}$ from the optimal solutions to instances in which i has been removed from the set of candidates. Thereby, decisions from $s^{\star}$ can be modified in the sense that facilities formerly operating can be either open or closed in $s^{\star}_{I\setminus i}$. Measuring the effect of removing a facility in the decision space is not straightforward as it requires the comparison of the vectors of binary location decisions $y^{\star}$ and $y^{\star}_{I\setminus i}$.

The Hamming distance is a popular way to determine the degree of similarity between two binary vectors. It measures the similarity between two strings of equal length by counting the positions at which their corresponding entries differ [Hamming, 1980]. It is widely used and simple to understand and compute. However, the Hamming distance is too imprecise

to interpret the effect of removing one facility on the remaining facilities in the optimal solution to a location problem, as the following Example D illustrates.

Example D Let $y^{5^\star} = y^{6^\star} = (1\ 1\ 0\ 0\ 0)$ be the optimal vectors of location decisions to two problem instances $\mathcal{P}_5$ and $\mathcal{P}_6$ with 5 candidate facilities each. We analyze the effect of removing optimal facility $i = 1$ on the optimal solution for both instances. Let $y^{5^\star}_{I\setminus 1} = (0\ 1\ 1\ 1\ 0)$ and $y^{6^\star}_{I\setminus 1} = (0\ 0\ 1\ 0\ 0)$ be the optimal vectors of location decisions to instances $\mathcal{P}_{5_{I\setminus 1}}$ and $\mathcal{P}_{6_{I\setminus 1}}$, respectively. The Hamming distance between the optimal solution and the solution without facility 1 is 3 for both instances. For $\mathcal{P}_{5_{I\setminus 1}}$, the entries of the vectors differ at positions 1, 3, and 4. For $\mathcal{P}_{6_{I\setminus 1}}$, the entries differ at positions as well as 1, 2, and 3. Thus, the Hamming distance implies that the effect the removal of the first candidate has on each of the two decision vectors is the same. However, from a decision-making perspective, this removal has very different effects. In the optimal solution to $\mathcal{P}_{5_{I\setminus 1}}$, facility 1 is replaced by two facilities (3 and 4) that were not operating before. Thereby, facility 2 continues to be optimal and is unaffected by the removal of facility 1. Meanwhile, for $\mathcal{P}_{6_{I\setminus 1}}$, the formerly optimal facility 2 is closed, and facilities 1 and 2 are replaced by facility 3. ▲

We present a similarity measure that allows distinguishing between effects with different implications for decision-makers. Let

$$I^\star_{I\setminus i} := \{i' \in I | y^\star_{I\setminus i_{i'}} = 1\} \tag{2.14}$$

be the index set of all optimal facilities operating in $s^\star_{I\setminus i}$. Obviously, $I^\star_{I\setminus i}$ cannot contain facility i by definition. Therefore, the two sets $I^\star$ and $I^\star_{I^\star\setminus i}$ cannot be identical. We distinguish between the following three effects the removal of an optimal facility may have on the remaining location decisions:

1. **No effect**: The removal of the facility does not affect the other location decisions, and the facility is not substituted. Other facilities either fill in for the missing capacity or the customer demand remains unsatisfied. Therefore, it holds that $I^\star_{I\setminus i} = I^\star \setminus i$.
2. **Substitution**: The removed facility is substituted by one or more facilities that did not operate in $s^\star$. Meanwhile, all facilities other than i operating in $s^\star$ continue to operate in $s^\star_{I\setminus i}$. This implies that $I^\star \setminus i \subset I^\star_{I\setminus i}$.
3. **Reconstruction**: The absence of facility i causes structural changes in the solution of different magnitudes. Facilities other than i operating in $s^\star$ no longer operate in $s^\star_{I\setminus i}$ and vice versa. Therefore, we have both $I^\star_{I\setminus i} \nsubseteq I^\star \setminus i$ and $I^\star \setminus i \nsubseteq I^\star_{I\setminus i}$.

In the following, we show that the overlap coefficient [Vijaymeena and Kavitha, 2016], or Szymkiewicz-Spimpson coefficient, a similarity measure from computer science restricted to the interval $[0, 1]$, allows us to distinguish between these three effects.

Definition 2.7. *The **overlap coefficient** of two finite sets A and B, overlap(A, B), is the ratio between the size of their intersection and the size of the smaller of the two sets such*

that

$$overlap(A, B) := \frac{|A \cap B|}{\min(|A|, |B|)}. \tag{2.15}$$

If one of the two sets is a subset of the other, $overlap(A, B)$ equals 1. If the intersection is the empty set and the two sets are entirely different, $overlap(A, B)$ equals 0.

Applying the overlap coefficient to measure the similarity between $I^\star$ and $I^\star_{I\setminus i}$ allows distinguishing between the above effects. For ease of notation, we denote $|I^\star| = n$. When the removal of a facility has no effect, $I^\star_{I\setminus i} = I^\star \setminus i$ and therefore $I^\star_{I\setminus i} \subset I^\star$, the overlap coefficient will be equal to one, as

$$overlap(I^\star, I^\star_{I\setminus i}) = \frac{n-1}{\min\{n, n-1\}} = \frac{n-1}{n-1} = 1. \tag{2.16}$$

If facility i is substituted, $I^\star_{I\setminus i}$ contains all facilities operating in $I^\star$ except facility i plus a positive number of candidate facilities which are not contained in $I^\star$. Therefore, $I^\star_{I\setminus i}$ can be considered the union of two sets such that $I^\star_{I\setminus i} = \{I^\star \setminus i\} \cup \overline{I}$ with $\overline{I} \subseteq I \setminus I^\star$. Thereby, it must hold that $|\overline{I}| = m \geq 1,\ m \in \mathbb{N}$ as otherwise we would be in the first case. It follows that

$$overlap(I^\star, I^\star_{I\setminus i}) = \frac{n-1}{\min(n, (n-1)+m)} = \frac{n-1}{n} < 1. \tag{2.17}$$

Lastly, we have the third effect: reconstruction. A number of facilities other than i operating $s^\star$ are closed in $s^\star_{I\setminus i}$, and an arbitrary number of facilities not operating in $s^\star$ is opened in addition. Again, let $m \in \mathbb{N}$ denote the number of facilities other than i that are operating in $s^\star$ and not operating in $s^\star_{I\setminus i}$. We can rewrite $I^\star_{I\setminus i}$ as the union of two sets, such that $I^\star_{I\setminus i} = I^\star \setminus \{I' \cup i\} \cup \overline{I}$, with $\overline{I}$ defined as before and $I' := \{i' \in I \setminus i | y^\star_{i'} = 1, y^\star_{I\setminus i_{i'}} = 0\}$. Let $|I'| = p \geq 1,\ p \in \mathbb{N}$ as when $p = 0$ we are back in the second effect, substitution. The size of the intersection of $I^\star$ and $I^\star_{I\setminus i}$ therefore reduces to $n - 1 - p$ and we obtain

$$overlap(I^\star, I^\star_{I\setminus i}) = \frac{n-1-p}{\min(n, n-1-p+m)}. \tag{2.18}$$

We want to know whether the resulting value can always be distinguished from the first two effects. Therefore, we will show that it holds that $\frac{n-1-p}{\min(n,n-1-p+m)} < \frac{n-1}{n}$.

We distinguish between three cases:

Case 1: $p = m$ The same amount of facilities are newly operating as are being removed from the solution, such that we obtain

$$overlap(I^\star, I^\star_{I\setminus i}) = \frac{n-1-p}{\min(n, n-1-p+m)} = \frac{n-1-p}{\min(n, n-1)} = \frac{n-1-p}{n-1}. \tag{2.19}$$

Together with $p \geq 1$ we obtain that

$$\frac{n-1-p}{n-1} \leq \frac{n-2}{n-1} < \frac{n-1}{n}. \tag{2.20}$$

Case 2: $m > p$ More facilities are added than removed:

$$overlap(I^\star, I^\star_{I\setminus i}) = \frac{n-1-p}{\underbrace{\min(n, \underbrace{n-1\underbrace{-p+m}_{>0}}_{\geq n})}} = \frac{n-1-p}{n} < \frac{n-1}{n}. \tag{2.21}$$

Case 3: $m < p$ More facilities are removed than facilities added. As $p, m \in \mathbb{N}$, from $m < p$ it follows that $q := p - m \geq 1$, which implies that $p = q + m$ and therefore

$$overlap(I^\star, I^\star_{I\setminus i}) = \frac{n-1-p}{\min(n, n-1-p+m)} = \frac{n-1-q-m}{n-1-q}. \tag{2.22}$$

Now setting $n' = n - 1 - q < n$ implies that

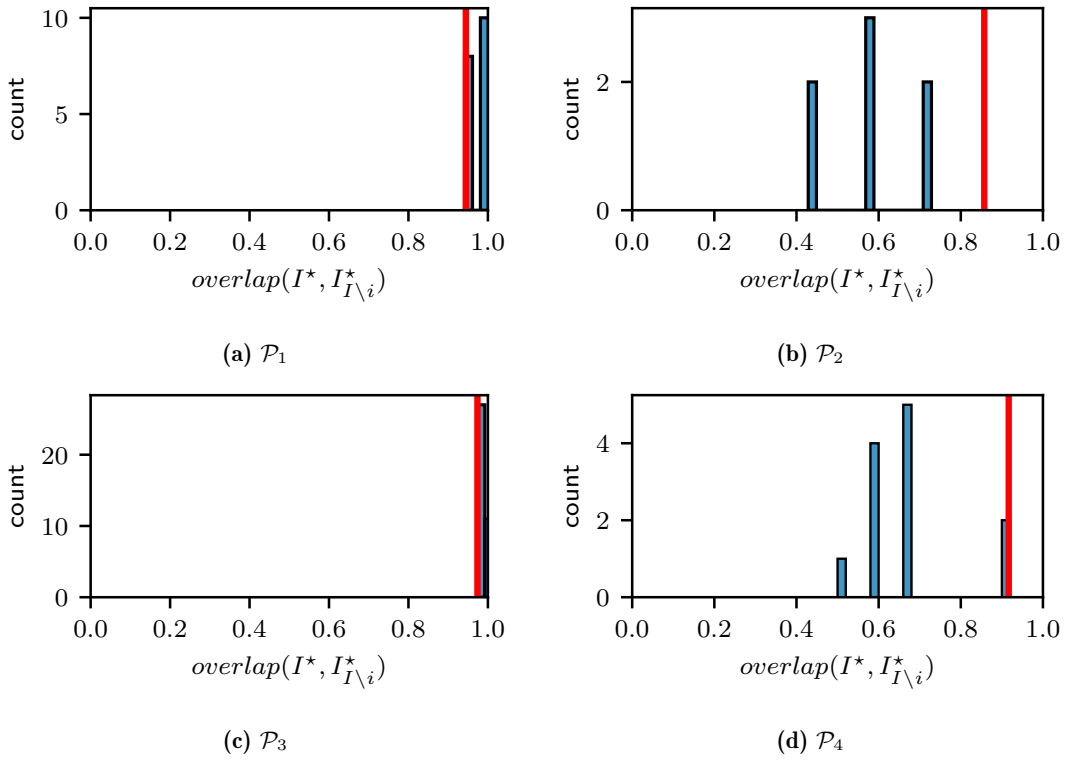
$$\frac{n'-p}{n'} \leq \frac{n'-1}{n'} < \frac{n-1}{n}. \tag{2.23}$$

Together, in the case of reconstruction, the value of the overlap coefficient will be strictly smaller than in the case of substitution. In conclusion, the overlap coefficient is a measure that, together with the cardinality of $I^\star$, allows distinguishing between the three effects that removing a facility may have on the other facilities operating in the optimal solution to a location problem. The distinction is only valid for $n > 1$ since, when only one facility operates in the optimal solution, removing that facility will always produce an overlap coefficient of 0.

Example D continued We obtain $I^\star_5 = I^\star_6 = \{1, 2\}$, $I^\star_{5_{I\setminus 1}} = \{2, 3, 4\}$ and $I^\star_{6_{I\setminus 1}} = \{3\}$ as the sets of optimal facilities for the different solutions. As we have two facilities operating in $s^\star_5$ and $s^\star_6$, we have $n = 2$. For $\mathcal{P}_5$, we obtain $overlap\left(I^\star_5, I^\star_{5_{I\setminus 1}}\right) = \frac{1}{2} = \frac{n-1}{n}$. Thus, we see that the effect of the removal is a substitution of facility 1. Two facilities, 3 and 4, that did not operate before, replaced it. For $\mathcal{P}_6$, we obtain $overlap\left(I^\star_6, I^\star_{6_{I\setminus 1}}\right) = 0 < \frac{n-1}{n}$. Thus, removing facility 1 affected the optimality status of another facility, namely facility 2, resulting in a reconstruction. ▲

In the subsequent Example A 2.4, we examine the overlap coefficients of the optimal facilities in instances $\mathcal{P}_1$-$\mathcal{P}_4$ from Example A and relate them to their persistence values in the 20 best alternative solutions.

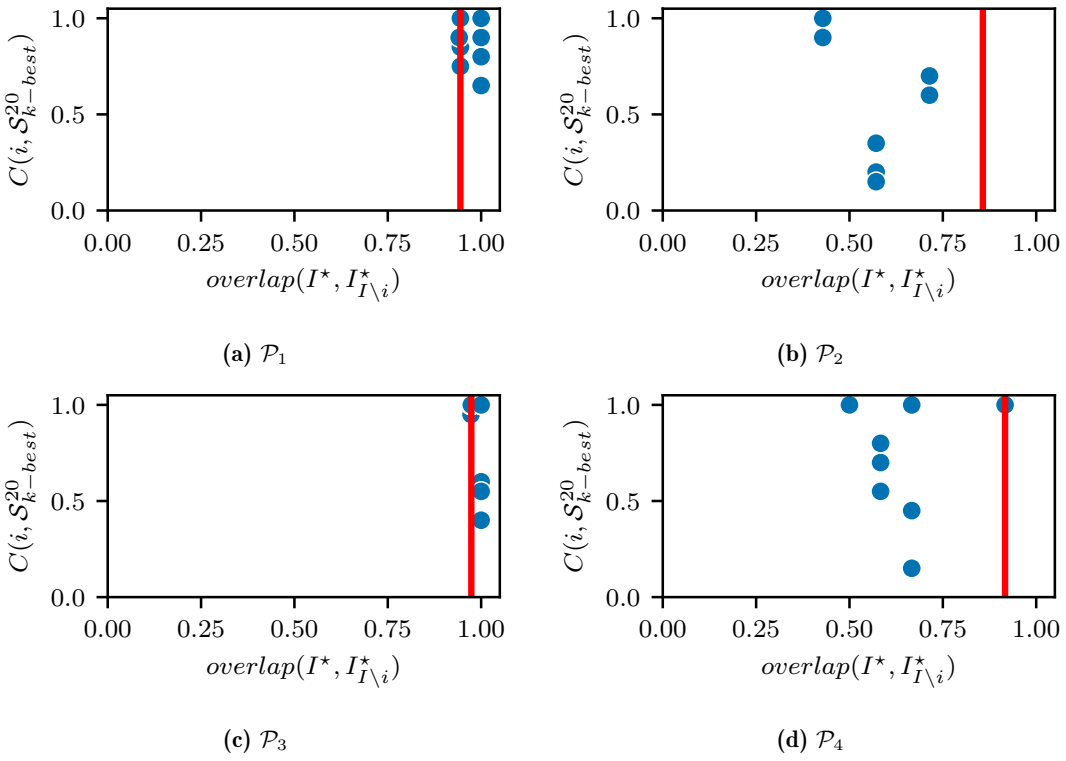
Example A 2.4 (Overlap coefficient and persistence) Figure 2.6 displays histograms for the distribution of the overlap coefficients for facilities operating in the optimal solution to $\mathcal{P}_1$-$\mathcal{P}_4$ from Example A. The red vertical line marks $(n-1)/n$ on the x-axis. All values below that mark indicate a reconstruction of the solutions. We point out the following.



(a) $\mathcal{P}_1$ **(b)** $\mathcal{P}_2$

(c) $\mathcal{P}_3$ **(d)** $\mathcal{P}_4$

Figure 2.6.: Distribution of overlap coefficients of the optimal facilities ($\mathcal{P}_1$-$\mathcal{P}_4$, Ex. A))



(a) $\mathcal{P}_1$ **(b)** $\mathcal{P}_2$

(c) $\mathcal{P}_3$ **(d)** $\mathcal{P}_4$

Figure 2.7.: Overlap coefficients versus persistence in 20 best alternative solutions ($\mathcal{P}_1$-$\mathcal{P}_4$, Ex. A)

There is a clear distinction between the effect that the removal of individual candidates has on the optimal location decisions in instances $\mathcal{P}_1$ and $\mathcal{P}_3$ compared to instances $\mathcal{P}_2$ and $\mathcal{P}_4$. While for the first, the removal of individual facilities had either no effect or resulted in their substitution, the removal of the majority of optimal facilities in the latter resulted in a reconstruction of the solution. Thus, while in $\mathcal{P}_1$ and $\mathcal{P}_3$ the removal of individual facilities had a more substantial effect on the objective value, in $\mathcal{P}_2$ and $\mathcal{P}_4$ the impact is more significant in the decision space.

Figure 2.7 sets the overlap coefficients in relation to the persistence of individual facilities in the 20 best alternative solutions. No systematic relationship between these two measures is visible. ▲

We characterize two different types of instances regarding the composition of facilities operating in their optimal solutions. In the first, individual facilities form persistent cores of favorable facilities operating throughout all well-performing solutions. Several of these facilities contribute significantly to the optimal objective value. In the decision space, the effect of their removal remains local in the sense the customers are either redistributed to other optimal facilities or new facilities that take on these customers are opened. In the second type, facilities are less persistent. Removing individual candidates leads to a reconstruction of a significant part of the optimal solution. Meanwhile, the effect of the removal on the objective value is marginal. This inevitably raises the question of the degree to which the objective value would deteriorate if such a reconstruction was not possible.

2.3.1.4. Damage potential

In the following, we look at the loss in objective value that results from removing a facility i when the rest of the optimal facilities $i' \in I^\star$, $i \neq i'$, must remain open. Thus, we look at the loss induced when one restricts the structural changes in the decision vector y to changes of types "no effect" and "substitution". To obtain this loss, for each facility i, we resolve the problem with the additional constraints that it must not operate in the optimal solution ($y_i = 0$) while all other facilities operating in the optimal solution must operate ($y_{i'} = 1,\ \forall i' \in I^\star \setminus i$). The resulting optimal solution will be denoted as $s^\star_{I^\star \setminus i}$ with an optimal objective value $z^\star_{I^\star \setminus i}$.

Definition 2.8. *The **damage potential** of a candidate facility i, $dp_\%(i)$, denotes the relative arithmetic difference between the optimal objective value obtained with the modified problem instance in which facility i has been removed from the set of candidate locations, $z^\star_{I \setminus i}$, and the optimal objective value obtained when in addition to removing i, all other facilities operating in the optimal solution to the original instance must operate, $z^\star_{I^\star \setminus i}$, such that*

$$dp_\%(i) := \frac{z^\star_{I \setminus i} - z^\star_{I^\star \setminus i}}{|z^\star|}. \tag{2.24}$$

The damage potential captures the additional loss incurred by the inability to adapt the solution upon removing candidate i. It separates this loss from the loss resulting purely

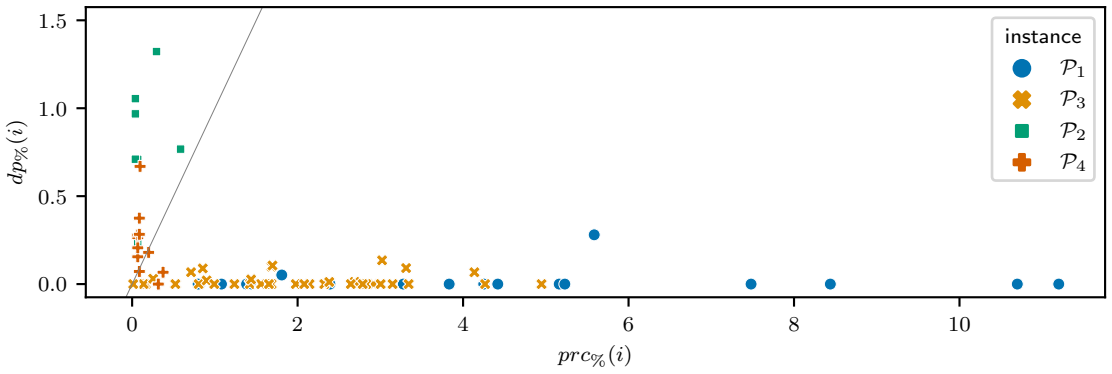


Figure 2.8.: $prc_{\%}$-values versus $dp_{\%}$-values ($\mathcal{P}_1$-$\mathcal{P}_4$, Ex. A)

from the absence of facility i, which is captured by $prc_{\%}(i)$. As $s^{\star}_{I^{\star}\setminus i}$ is obtained from a problem that is even more restricted as $s^{\star}_{I\setminus i}$, it holds that $z^{\star}_{I^{\star}\setminus i} \leq z^{\star}_{I\setminus i}$. To facilitate comparison between instances and between $prc_{\%}(i)$ and $dp_{\%}(i)$, the difference is again scaled by the optimal objective value of the original problem. Once again, we evaluate the damage potential on instances from Example A.

Example A 2.5 (Damage potential) Figure 2.8 displays the $dp_{\%}$-values plotted against the $prc_{\%}$-values of the individual facilities operating in the optimal solutions to instances $\mathcal{P}_1$-$\mathcal{P}_4$. In instances $\mathcal{P}_1$ and $\mathcal{P}_3$, most facilities are omitted or substituted, and the associated damage potential is 0. For those facilities whose removal induces a reconstruction, the damage potential is relatively low and significantly lower than the associated $prc_{\%}$-value. The opposite is true for facilities in instances $\mathcal{P}_2$ and $\mathcal{P}_4$ because their $dp_{\%}$-values are significantly larger than the $prc_{\%}$-values and reach up to 1.3%. In fact, for $\mathcal{P}_2$ and $\mathcal{P}_4$ the $dp_{\%}$-value of individual facilities exceeds their $prc_{\%}$-value in 7 out of 7 facilities for $\mathcal{P}_2$ and 7 out of 12 facilities in $\mathcal{P}_4$, respectively. Meanwhile, for $\mathcal{P}_1$ and $\mathcal{P}_3$ the $prc_{\%}$-values exceed the associated $dp_{\%}$-values for all facilities. In conclusion, in those instances where removing certain facilities makes it necessary to modify the remaining decisions, reconstruction is necessary to preserve a near-optimal objective value. ▲

The presented measures allow for the detection of differences in the role of individual facilities operating in the optimal solutions to different instances. In particular, the persistence of facilities in well-performing solutions cannot be attributed to these facilities' high contribution to the optimal objective value. Instead, we observe that in those instances where the optimal solution exhibits a persistent subset of favorable core facilities, regaining optimality upon removing an individual facility can be achieved locally by redistributing the customers previously served by this facility to new or existing ones. This is different for optimal solutions in which facilities exhibit low persistence values. Once a facility is removed, the vector of optimal location decisions changes substantially to preserve optimality, and a larger subset of customers is reallocated. This implies that in instances with low kernel persistence, some facilities are optimal only in combination. Although this is an expected property of solutions to mixed-binary problems, we examine these relationships more closely in the context of discrete location problems in the following.

2.3.2. Independent and interdependent facilities

The removal of individual facilities has different effects on the optimality status of other facilities. In particular, it may happen that a subset of optimal facilities, other than the one being removed, is no longer part of the optimal solution to the modified problem instance. One could say that these facilities depend on the removed facility. The overlap coefficient allows observing this effect when directly comparing a particular solution $s^\star_{I\setminus i}$ to $s^\star$. In the following, we formally describe these dependence relationships and introduce measures that allow us to assess the level of interdependence of the location decisions in the optimal solution of a problem instance as a whole. In the previous subsection, we evaluated the role of individual optimal facilities by pairwise comparison of $s^\star_{I\setminus i}$ and $s^\star$. In the following, we consider the entire information on the dependence relationships contained in $\mathcal{S}_{I\setminus i}$.

Let

$$Dep(i) := \{i' \in I | i' \in I^\star \setminus \{I^\star_{I\setminus i} \cup i\}\} \tag{2.25}$$

denote the set of optimal facilities whose optimality status depends on facility i in the sense that when i is removed from the set of candidates, the facility turns from open to closed in the corresponding optimal solution. From $\mathcal{S}_{I\setminus i}$ we can derive $|I^\star|$ sets $Dep(i)$. We refer to the union of the sets $Dep(i)$ as Dep. It holds that $Dep \subseteq I^\star$. Furthermore, let

$$Rep(i) := \left\{i' \in I | i' \in I^\star_{I\setminus i} \setminus I^\star\right\} \tag{2.26}$$

denote the set of newly operating facilities when i is removed from the set of candidates. We refer to the union of all sets $Rep(i)$ as Rep, the set of replacements.

Definition 2.9. *A facility $i' \in I^\star$ **depends** on a facility $i \in I^\star$ with $i \neq i'$ if $i' \in Dep(i)$.*

Definition 2.10. *Two facilities $i, i' \in I^\star$ with $i \neq i'$ **interdepend** on each other, if $i' \in Dep(i)$ and $i \in Dep(i')$.*

Definition 2.11. *A facility $i \in I^\star$ is **independent** if $i \notin \bigcup_{i' \in I^\star \setminus i} Dep(i')$.*

We derive the dependency graph of an optimal solution to depict these dependence structures.

Definition 2.12. *The **dependency graph** of an optimal solution $s^\star$, $G^{Dep}(s^\star)$, is a directed graph that depicts the dependence relationships between facilities operating in $s^\star$. The set of vertices equals the set of optimal facilities $I^\star$. An arc (i, i') with $i, i' \in I^\star$ and $i \neq i'$ exists, if facility i' depends on facility i.*

The degree of each vertex in $G^{Dep}(s^\star)$ indicates the dependence or independence of the associated facility i. The in-degree equals the number of facilities that the facility depends on. The out-degree equals the number of facilities its removal affects. Independent facilities have an in-degree of 0 but may have an out-degree greater than 0.

The fraction $|E|/|I^\star|$ denotes the average number of ties any optimal facility in i has with any other facility. As each facility i may affect at most all of the $|I^\star|-1$ remaining facilities, we have $|E| \in [0, |I^\star| \cdot (|I^\star| - 1)]$, $|E| \in \mathbb{N}_0$.

Definition 2.13. *The **average dependence density** of an optimal solution $s^\star$, $Dep^{avg}(s^\star)$, quantifies the degree to which the facilities operating in that solution depend on each other. It sets the total number of edges in $G^{Dep}(s^\star)$ in relation to the number of potential edges such that*

$$Dep^{avg}(s^\star) := \frac{|E|}{|I^\star| \cdot (|I^\star| - 1)}, \text{ with } Dep^{avg} \in [0, 1]. \tag{2.27}$$

If $Dep^{avg}(s^\star)$ equals 0, all facilities operating in the solution are independent. If $Dep^{avg}(s^\star)$ equals 1, all pairs of facilities interdepend.

The set of weakly connected components $\mathcal{W}$ in $G^{Dep}(s^\star)$ gives insight into how the set of facilities can be divided into smaller interdependent subsets. Notice that when two facilities $i, i' \in I^\star$ with $i \neq i'$ are in the same weakly connected component, that does not necessarily mean that one depends on the other. Instead, this can also result from the fact that, e.g., both facilities depend on a third facility. Facilities that are neither affected by any other facility nor affecting any other facility will form a weakly connected component of size 1 by themselves.

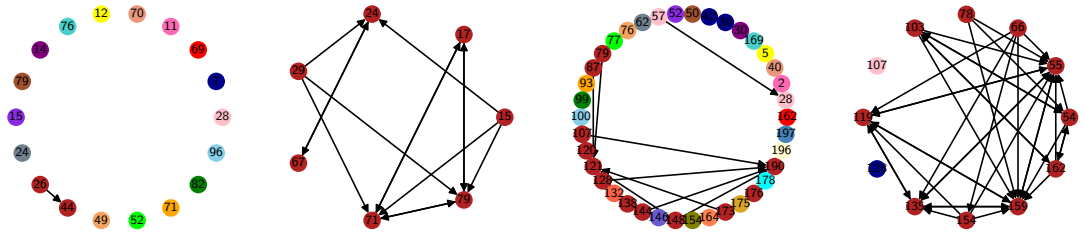
Definition 2.14. *The **strong subset coefficient** of an optimal solution $s^\star$, $\Psi^{Dep}(s^\star)$, measures the degree to which the dependency structures in a graph are separable from one another by setting the number of weakly connected components $\mathcal{W}$ in $G^{Dep}(s^\star)$ in relation to the number of operating facilities such that*

$$\Psi^{Dep}(s^\star) := \rho \left(\frac{|\mathcal{W}|}{|I^\star|} - \frac{1}{|I^\star|} \right), \qquad \text{with } \rho = \frac{|I^\star|}{|I^\star| - 1}. \tag{2.28}$$

When $|I^\star| > 0$, there are at least one and at most $|I^\star|$ weakly connected components. Therefore, to scale $\Psi^{Dep}(s^\star)$ to the interval $[0, 1]$, we must first subtract $\frac{1}{|I^\star|}$ and subsequently multiply it by ρ. The higher $\Psi^{Dep}(s^\star)$, the higher the implied separability of the solution in terms of subsets of facilities that operate independently from one another. If no facility is dependend on any other facility, then there are no connections, and $|\mathcal{W}| = |I^\star|$ which leads to $Dep(s^\star)$ equal to 1.0. This implies that all facilities are independent.

Example A 2.6 (Dependence and separability) Figure 2.9 depicts the dependency graphs of the optimal solutions of instances $\mathcal{P}_1$-$\mathcal{P}_4$ from Example A. The vertices of facilities belonging to the same weakly connected component have the same color.

The differences between the optimal solutions to instances $\mathcal{P}_1$ and $\mathcal{P}_3$ compared to the optimal solutions to instances $\mathcal{P}_2$ and $\mathcal{P}_4$ are apparent. The first two primarily comprise independent facilities or small, weakly connected components. Meanwhile, in $\mathcal{P}_2$, all facilities belong to the same weakly connected component, while for $\mathcal{P}_4$ all but 2 optimal



(a) $\mathcal{P}_1$ **(b)** $\mathcal{P}_2$ **(c)** $\mathcal{P}_3$ **(d)** $\mathcal{P}_4$

Figure 2.9.: $G^{Dep}(s^\star)$ with colored weakly connected components ($\mathcal{P}_1$-$\mathcal{P}_4$, Ex. A)

facilities belong to the same weakly connected component. However, particularly for the optimal solution to $\mathcal{P}_2$, there are some subsets of facilities where all facilities interdepend. In contrast, the dependence relationship is one-sided for other pairs of facilities.

Instance	$\mathcal{P}_1$	$\mathcal{P}_2$	$\mathcal{P}_3$	$\mathcal{P}_4$
$Dep^{avg}(s^\star)$	0.00	0.33	0.01	0.27
$\Psi^{Dep}(s^\star)$	0.94	0.00	0.65	0.18

Table 2.3.: Average dependence density and strong subset coefficient ($\mathcal{P}_1$-$\mathcal{P}_4$, Ex. A)

Table 2.3 shows the average dependence densities. Most facilities in $\mathcal{P}_1$ and $\mathcal{P}_3$ are independent. For $\mathcal{P}_1$, only facility 44 depends on facility 26. This is reflected in the low average dependence densities of 0.00 and 0.01, respectively. For instances $\mathcal{P}_2$ and $\mathcal{P}_4$, the average dependence densities are significantly larger. The table also displays the strong subset coefficients. They adequately represent the high level of separability of the independent facilities of $\mathcal{P}_1$ and $\mathcal{P}_3$ as well as the high interdependence of $\mathcal{P}_2$ and $\mathcal{P}_4$. ▲

We introduce the following to distinguish between subsets of facilities of stronger and weaker coherence.

Definition 2.15. *The* ***interdependency graph*** *of an optimal solution $s^\star$, $G^{Dep'}(s^\star)$, is an undirected graph that depicts the interdependence relationships between facilities operating in the optimal solution. The set of vertices equals the set of optimal facilities $I^\star$. An arc (i, i') with $i, i' \in I^\star$ and $i \neq i'$ exists, if facilities i and i' interdepend.*

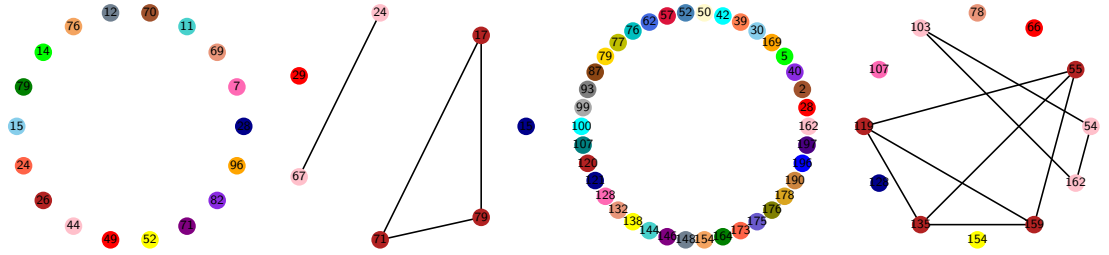
Determining the degree to which the problem is separable into subsets of facilities that mutually depend on one another equals determining the set of maximal cliques $\mathcal{C}$ in $G^{Dep'}(s^\star)$.

Definition 2.16. *The* ***weak subset coefficient*** *of an optimal solution $s^\star$, $\Psi^{Dep'}(s^\star)$, sets the number of cliques in $G^{Dep'}(s^\star)$ in relation to the number of operating facilities*

$$\Psi^{Dep'}(s^\star) := \rho \left(\frac{|\mathcal{C}|}{|I^\star|} - \frac{1}{|I^\star|} \right) \qquad \text{with } \rho = \frac{|I^\star|}{|I^\star| - 1}. \tag{2.29}$$

Again $\Psi^{Dep'}(s^\star)$ is restricted to the $[0, 1]$-interval. When $\Psi^{Dep'}(s^\star)$ is 0, it means that all facilities interdepend on one another. The larger $\Psi^{Dep'}(s^\star)$, the more is the set of facilities separable into subsets of facilities that may depend but not interdepend on one another.

Example A 2.7 (Interdependence and separability) Figure 2.10 displays the interdependency graph $G^{Dep'}(s^\star)$ of $\mathcal{P}_1$-$\mathcal{P}_4$ from Example A. There are no interdependent facilities in the optimal solutions to $\mathcal{P}_1$ and $\mathcal{P}_3$. In contrast, several facilities do not interdepend with any other facilities in the optimal solutions to instances $\mathcal{P}_2$ and $\mathcal{P}_4$. Yet, for both instances one can observe distinct subsets of facilities that mutually interdepend.



(a) $\mathcal{P}_1$ **(b)** $\mathcal{P}_2$ **(c)** $\mathcal{P}_3$ **(d)** $\mathcal{P}_4$

Figure 2.10.: $G^{Dep'}(s^\star)$ with colored maximal cliques ($\mathcal{P}_1$-$\mathcal{P}_4$, Ex. A)

Instance	$\mathcal{P}_1$	$\mathcal{P}_2$	$\mathcal{P}_3$	$\mathcal{P}_4$
$\Psi^{Dep'}(s^\star)$	1.0	0.5	1.0	0.55

Table 2.4.: Weak subset coefficient ($\mathcal{P}_1$-$\mathcal{P}_4$, Ex. A)

The values of $\Psi^{Dep'}(s^\star)$ are displayed in Table 2.4. For instances $\mathcal{P}_1$ and $\mathcal{P}_3$, no two facilities interdepend. Thus, the associated weak subset coefficients are 0. Meanwhile, for $\mathcal{P}_2$ and $\mathcal{P}_4$, values greater or equal to 0.5 suggest that certain subsets in the problem interdepend more strongly than others. ▲

The above raises the question of whether the independence or interdependence of individual facilities is related to the persistence of this facility in well-performing solutions. The following Example A 2.8 relates the persistence of facilities to their interdependence relationships.

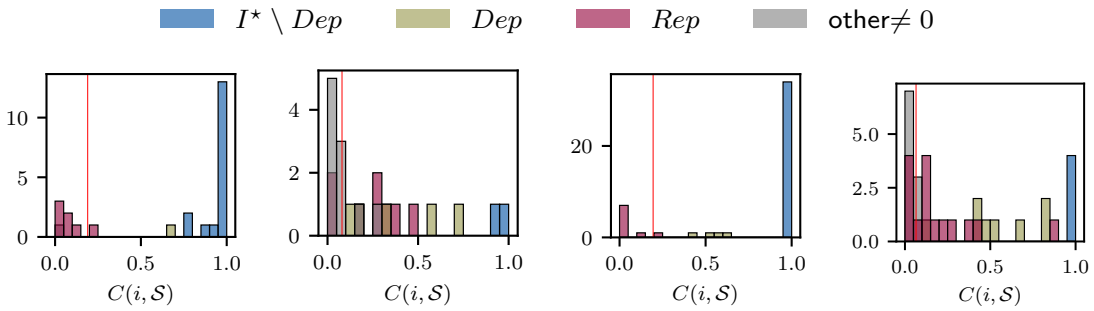
Example A 2.8 (Persistence of independent and interdependent facilities) Table 2.5 displays the kernel persistence in the 20 best alternative solutions and the average dependence density of facilities operating in the optimal solution to instances $\mathcal{P}_1$ to $\mathcal{P}_4$ from Example A. One can clearly see that instances $\mathcal{P}_1$ and $\mathcal{P}_3$ with a higher kernel persistence have a significantly lower dependence density compared to instances $\mathcal{P}_2$ and $\mathcal{P}_4$.

Instance	$\mathcal{P}_1$	$\mathcal{P}_2$	$\mathcal{P}_3$	$\mathcal{P}_4$
$KER(\mathcal{S}^{20}_{k-best}))$	0.73	0.29	0.93	0.35
$Dep^{avg}(s^\star)$	0.00	0.33	0.01	0.27

Table 2.5.: Average dependence density and kernel persistence in the 20 best alternative solutions ($\mathcal{P}_1$-$\mathcal{P}_4$, Ex. A)

Furthermore, Figure 2.11 depicts the distribution of the persistence of individual facilities in the 20 best alternative solutions grouped according to whether these facilities are optimal

and independent ($i \in I^\star \setminus Dep$), optimal and dependent on some other facility ($i \in Dep$), or an element of the set of replacements ($i \in Rep$).



(a) $\mathcal{P}_1$ **(b)** $\mathcal{P}_2$ **(c)** $\mathcal{P}_3$ **(d)** $\mathcal{P}_4$

Figure 2.11.: Distribution of persistence values in 20 best alternative solutions ($\mathcal{P}_1$-$\mathcal{P}_4$, Ex. A)

A clear hierarchy is observable. The independent, optimal facilities have the highest persistence values throughout all instances, with values close to 1.0 throughout. The optimal, dependent facilities have a lower persistence value, often not significantly higher than the persistence value of facilities in the set of replacements. ▲

We conclude a strong relationship exists between facilities being independently optimal and these facilities operating consistently throughout well-performing solutions for a particular problem instance. At the same time, we observe that the facilities operating in interdependent subsets operate less often and are less persistent.

In conclusion, one can distinguish between independent and interdependent facilities operating in an optimal solution. In case a solution contains interdependent facilities, certain subsets of facilities are more strongly connected than others. The previous subsection provided formal definitions and concise measures to quantify these dependence relationships in an optimal solution $s^\star$ based on the set of alternative solutions $\mathcal{S}_{I\setminus i}$. The fact that binary location decisions interdepend to be optimal is not surprising as it is a simple expression of the combinatorial nature of the mixed-binary CFLP. However, what that interdependence implies for individual location decisions in a discrete location problem is not obvious. Furthermore, the ability to identify subsets of stronger and weaker coherence implies an interesting interplay between combinatorial dynamics and spatial patterns underlying the problem instance. We examine this interplay in the following.

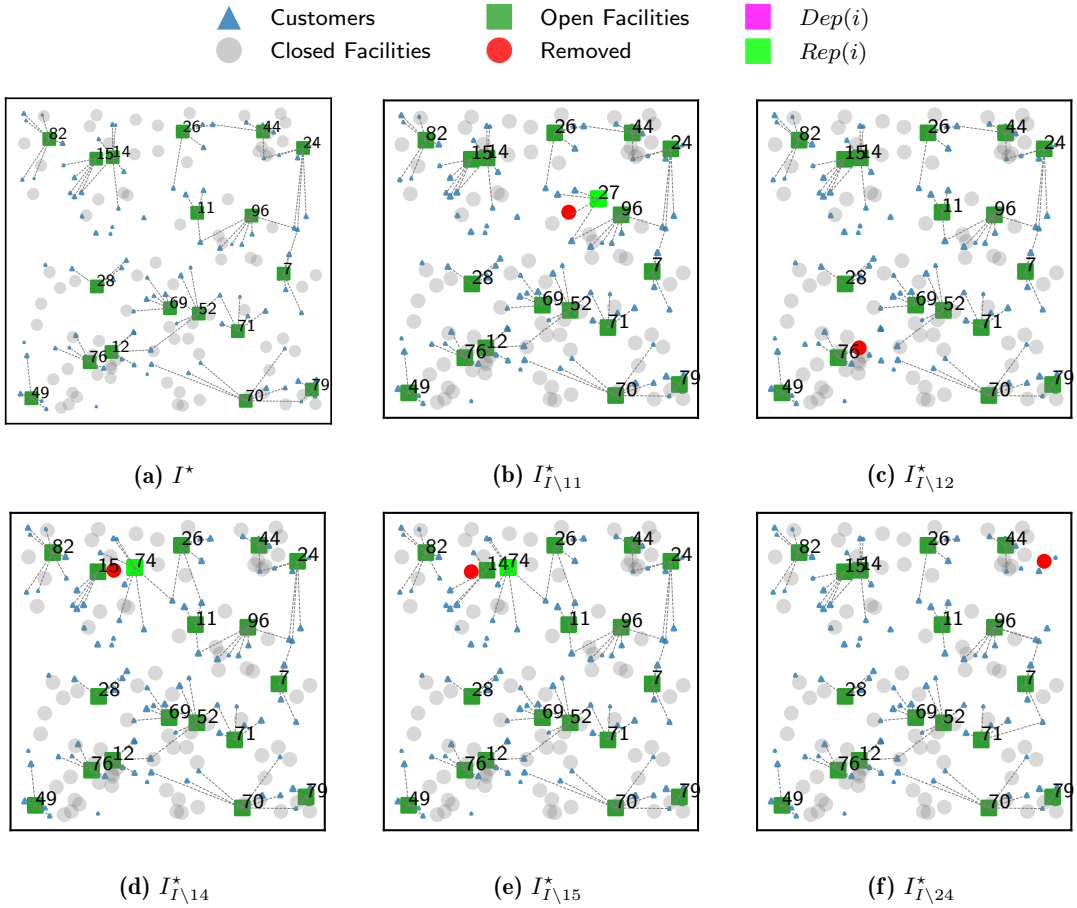
2.3.3. Implicit divisions of the facility-customer space

We refer to the hypothetical embedding that maps candidates and customers on a plane according to their implied spatial relationships derived from the transport-cost matrix as the facility-customer space. The relative positions in this embedding allow us to derive insights into the spatial relationships between candidates and customers solely from the information in the transport-cost matrix without knowing actual geographical positions. It is intuitive to assume that facilities that interdependend are somewhat close to each other in the facility-customer space. In the following, we take a closer look at the effect of

removing individual facilities and the resulting dependence relationships by examining the visual representations of the facility-customer space. In Example A 2.9, we compare the visual representations of the replacements of independent facilities with the replacements of interdependent subsets of facilities based on instances $\mathcal{P}_1$-$\mathcal{P}_4$ from Example A.

Example A 2.9 (Implied spatial relationship between independent and interdependent facilities) Figure 2.12 depicts the optimal solution (Figure 2.12a) and the optimal solutions after the removal of several candidates (Figure 2.12b-Figure 2.12f) to instance $\mathcal{P}_1$ from Example A. We previously established that the optimal solution to $\mathcal{P}_1$ comprises predominantly independent facilities. For example, the removal of facility 12 or facility 24 results in these facilities simply being omitted from the optimal solution. Figure 2.12c and Figure 2.12f illustrate that the customers originally served by these facilities are no longer served upon their removal. Other facilities are substituted by a candidate who then takes over their customers. For example, Figure 2.12b shows how the customers originally served by facility 11 are allocated to facility 27, newly operating after its removal. Another effect can be observed for the removal of facilities 14 and 15 (see Figure 2.12d or Figure 2.12e), respectively. In both cases, facility 74 is opened to replace the removed facility. However, while when replacing facility 14, facility 74 predominantly serves the set of customers previously served by facility 14, the removal of facility 15 leads to facility 14 taking over the customers of facility 15 and facility 74 taking over the customers of facility 14. The customers are redistributed among the set of operating facilities. In summary, Figure 2.12 shows that for $\mathcal{P}_1$ the effects of the removal of individual candidates are restricted to a "local" level, and customers in immediate proximity to that facility are reallocated.

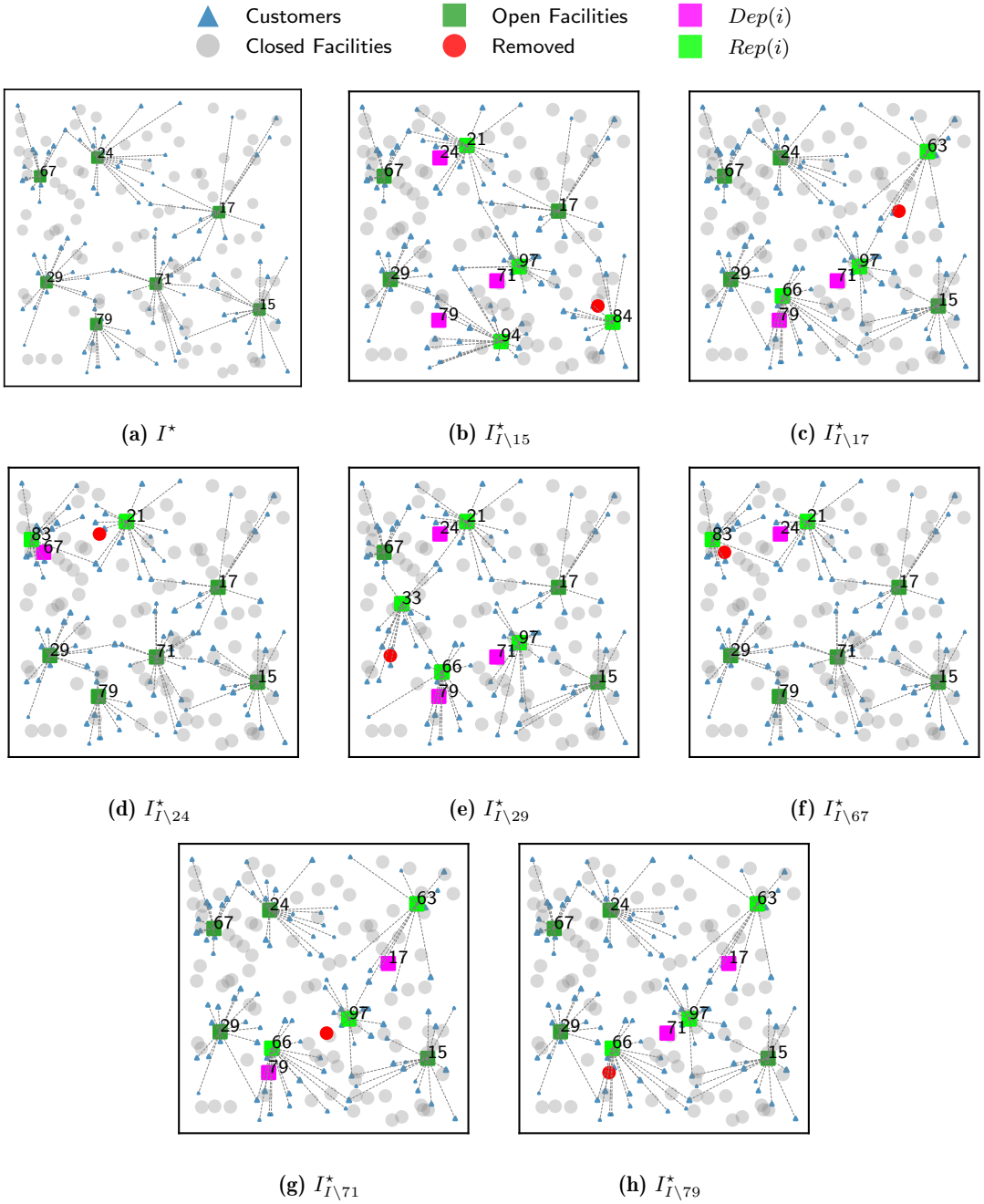
Figure 2.13 and Figure 2.14 depict the effect of the removal of individual candidates in $\mathcal{P}_2$ and $\mathcal{P}_4$, whose optimal solutions comprise mostly interdependent facilities. The figures show that interdependent subsets of facilities are often replaced by the same set $Rep(i)$ of replacement facilities. For example, in $\mathcal{P}_2$ the facilities in the interdependent set $\{17, 71, 79\}$ are always replaced by the set $\{63, 66, 97\}$. Meanwhile, both facilities from $\{24, 67\}$ are replaced by $\{21, 83\}$. As Figure 2.13 shows, in both cases, the replacing subset of facilities serves a similar subset of customers as the facilities operating in the original optimal solutions. However, the allocation of this subset of customers to the individual facilities differs. For example, 87.5% of the customers served by any of the three originally operating facilities 17, 71, or 79 in $s^\star$ are also served by one of the three facilities $\{63, 66, 97\}$ in $s^\star_{I\setminus 17}$, $s^\star_{I\setminus 71}$, or $s^\star_{I\setminus 79}$. However, when we try to map the customers served by either facility 17, 71, or 79 to one of the three replacing facilities individually, that percentage is much lower. For example, facility 63 serves 75% of the same customers as facility 17, and facility 66 serves only 58% of the same customers as facility 71. Thus, while the subset of facilities largely takes over the customers, there is a different allocation of customers within that subset.



(a) $I^\star$ (b) $I^\star_{I\setminus 11}$ (c) $I^\star_{I\setminus 12}$

(d) $I^\star_{I\setminus 14}$ (e) $I^\star_{I\setminus 15}$ (f) $I^\star_{I\setminus 24}$

Figure 2.12.: Opt. solution and opt. solutions after removing candidates ($\mathcal{P}_1$, Ex. A)

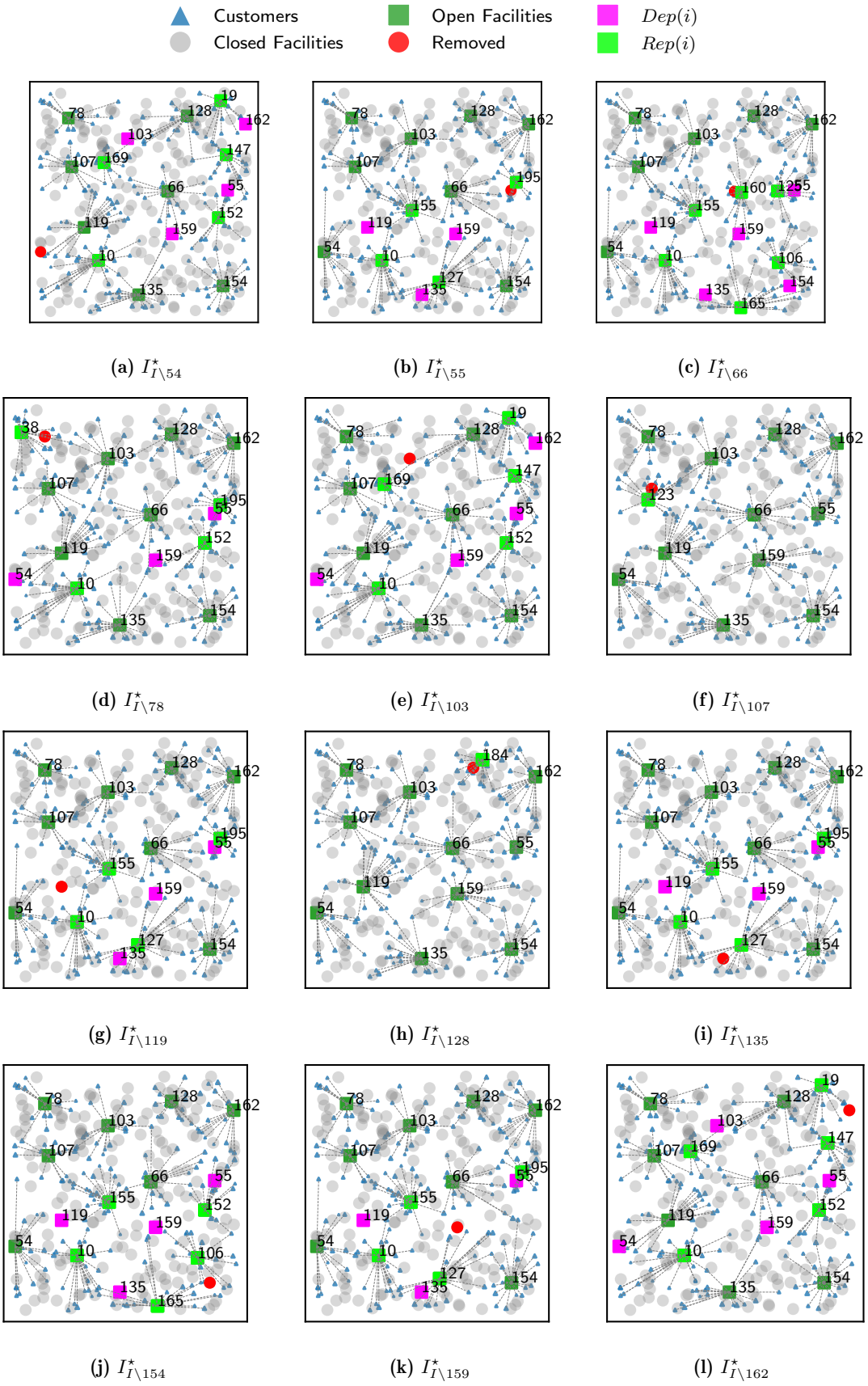


(a) $I^\star$ (b) $I^\star_{I\setminus 15}$ (c) $I^\star_{I\setminus 17}$

(d) $I^\star_{I\setminus 24}$ (e) $I^\star_{I\setminus 29}$ (f) $I^\star_{I\setminus 67}$

(g) $I^\star_{I\setminus 71}$ (h) $I^\star_{I\setminus 79}$

Figure 2.13.: Opt. solution and opt. solutions after removing candidates ($\mathcal{P}_2$, Ex. A)



(a) $I^{\star}_{I\setminus 54}$ (b) $I^{\star}_{I\setminus 55}$ (c) $I^{\star}_{I\setminus 66}$

(d) $I^{\star}_{I\setminus 78}$ (e) $I^{\star}_{I\setminus 103}$ (f) $I^{\star}_{I\setminus 107}$

(g) $I^{\star}_{I\setminus 119}$ (h) $I^{\star}_{I\setminus 128}$ (i) $I^{\star}_{I\setminus 135}$

(j) $I^{\star}_{I\setminus 154}$ (k) $I^{\star}_{I\setminus 159}$ (l) $I^{\star}_{I\setminus 162}$

Figure 2.14.: Opt. solution and opt. solutions after removing candidates ($\mathcal{P}_4$, Ex. A)

Figure 2.14 shows the effect of removing some of the optimal facilities in $\mathcal{P}_4$. Again, we can observe that e.g., for facilities 55, 119, 135, and 159, the reconstruction of the solution is always the same. When one of these four facilities is removed, the other four also no longer operate but are replaced by the set $\{10, 127, 155, 195\}$.

The effect of removing a facility is no longer restricted to a local level, but larger regions of the facility-customer space are affected by the removal. Meanwhile, these regions are not visibly distinct in the form of a cluster or another recognizable pattern in the implied spatial structure. Instead, the affected area can be determined by a larger coherent set of customers that are reallocated compared to the original optimal solution. ▲

The visual representations of the solutions in Example A 2.9 show that interdependent subsets of facilities are located close to each other based on their inferred positions in the facility-customer space. This means that interdependent facilities serve customers in an implied spatial proximity. We conclude that interdependent subsets of facilities offer optimal coverage for sets of customers when operating in combination. If one facility is removed, the interdependent subset is replaced by the next best combination of facilities to cover the entire set of customers. At the same time, independent facilities offer optimal service for smaller sets of customers served optimally by a single facility. This implies a division of the facility-customer space into coherent service regions. To validate this observation on a broader set of instances, we present a measure to quantify the level of customer redistribution in the following.

2.3.3.1. Joint service of customer regions

Every solution implicitly divides the set of customers into subsets that are allocated to the different operating facilities. We can reformulate the observation of the previous subsection as follows. When independent facilities are removed from the set of candidates, this implicit division of the customer space is preserved. In contrast, this implicit division changes when interdependent facilities are removed. It changes particularly strongly within the union of the subsets of customers allocated to one of the facilities in the interdependent set. This perspective allows quantifying the degree to which this effect can be observed in the optimal solution to arbitrary instances.

Given an optimal solution $s^\star$ and the associated set of solutions $\mathcal{S}_{I\setminus i}$, we want to evaluate the level to which the interdependence of facilities operating in the optimal solution is associated with these facilities serving a larger subset of customers jointly, in the sense that, upon the removal of one facility, these customers are redistributed to facilities within the set of replacements. We quantify this observation as follows. First, we evaluate the level to which we can map the implicit division in a solution $s^\star_{I\setminus i}$ to the implicit division in $s^\star$ when considering facilities only individually. To account for the fact that individual facilities in $I^\star$ may be replaced by more than one facility, we search for a one-to-many mapping of the facilities $i' \in I^\star_{I\setminus i}$ to the facilities $i'' \in I^\star$. Thereby, we map each facility

$i' \in I^\star_{I\setminus i}$ to that facility i'' in $I^\star$ with which it shares the most customers such that

$$i' \longrightarrow i'' = \arg\max_{i\in I} \sum\nolimits_j \left\lceil x^{s^\star_{I\setminus i}}_{i'j} \right\rceil \cdot \left\lceil x^{s^\star}_{ij} \right\rceil . \tag{2.30}$$

Then, we can determine the degree to which these implicit divisions match.

Definition 2.17. *The **individual matching score** of the optimal solution $s^\star_{I\setminus i}$ obtained without facility i and the optimal solution $s^\star$, $IM\left(s^\star_{I\setminus i}, s^\star\right)$, measures the proportion of customers allocated to the same facility in $s^\star$ that are also allocated to the same facility in $s^\star_{I\setminus i}$ such that*

$$IM\left(s^\star_{I\setminus i}, s^\star\right) := \left(\sum\nolimits_{i\in I}\sum\nolimits_{j\in J}\left\lceil x^{s^\star}_{ij}\right\rceil\right)^{-1} \sum\nolimits_{i'\in I^\star_{I\setminus i}} \max_{i''\in I^\star} \sum\nolimits_{j\in J} \left\lceil x^{s^\star_{I\setminus i}}_{i'j}\right\rceil \cdot \left\lceil x^{s^\star}_{i''j}\right\rceil . \tag{2.31}$$

Notice that ceiling all allocation decisions implies that customers assigned to multiple facilities may be counted more than once. Therefore, to ensure an upper bound for IM of 1.0, we divide not by the number of customers but by the total number of allocations.

As customers and allocation volumes may be very heterogeneous, a better evaluation of this mapping might be achieved when one considers the shared demand instead of the customer count shared by i' and i''.

Definition 2.18. *The **individual demand matching score** of the optimal solution $s^\star_{I\setminus i}$ obtained without facility i and the optimal solution $s^\star$, $IDM\left(s^\star_{I\setminus i}, s^\star\right)$, measures the proportion of the demand served by the same facility in $s^\star$ that is also served by the same facility in $s^\star_{I\setminus i}$ such that*

$$IDM\left(s^\star_{I\setminus i}, s^\star\right) := \frac{1}{\sum_{i\in I}\sum_{j\in J} x^{s^\star}_{ij} D_j} \sum\nolimits_{i'\in I^\star_{I\setminus i}} \max_{i''\in I^\star} \sum\nolimits_{j\in J} \min\{x^{s^\star_{I\setminus i}}_{i'j}, x^{s^\star}_{i''j}\} \cdot D_j. \tag{2.32}$$

IM and IDM map every facility individually. Consequently, according to previous observations, we expect it to be high when comparing two solutions predominantly composed of independent facilities. We expect it to be low when comparing solutions composed predominantly of subsets of interdependent facilities. For the latter, we assume that a good matching requires jointly mapping facilities in $Rep(i)$ to the facilities in $Dep(i)$.

Definition 2.19. *The **joint matching score** of the optimal solution $s^\star_{I\setminus i}$ obtained without facility i and the optimal solution $s^\star$, $JM\left(s^\star_{I\setminus i}, s^\star\right)$, measures the proportion of customers allocated jointly to a subset of facilities in $s^\star$ that is also jointly allocated to that subset's set of replacements in $s^\star_{I\setminus i}$ such that*

$$\begin{aligned} JM\left(s^\star_{I\setminus i}, s^\star\right) := \left(\sum\nolimits_{i\in I}\sum\nolimits_{j\in J}\lceil x_{ij}\rceil\right)^{-1} & \sum\nolimits_{j\in J} \left[\sum\nolimits_{i'\in Rep(i)} x^{s^\star_{I\setminus i}}_{i'j}\right] \cdot \left[\sum\nolimits_{i''\in Dep(i)\cup\{i\}} x^{s^\star}_{i''j}\right] \\ & + \sum\nolimits_{i'\in I^\star_{I\setminus i}\setminus\{Rep(i)\}} \max_{i''\in I^\star\setminus Dep(i)\cup\{i\}} \sum\nolimits_{j\in J} \left\lceil x^{s^\star_{I\setminus i}}_{i'j}\right\rceil \cdot \left\lceil x^{s^\star}_{i''j}\right\rceil . \end{aligned} \tag{2.33}$$

Definition 2.20. *The **joint demand matching score** of the optimal solution $s^{\star}_{I\setminus i}$ obtained without facility i and the optimal solution $s^{\star}$, $JDM\left(s^{\star}_{I\setminus i}, s^{\star}\right)$, measures the proportion of the demand jointly served by a subset of facilities in $s^{\star}$ that is also jointly served by that subset's set of replacements in $s^{\star}_{I\setminus i}$ such that*

$$JDM\left(s^{\star}_{I\setminus i}, s^{\star}\right) := \frac{1}{\sum_{i\in I}\sum_{j\in J} x^{s^{\star}}_{ij} D_j}\Bigg(\sum\nolimits_j \min\left\{\sum\nolimits_{i''\in Dep(i)\cup\{i\}} x^{s^{\star}}_{i''j}, \sum\nolimits_{i'\in Rep(i)} x^{s^{\star}_{I\setminus i}}_{i'j}\right\}\cdot D_j + \sum\nolimits_{i'\in I^{\star}_{I\setminus i}\setminus\{Rep(i)\}} \max_{i''\in I^{\star}\setminus Dep(i)\cup\{i\}} \sum\nolimits_j \min\{x^{s^{\star}_{I\setminus i}}_{i'j}, x^{s^{\star}}_{i''j}\}\cdot D_j\Bigg). \tag{2.34}$$

When we determine the quality of the joint matching, we consider the facilities in $Dep(i)$ and $Rep(i)$ jointly. We determine the degree to which these larger customer regions match and then determine the matching for the remaining facilities in $I^{\star}_{I\setminus i}$ as before. Since in JM and JDM it is not possible to match the customers of a facility operating in $I^{\star}_{I\setminus i}$ which is not part of $Rep(i)$ to a facility that is part of $Dep(i)$, the matching achieved by JM or JDM may be of lower quality than that of IM or IDM, in particular, when there are several independent facilities. Taking the average differences of the scores obtained for all facilities operating in the optimal solution yields the following concise measures.

Definition 2.21. *The **customer reallocation** denotes the average arithmetic difference between $IM\left(s^{\star}_{I\setminus i}, s^{\star}\right)$ and $JM\left(s^{\star}_{I\setminus i}, s^{\star}\right)$ across all solutions in $\mathcal{S}_{I\setminus i}$ such that*

$$CRA\left(s^{\star}, \mathcal{S}_{I\setminus i}\right) := \left(|I^{\star}|\right)^{-1}\cdot\sum\nolimits_{i\in I^{\star}}\left[JM\left(s^{\star}_{I\setminus i}, s^{\star}\right) - IM\left(s^{\star}_{I\setminus i}, s^{\star}\right)\right]. \tag{2.35}$$

Definition 2.22. *The **demand reallocation** denotes the average arithmetic difference between $IDM\left(s^{\star}_{I\setminus i}, s^{\star}\right)$ and $JDM\left(s^{\star}_{I\setminus i}, s^{\star}\right)$ across all solutions in $\mathcal{S}_{I\setminus i}$ such that*

$$DRA\left(s^{\star}, \mathcal{S}_{I\setminus i}\right) := \left(|I^{\star}|\right)^{-1}\cdot\sum\nolimits_{i\in I^{\star}}\left[JDM\left(s^{\star}_{I\setminus i}, s^{\star}\right) - IDM\left(s^{\star}_{I\setminus i}, s^{\star}\right)\right]. \tag{2.36}$$

The larger CRA and DRA, the larger the level of reallocation of customers within interdependent subsets of facilities between the solutions in $\mathcal{S}_{I\setminus i}$ compared to the optimal solution $s^{\star}$.

Example A 2.10 (Level of demand and customer reallocation) Table 2.6, Table 2.7, and Table 2.8 display the results of the individual and joint mapping of the implied customer regions for every pair $s^{\star}_{I\setminus i}$ and $s^{\star}$, with $i \in I^{\star}$ for instances $\mathcal{P}_1$, $\mathcal{P}_2$, and $\mathcal{P}_4$, from Example A, respectively.

For $\mathcal{P}_1$, we have that $CRA\left(s^{\star}, \mathcal{S}_{I\setminus i}\right) = 0.94 - 0.94 = 0.00$. Similarly, $DRA\left(s^{\star}, \mathcal{S}_{I\setminus i}\right) = 0.94 - 0.94 = 0.0$. We know from previous experiments that the optimal solution to $\mathcal{P}_1$ is primarily composed of independent facilities. Thus, CRA and DRA equaling 0 confirms our hypothesis that in solutions primarily composed of independent facilities, the implicit division of the customer space into subsets allocated to the same facility observable in $s^{\star}$ is preserved throughout the solutions in $\mathcal{S}_{I\setminus i}$.

i	$Dep(i)$	$Rep(i)$	IM	IDM	JM	JDM
7	{7}	∅	0.93	0.93	0.93	0.93
11	{11}	{27}	1.00	1.00	1.00	1.00
12	{12}	∅	0.93	0.93	0.93	0.93
14	{14}	{74}	0.97	0.96	0.97	0.96
15	{15}	{74}	0.95	0.95	0.90	0.90
24	{24}	∅	0.89	0.92	0.89	0.92
26	{26, 44}	{54}	0.88	0.89	0.90	0.91
28	{28}	∅	0.96	0.94	0.95	0.94
44	{44}	∅	0.89	0.91	0.89	0.90
49	{49}	∅	0.95	0.94	0.95	0.94
52	{52}	∅	0.91	0.91	0.91	0.91
69	{69}	∅	0.94	0.93	0.94	0.93
70	{70}	{13}	0.93	0.93	0.93	0.93
71	{71}	{57}	0.98	0.97	0.98	0.97
76	{76}	∅	0.91	0.92	0.91	0.92
79	{79}	∅	0.90	0.90	0.90	0.90
82	{82}	{22}	0.94	0.96	0.94	0.96
96	{96}	{38}	0.97	0.97	0.97	0.97
Average:			0.94	0.94	0.94	0.94

Table 2.6.: Pairwise matching scores for $s^{\star}_{I\setminus i}$ and $s^{\star}$ ($\mathcal{P}_1$, Ex. A)

i	$Dep(i)$	$Rep(i)$	IM	IDM	JM	JDM
15	{15, 24, 71, 79}	{21, 84, 94, 97}	0.87	0.86	0.95	0.96
17	{17, 71, 79}	{63, 66, 97}	0.85	0.85	0.92	0.94
24	{24, 67}	{21, 83}	0.96	0.98	0.98	1.00
29	{24, 29, 71, 79}	{21, 33, 66, 97}	0.85	0.88	0.92	0.95
67	{24, 67}	{21, 83}	0.96	0.98	0.98	1.00
71	{17, 71, 79}	{63, 66, 97}	0.85	0.85	0.92	0.94
79	{17, 71, 79}	{63, 66, 97}	0.85	0.85	0.92	0.94
Average:			0.88	0.89	0.94	0.96

Table 2.7.: Pairwise matching scores for $s^{\star}_{I\setminus i}$ and $s^{\star}$ ($\mathcal{P}_2$, Ex. A)

The optimal solutions to $\mathcal{P}_2$ and $\mathcal{P}_4$ are predominantly composed of interdependent facilities. In Table 2.7 and Table 2.8, we see that for these instances JM and JDM yield a mapping score significantly above IM and IDM. In particular, for $\mathcal{P}_2$, the mapping score could be improved significantly for the larger subsets of facilities like $Dep(15)$, where the difference between JM and IM is 0.08. In total, we have that $CRA = 0.06$ and $DRA = 0.07$ for $\mathcal{P}_2$.

i	$Dep(i)$	$Rep(i)$	IM	IDM	JM	JDM
54	{54, 55, 103, 159, 162}	{10, 19, 147, 152, 169}	0.71	0.71	0.76	0.77
55	{55, 119, 135, 159}	{10, 127, 155, 195}	0.80	0.79	0.87	0.87
66	{55, 66, 119, 135, 154, 159}	{10, 106, 125, 155, 160, 165}	0.76	0.77	0.94	0.96
78	{54, 55, 78, 159}	{10, 38, 152, 195}	0.82	0.81	0.83	0.83
103	{54, 55, 103, 159, 162}	{10, 19, 147, 152, 169}	0.71	0.71	0.76	0.77
107	{107}	{123}	0.99	0.99	0.99	0.99
119	{55, 119, 135, 159}	{10, 127, 155, 195}	0.80	0.79	0.87	0.87
128	{128}	{184}	0.99	0.98	0.99	0.98
135	{55, 119, 135, 159}	{10, 127, 155, 195}	0.80	0.79	0.87	0.87
154	{55, 119, 135, 154, 159}	{10, 106, 152, 155, 165}	0.76	0.76	0.88	0.88
159	{55, 119, 135, 159}	{10, 127, 155, 195}	0.80	0.79	0.87	0.87
162	{54, 55, 103, 159, 162}	{10, 19, 147, 152, 169}	0.71	0.71	0.76	0.77
Average:			0.80	0.80	0.86	0.87

Table 2.8.: Pairwise matching scores for $s^\star_{I\setminus i}$ and $s^\star$ ($\mathcal{P}_4$, Ex. A)

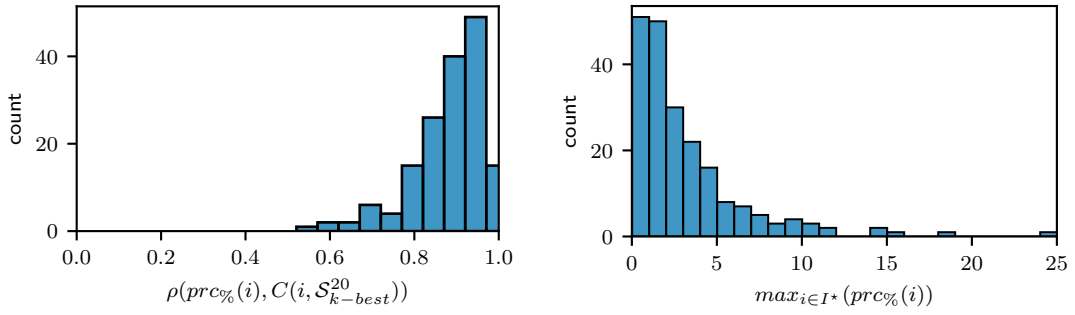
For $\mathcal{P}_4$, $CRA = 0.06$ and $DRA = 0.07$. The higher levels of reallocation support the idea that interdependent facilities serve larger customer regions jointly, leading to a reallocation of customers or demands upon the removal of individual candidates within these interdependent subsets. ▲

The experiments on the problem instances from Example A in Example A 2.3-Example A 2.8 suggest that spatial patterns implied by the transport cost matrix combined with the capacity-demand ratio imply a separation of the facility-customer space. Subsets of customers that are served jointly by one or more facilities. When these subsets are small enough that a single facility optimally serves them, these regions manifest as persistent subsets of favorable core facilities. These facilities operate independently from each other, and upon removal, the customers in the set are redistributed to existing or new facilities. In contrast, when these subsets are larger, they are served best by a subset of facilities. Upon removing a single facility of this subset, the entire subset is replaced by the second-best subset to serve this customer region. Individual facilities will likely operate less persistently throughout well-performing solutions, e.g., the set of k best alternative solutions. In either case, the persistent decision pattern is the implied division of the customer space into distinct service regions rather than individual operating facilities. We refer to these persistent subsets as service regions.

2.3.4. Experimental validation: persistent service regions

We evaluate to which degree the main conclusions drawn from the experiments on the four problem instances from Example A can be confirmed by repeating the same experiments on the data sets introduced in Section 1.3.

We examined the relationship between the persistence of individual facilities throughout k best alternative solutions and their contribution to the optimal objective value quantified via the pseudo-reduced costs for instances $\mathcal{P}_1$ to $\mathcal{P}_4$. We observed that while facilities exhibiting a higher persistence level also exhibit higher pseudo-reduced costs, the range of the latter varies tremendously amongst the optimal solutions to different instances, making them unfit as an indicator for high persistence levels. We determine $C(i, \mathcal{S}^{20}_{k-best}\})$ and $prc_{\%}(i)$ for every facility operating in the optimal solution to every instance from the data sets described in Section 1.3. Figure 2.15 depicts the results. Figure 2.15a displays a histogram of the Spearman's correlation coefficients obtained for every instance between the persistence values and the pseudo-reduced costs of individual facilities denoted by $\rho(prc_{\%}(i),\ C(i, \mathcal{S}^{20}_{k-best})$. We only included correlation coefficients for whom the p-value of the associated two-sided t-test testing the null hypothesis that the correlation coefficient is 0 was below 1%. The latter means the null hypothesis can be rejected. We had to remove 46 of 206 reported entries. For the remaining 160 entries, a strong correlation is visible within individual instances. All values exceed 0.5, and the average is 0.88. Meanwhile, Figure 2.15b displays the distribution of the maximum pseudo-reduced costs of the facilities operating in the optimal solution to different instances. An immense spread is observable. Several instances do not contain any facility whose associated pseudo-reduced costs exceed 1%. For other facilities, individual facilities have a $prc_{\%}$-value of over 20%. This confirms our previous observation. Within individual instances, those facilities making a huge contribution to the objective value exhibit a relatively higher persistence across well-performing solutions to that instance. Still, the pseudo-reduced costs differ tremendously between different instances. Thus, they can neither be used to anticipate the persistence of a facility in a new, yet unexamined, problem instance nor serve as an explanation for the high persistence of individual facilities.



(a) **(b)**

Figure 2.15.: Distribution of $prc_{\%}$-values and their relation to the persistence in 20 best alternative solutions

The facilities operating in the optimal solution to different instances exhibit different levels of interdependence. However, Example A 2.6 and Example A 2.7 have shown that even in instances with significant interdependence relationships between facilities, one can distinguish regions of stronger and weaker coherence. We presented the average dependence density as a measure that indicates the level of interdependence between facilities in the optimal solution to an instance. The closer it is to 1.0, the more dependency relationships

exist between optimal facilities. Furthermore, we presented the weak and strong subset coefficients. These measures are also restricted to the interval $[0, 1]$. The larger they are, the more distinctly individual subsets of facilities can be separated from each other in the underlying dependence and interdependence graphs. Figure 2.16 displays the distribution of these three measures for the instances from Section 1.3. We observe that the strong and weak subset coefficients decrease with increasing average dependence density. This implies that the more dependency relationships, the fewer weakly or strongly connected components can be derived from the interdependence graphs. However, it must be emphasized that even for relatively high average dependence densities, the strong and weak subset coefficients exceed the minimum they would take if all facilities belonged to the same component. This shows that regions of stronger and, in particular, weaker coherence can be found throughout all considered instances.

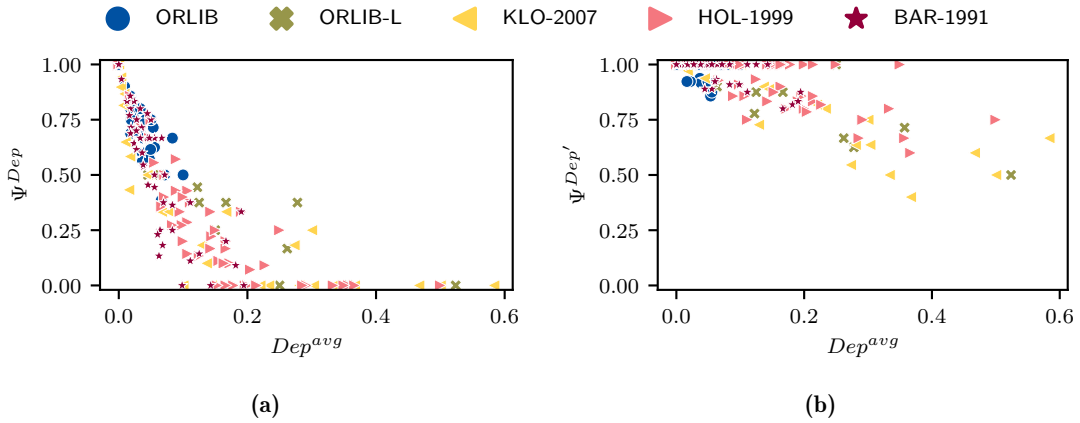


Figure 2.16.: Separability of optimal solution into individual components

We hypothesized that the underlying explanation for the interdependence of some subsets of facilities is the fact that they and their replacements serve subsets of customers jointly in the sense that customers are redistributed among facilities in that subset when the subset is replaced. This implicitly induces a separability of the problem into different service regions. Figure 2.17a and Figure 2.17b display the development of the CRA and DRA for increasing dependency levels in the optimal solutions as indicated by Dep^{avg}. For both, there is a clear tendency that with increasing interdependence, a joint mapping is better suited to capture the customer service regions and that the level of reallocation increases. Spearman's correlation coefficient between CRA and Dep^{avg} is 0.69 with a p-value of 0.00, and for DRA and Dep^{avg}, the correlation coefficient between is 0.72 with a p-value of 0.00.

In conclusion, the experiments show that the main observations on instances from Example A are representative of other instances of interest.

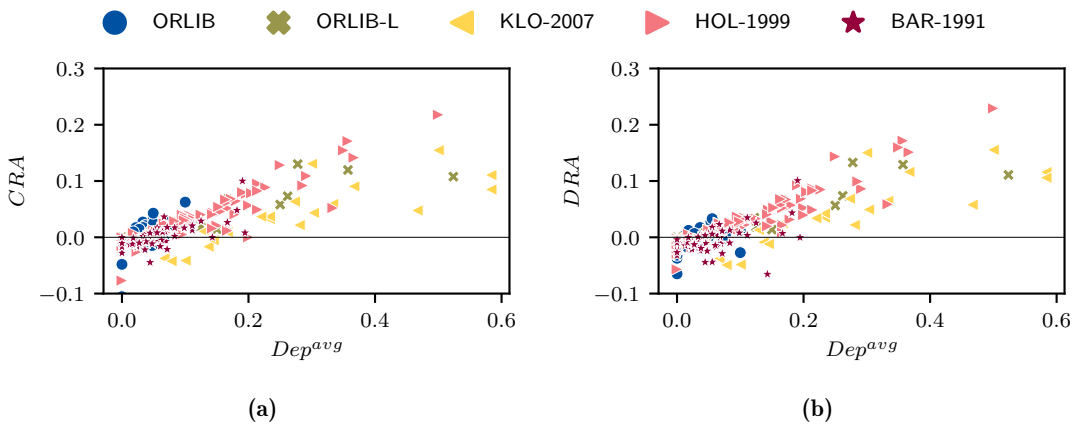


Figure 2.17.: Level of customer reallocation versus level of interdependence

2.4. Conclusion

We analyze multi-sets of well-performing solutions to instances of the CFLP to derive characteristic decision patterns these solutions share. The ability to detect and measure characteristic patterns in the decision space is the foundation for further analysis linking properties of the problem's input data or the performance of certain solution algorithms to the decisions. To the best of our knowledge, a structured analysis of patterns in the decision space has not been done before. In particular, when reviewing literature analyzing the decision space of discrete location problems, one often encounters an implied notion that subsets of core facilities operate in all optimal or near-optimal solutions. We present a concise measure that quantifies the size and relevance of such subsets and perform extensive analyses on sets of k best alternative solutions to problem instances from literature. Results show that some but not all instances exhibit subsets of favorable core facilities. Furthermore, their existence is not necessarily linked to a particularly "structured" spatial pattern of the distribution of candidates and customers.

We evaluate a particular multi-set of solutions for a given instance to explore persistent patterns across well-performing solutions. We first derive concise measures that quantify the relevance of individual facilities in the objective value space and link them to their persistence in well-performing solutions. As no systematic relationship can be observed, we turn to the decision space to evaluate the dependence relationships between individual optimal facilities. We observe that in solutions that exhibit persistent subsets of favorable core facilities, these facilities are independent in that, upon their removal, they are either substituted, or customers are redistributed to existing facilities. In contrast, facilities operating in solutions that do not exhibit persistent cores often interdepend. Upon removing a single facility, other facilities are no longer optimal, and an entire subset of facilities is replaced. We present concise measures to quantify these interdependence relationships and link them to the persistence of core facilities. We then show that the interdependence of facilities in a solution is related to these facilities serving larger subsets of customers jointly. Upon removing a single facility, customers served by any facility in the associated interdependent subset are reallocated to facilities in the set of replacements. Again, we develop a means to quantify this observation and validate it on sets of problem instances

from literature.

The main results from the previous chapter can be summarized as follows:

- Persistent subsets of favorable core facilities characterize well-performing solutions to some but not all instances of the CFLP.
- The persistence of facilities throughout well-performing solutions is not necessarily linked to these facilities' contribution to the optimal objective value. Instead, it is affected by the degree to which this facility depends on other facilities to be optimal.
- The combination of spatial patterns implied by the variation in the unit net profit, the ratio between demand and capacity, and the mixed-binary component of the CFLP result in facilities in the optimal solutions of different instances interdepending to varying degrees. Independent facilities serve smaller subsets of customers optimally. Interdependent facilities serve larger subsets of customers optimally in combination. This induces an implicit division of candidates and customers into distinct service regions. These service regions persist through well-performing solutions. This leads instances whose optimal solutions are primarily composed of independent facilities to exhibit persistent subsets of favorable core facilities.

Up to this point, we inferred customer service regions from a particular set of solution $\mathcal{S}_{I\setminus i}$. Thereby, the inferred regions have varying levels of coherence and sometimes overlap. In the following Chapter 3, we present a means to derive service regions from arbitrary sets of solutions to a particular instance. This separation into regions of stronger and weaker coherence offers new opportunities for solving and understanding the problem. In Chapter 4, we evaluate whether service regions can be detected already from the problem's input data without solving the CFLP even once.

3. Identifying service regions

In the previous Chapter 2, we established that sets of well-performing solutions to a particular instance of a CFLP exhibit a persistent division of the facility-customer space into service regions. These regions were observed in a specific set of solutions. However, they largely overlapped and exhibited varying coherence levels regarding the interdependence of associated location decisions. The upcoming chapter aims to find an algorithmic approach to derive service regions of a defined coherence level from an arbitrary set of solutions, in particular, from a set of solutions that does not necessarily include the optimal solution to the problem instance. This leads us to the following research question.

RQ2: How can persistent service regions be identified in arbitrary sets of solutions?

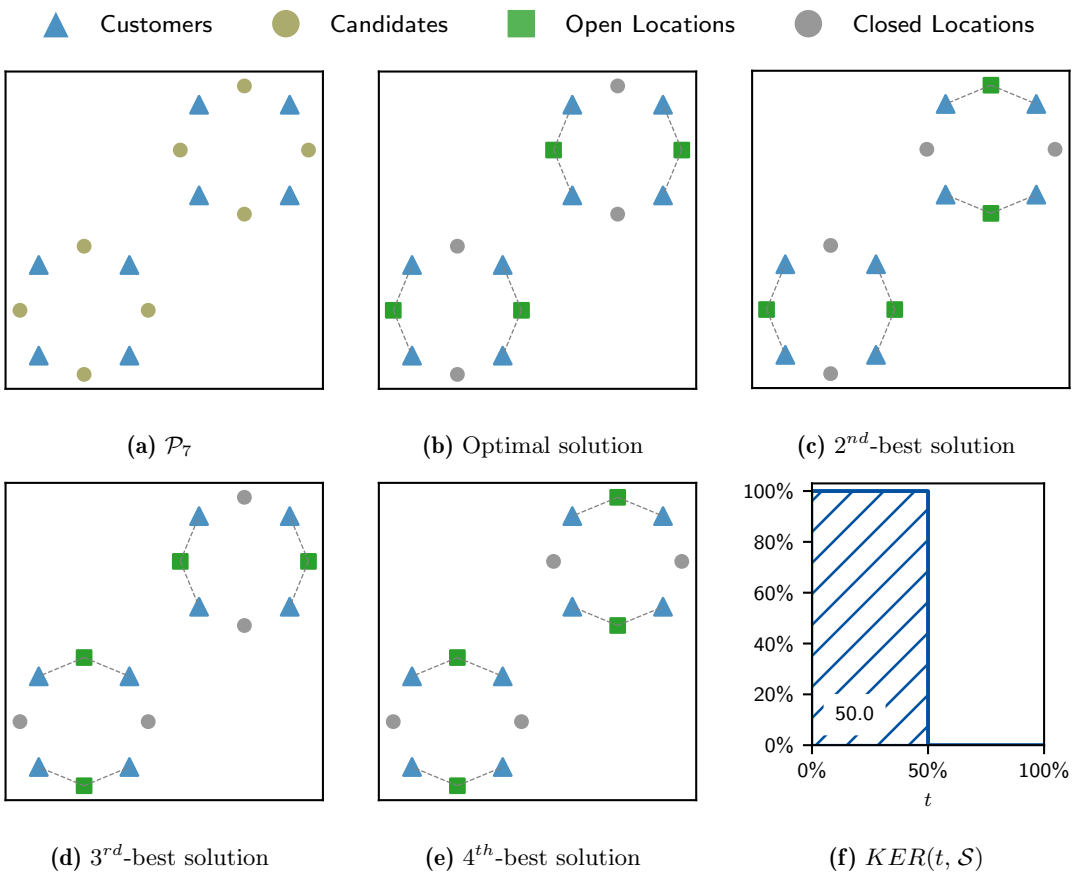
Detecting service regions without using the visual representation of the facility-customer space and without knowledge of the interdependence relationships implied by $\mathcal{S}_{I\setminus i}$ is a non-trivial task. Measures that are based solely on location decisions, like kernel persistence, cannot identify commonalities in solutions composed of interdependent subsets of facilities. In particular, these measures are oblivious to the implicit division into service regions that may persist across most or even all optimal and near-optimal solutions to a particular problem instance. The following example demonstrates this.

Example E Let there be an instance $\mathcal{P}_7$ with 8 candidate facilities and 8 customers. All customers have the same demand, and each candidate facility has the capacity to serve exactly two customers, such that $Q_i = 2D_j$ for all $j \in J,\ i \in I$. Furthermore, the locations of the candidates and customers are such that the problem can be separated into 2 visibly distinct regions as depicted in Figure 3.1a. Variable transportation costs are proportional to the Euclidean distances, and profits are the same for all customers.

Figure 3.1 depicts the 4 best solutions to $\mathcal{P}_7$. As a direct consequence of the described spatial pattern, the customers can be separated into two service regions that are served by distinct subsets of interdependent facilities. However, $KER(\mathcal{S})$ has no way to identify this common structure and is not only relatively low with a value of 0.5 but equals its expected value under symmetry, $\mathbb{E}(KER(\tilde{\mathcal{S}}))$. ▲

Persistent service regions cannot be identified solely from the location decisions. The kernel persistence is oblivious to the fact that certain subsets of customers are served jointly by subsets of facilities. Information on the allocations of customers to facilities is necessary. However, counting the frequency with which individual allocations occur similarly to $KER(\mathcal{S})$ does not yield additional insights as it fails to detect the interdependence between individual facilities. Still, coherent patterns in the matrices of the allocation decisions that represent the persistent service regions must exist. These are obvious when looking at the graphic representation of the problem instance and its solutions. Yet, identifying them purely from the data is a non-trivial task linked to the field of pattern recognition.

Pattern recognition is the theory of developing algorithms and statistical models to automatically recognize patterns in data and make predictions or decisions based on these



(a) $\mathcal{P}_7$ **(b)** Optimal solution **(c)** 2^{nd}-best solution

(d) 3^{rd}-best solution **(e)** 4^{th}-best solution **(f)** $KER(t, \mathcal{S})$

Figure 3.1.: 4 best alternative solutions and the development of $KER(t, \mathcal{S})$ ($\mathcal{P}_7$, Ex. E)

patterns. It is closely related to data mining and machine learning, yet each field has a different focus and set of techniques. While machine learning is mainly concerned with developing models that can learn from data and make predictions or decisions based on that, data mining has a stronger emphasis on techniques for exploratory data analysis. The grouping of data according to some similarity criteria is a core task of pattern recognition. When predefined group labels exist, this task is termed classification. When no labeling information exists, it is called clustering [Bishop, 2007].

Our objective to identify coherent service regions in a set of solutions to discrete location problems falls under the field of pattern recognition, particularly under the specific task of clustering. We want to assign customers and candidate locations (the data inputs) to different service regions (the groups). No previous knowledge about these groups exists. Clustering does not denote a specific algorithm but refers to the general task. Rokach and Maimon [2005] describe clustering as the task of grouping data instances into subsets so that similar instances are grouped together while different instances belong to different groups. In pattern recognition, the input data usually comes in the form of a feature matrix in which the rows correspond to a set of observations or data instances described by a set of features – the columns of the matrix.

Definition 3.1. *Given a data set of n samples and m features, the* ***feature matrix****, $A = (a_{ij})_{n \times m}$, is a rectangular matrix where a_{ij} is the value of the j-th feature in the i-th sample.*

Definition 3.2. *A **clustering** is a collection of subsets of samples, $\mathcal{C} = (C_1, \ldots, C_r)$, such that*

$$\begin{aligned} &C_k \subseteq \{1, \ldots, n\} && k \in \{1, \ldots, r\}, \\ &C_1 \cup C_2 \cup \cdots \cup C_r = \{1, \ldots, n\}, \text{ and} && \\ &C_k \cap C_l = \emptyset && k, l \in \{1, \ldots, r\}, k \neq l. \end{aligned} \tag{3.1}$$

Every sample belongs to exactly one cluster.

In the present application, a feature matrix as such is not apparent. Different possibilities to derive such a matrix from a set of solutions exist, and we evaluate several approaches in Section 3.2. However, the problem of identifying coherent regions of customers that are jointly served bears the particularity that we are not only interested in grouping the customers but also the facilities that serve them. Thereby, customers and facilities are different data inputs with different characteristic features. Listing customers and facilities as data instances (rows) of the same feature matrix raises the question regarding what common features that induce a grouping should look like. We avoid this question by approaching the problem with a variant of traditional clustering, biclustering, which allows us to group the facilities together with the customers.

Biclustering is a variant of clustering, also known as two-dimensional clustering or co-clustering. It extends the idea of conventional clustering by simultaneously grouping the set of samples (rows) and the set of features (columns) of a feature matrix. Thereby, samples and features are grouped together so that they have high relevance to each other. The samples and the features are divided into a collection of sample and corresponding feature clusters. The sample clusters are defined according to (3.1). Similarly, the feature clusters obey

$$\begin{aligned} &F_k \subseteq \{1, \ldots, n\} && k \in \{1, \ldots, r\}, \\ &F_1 \cup F_2 \cup \cdots \cup F_r = \{1, \ldots, m\}, \text{ and} && \\ &F_k \cap F_l = \emptyset && k, l \in \{1, \ldots, r\}, k \neq l, . \end{aligned} \tag{3.2}$$

The rationale behind creating these clusters is that the features in cluster F_k are "responsible" for creating the group of samples C_k. Busygin et al. [2008] provide the following formal definition of biclustering:

Definition 3.3. *A **biclustering** of a data set is a collection of pairs of sample and feature subsets $\mathcal{R} = ((C_1, F_1), (C_2, F_2), \ldots, (C_r, F_r))$ such that the collection $(C_1, C_2, \ldots, C_r)$ forms a partition of the set of samples, and the collection $(F_1, F_2, \ldots, F_r)$ forms a partition of the set of features. A pair (C_k, F_k) will be called a **bicluster**.*

The pattern recognition task of biclustering is the methodological concept for identifying coherent service regions. Like clustering, biclustering refers to the task, not a specific algorithm. Several algorithms and performance metrics exist. As it is a concept alien

to conventional OR literature, we provide a general overview of the idea of biclustering, commonly used algorithms, and evaluation criteria in Section 3.1. Thereby, we provide the necessary background on the algorithms and metrics used in what follows and explain why we chose them for our application. In Section 3.2, we proceed with the present problem and apply biclustering to identify coherent service regions in sets of solutions to discrete location problems.

3.1. Related work on the basic concepts of biclustering

In contrast to clustering, which focuses on identifying global patterns in the sense that when clustering rows of the feature matrix, all columns are considered, biclustering is focused on local patterns. Rows are clustered into different groups based on different subsets of columns [Fraiman and Li, 2020].

Biclustering has many applications, particularly in bioinformatics in studying gene expression data [Pontes et al., 2015b], text mining to identify topics in collections of documents [Dhillon, 2001], or, very recently, market research to identify customer segments [Fang et al., 2022]. To the best of our knowledge, biclustering has never been used to identify patterns in the solutions of mathematical programs before. We provide an overview of biclustering algorithms in Subsection 3.1.1 and a detailed explanation of the spectral biclustering algorithm used further in this work in Subsection 3.1.2. In Subsection 3.1.3, we review evaluation criteria for produced biclusterings.

3.1.1. Algorithms for biclustering

Biclustering is the task of identifying biclusters. A wide variety of algorithms have been proposed over the years and are summarized in several review papers. For example, Madeira and Oliveira [2004] classify different bicluster structures in data sets depending on the persistence of the numeric values in the bicluster and group algorithms according to which cluster type they identify and which heuristic approach (greedy, divide-and-conquer, etc.) they follow. Zhao et al. [2012] distinguish between distance-based clustering algorithms, which perform an iterative search for biclusters minimizing a particular variance metric, probabilistic-based biclustering, biclustering for coherent evolution, which detects biclusters that reflect consistent trends in the data rather than consistent numeric values, geometric-based biclustering, which is based on a spatial interpretation of biclusters and identifies biclusters as linear geometric patterns in the high-dimensional data-space, as well as factorization based biclustering which refers to spectral decomposition of the feature matrix to uncover "natural" substructures that are related to the main patterns of the matrix. Recent reviews on biclustering can be found, e.g., in Pontes et al. [2015b], Henriques et al. [2015], and, with a special emphasis on meta-heuristics, in José-García et al. [2022].

While several possible algorithms exist, many are unfit for this work's particular biclustering application. For example, there is no point in identifying sub-spaces in the feature matrix with constant values as the customer demands and facility capacities may induce

different allocation volumes. We want to group those customers served jointly by different subsets of facilities and not those that exhibit similar demand volumes or generate similar profits. However, rather than reasoning against many algorithms from the literature, we explain why the spectral biclustering algorithm chosen in the following is particularly well-suited for our application.

3.1.2. Spectral biclustering

Spectral biclustering (factorization-based biclustering or biclustering via graph-partitioning) is a clustering technique from one-dimensional clustering that can be straightforwardly extended to the two-dimensional space. It relies on the fact that biclustering is closely related to the problem of Singular Value Decomposition (SVD). This matrix factorization generalizes the eigendecomposition of a symmetric matrix, providing the orthogonal basis of eigenvectors [Busygin et al., 2008]. It is based on the idea of an idealized feature matrix that exhibits a block-diagonal structure. In such a matrix, each block can be associated with a bicluster, and each pair of singular vectors designates one such bicluster by the nonzero components in the vectors. This idealized structure implies that the feature matrix A is of the form

$$A = \begin{pmatrix} A_1 & 0 & \dots & 0 \\ 0 & A_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & A_r \end{pmatrix},$$

whereby $\{A_k\}, k = 1, \dots, r$ are arbitrary matrices. Then, for each of these matrices, there will be a singular vector pair such that the nonzero components of these vectors correspond to the rows and columns occupied by A_k. In a less idealized case, when the elements outside the diagonal blocks are not necessarily zero, but diagonal blocks still contain dominating values, the SVD can reveal these biclusters, too, by the dominating components in the singular vector pairs.

The concept is closely related to bipartite spectral graph partitioning. In this context, it was first proposed by Dhillon [2001] for the simultaneous grouping of words and documents in text mining. The feature matrix $A_{n\times m}$ is converted into a bipartite graph $G := (n, m, \mathcal{E})$, with $\mathcal{E}$ a set of edges with edge weights a_{ij} being the entries from the feature matrix. The graph has no edges between two vertices representing samples or between two vertices representing features.

A partitioning of the set of vertices $\mathcal{V} := n \cup m$ into r subsets $V_1, V_2, \dots, V_r$, such that

$$\mathcal{V} = V_1 \cup V_2 \cup \dots \cup V_r \tag{3.3}$$

$$\tag{3.4}$$

and

$$V_k \cap V_l = \emptyset \qquad\qquad k, l \in \{1, \dots, r\}, k \neq l, \tag{3.5}$$

provides a biclustering of the data set. The cost of this partition can be defined as the total weight of the edges cut by it

$$\mathrm{cut}(V_1, V_2, \ldots, V_r) = \sum_{k=1}^{r-1} \sum_{l=k+1}^{r} \sum_{i \in V_k} \sum_{j \in V_l} a_{ij} = \sum_{1 \leq i < j \leq k} \mathrm{cut}(V_i, V_j). \quad (3.6)$$

Consequently, the problem of biclustering a data set can be translated into a problem of bipartite graph partitioning. This naturally fits our application of identifying coherent service regions. The aggregated matrices of allocation decisions of solutions to discrete location problems translate naturally into a bipartite graph with nodes I and J. The edge weights are derived from the aggregated allocation decisions. This makes spectral biclustering a well-fit algorithm for our application. Every bicluster derived from the aggregated matrix of allocation decisions corresponds to one service region.

Unless the graph contains multiple components, a minimization of Eq. (3.6) will likely produce a partition in which all but one bicluster contain zero vertices. Therefore, to detect underlying patterns of stronger and weaker coherence, each vertex $v \in \mathcal{V}$ is assigned a positive weight w_v such that minimizing the following function produces "balanced" biclusters, i.e., clusters of equal weight [Yu, 2003]

$$Q(V_1, V_2, \ldots, V_r) = \frac{1}{r} \sum_{i=1}^{r} \frac{\mathrm{cut}(V_i, \mathcal{V} \setminus V_i)}{\sum_{v \in V_i} w_v}. \quad (3.7)$$

Two standard definitions of the weight of the vertex are either a constant weight of 1, which produces the ratio cut, or the vertex degree, which produces the normalized cut. The minimization of the normalized cut is equivalent to the maximization of the proportion of edge weights that lie within each partition. It furthermore allows the direct link to SVD in graph theory as there is a direct connection between the partitioning of a graph and the singular value decomposition of its Laplacian matrix. The Laplacian matrix is a symmetric matrix that captures the connectivity structure of a graph. Its eigenvalues and eigenvectors provide information on the graph's properties. In particular, the Laplacian L is defined as $L = D - A$, where D is a diagonal matrix whose entries are the degrees of the vertices, and A is the adjacency matrix of the graph. The Laplacian matrix L can be factorized using SVD and $L = U \Sigma V^T$, where U and V are orthogonal matrices and Σ is a diagonal matrix whose entries are singular values of L. The columns of U and V are the left and the right singular vectors of L, respectively. They can be used to compute a low-dimensional embedding of the graph. In particular, the first k columns of U correspond to the k smallest singular values of L. These columns can be used to embed the graph into a k-dimensional space, where the k-th dimension corresponds to the sign of the k-th columns of U. The graph's vertices can then be clustered based on their position in this embedding. Thereby, it can be shown that the k-th smallest singular value of L is proportional to the minimum ratio cut that can be achieved using k clusters. Therefore, the singular values of L provide a natural way to rank the quality of different partitions. A comprehensive tutorial can be found in von Luxburg [2007]. We refer to de Abreu [2007] for background information on graph partitioning.

So, in other words, spectral biclustering uses SVD to transform the feature matrix into

a new space by decomposing it into singular values and eigenvectors that represent the underlying structure of the data. Then, by keeping only the k largest singular eigenvectors, this low-rank approximation captures the most important patterns and features in the data. In this approximation, rows and columns are clustered by a standard clustering algorithm as, e.g., k-means. k-means is a clustering algorithm that partitions a set of data points into k clusters based on their similarity. The algorithm works by randomly initializing k cluster centers, assigning each data point to the nearest cluster center, and then updating the cluster centers based on the mean of the data points in each cluster. This process is repeated until the cluster assignments no longer change or a maximum number of iterations is reached. The spectral biclustering algorithm is summarized in pseudo-code in Algorithm 1. It requires the feature matrix A and the number of biclusters the data is supposed to be partitioned into r as inputs.

Algorithm 1 *Spectral Biclustering*

Input: A, r
Output: $\mathcal{R}$

1: Convert the information in the feature matrix A into a bipartite graph G.
2: Calculate the Laplacian matrix L to G.
3: Calculate the first k eigenvectors, those eigenvectors corresponding to the k smallest eigenvalues of L.
4: Consider the matrix formed by the first k eigenvectors. The l-th row defines the features of graph node l.
5: Cluster the graph nodes based on these features using k-means clustering.
6: **return** $\mathcal{R}$

In summary, spectral biclustering is a matrix-factorization-based biclustering technique that is particularly well-suited to our application due to its natural connection to bipartite graph partitioning. It is a well-established method in data mining and bioinformatics. Several well-documented software packages provide ready-to-use implementations with open source code and descriptions of the algorithms available in peer-reviewed journals such as Role et al. [2019], Padilha and Campello [2017] and, e.g., Pedregosa et al. [2011]. As the contribution of this work is the application of biclustering to extract additional information from a set of solutions to discrete location problems and not the development of novel algorithms for biclustering, we did not implement the biclustering algorithm ourselves but used the open-source implementation of Dhillon [2001]'s algorithm in *Scikit-learn*. It contains a Python class that implements spectral biclustering with options for implementing both the SVD and the integrated k-means algorithm. We deploy the function with default parameter settings.

3.1.3. Evaluation criteria for biclustering

An essential characteristic of clustering problems is that information on the actual set of clusters is usually unavailable. Two key challenges arise in this context: first, to determine the appropriate number of biclusters, and second, to evaluate the quality of the resulting biclustering. Many metrics have been proposed to determine bicluster coherence and evaluate the effectiveness of biclustering algorithms. They can be distinguished into two

main categories: external and internal validation metrics [Liu et al., 2010]. While the former require external groups as a reference point, the latter evaluate the clustering structure based on the degree to which they achieve the generally targeted intra-cluster similarity and inter-cluster dissimilarity. Several different metrics for both aspects and joint metrics that balance these objectives have been proposed.

Internal validation metrics usually combine a measure of the compactness, or similarity, of the data points within a single bicluster with a measure of the separation, or dissimilarity, to the data points in other biclusters [Liu et al., 2010]. A comprehensive overview can be found in Pontes et al. [2015a]. Thereby, a generic way to guarantee the effectiveness of a biclustering algorithm does not exist. Instead, the evaluation metrics must meet the specific objectives of the application.

One measure to compare the similarity of two biclusterings is the Jaccard coefficient, which extends the logic of the Jaccard index for sets to biclusterings [Hochreiter et al., 2010]. It measures the similarity of two biclusters as the quotient of the number of matrix elements contained in the intersection of the biclusters and the number of matrix elements contained in their union.

Definition 3.4. *Let $\mathcal{R}_1$ and $\mathcal{R}_2$ denote two biclusterings on the same data sets with individual biclusters $R_1, R_2, \ldots, R_{r_1} \in \mathcal{R}_1$ and $R_1, R_2, \ldots, R_{r_2} \in \mathcal{R}_2$. The **Jaccard coefficient**, $JC(\mathcal{R}_1, \mathcal{R}_2)$, is defined as follows*

$$JC\left(\mathcal{R}_1, \mathcal{R}_2\right) = \frac{1}{\max\{|\mathcal{R}_1|, |\mathcal{R}_2|\}} \sum_{k=1}^{r_1} \sum_{l=1}^{r_2} \frac{|R_k \cap R_l|}{|R_k \cup R_l|}. \tag{3.8}$$

The Jaccard coefficient is a measure to compare the similarity of two biclusterings. It resides in the interval $[0, 1]$ and equals 1 if the two biclusterings are identical. When one of the two biclusters represents the "true" biclustering underlying the data, it can also be interpreted as an external validation criterion.

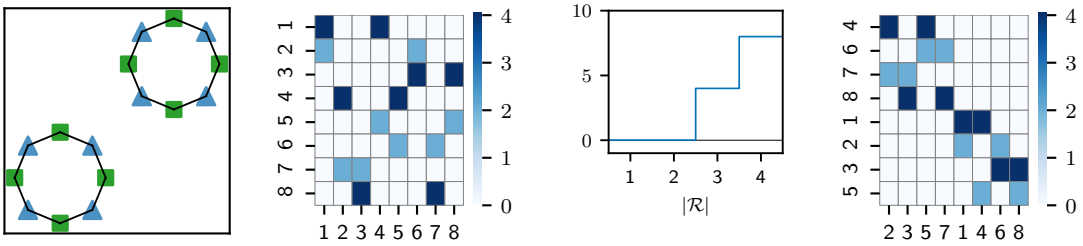
The above presented a general overview of biclustering, notably the spectral biclustering algorithm. The following will discuss how these concepts can be applied to detect coherent service regions from sets of solutions to a discrete location problem instance. To the best of our knowledge, this is the first approach aiming to detect regions of stronger and weaker coherence as a persistent decision pattern among different solutions to an instance of a discrete location problem in particular or any mathematical program in general.

3.2. Biclustering as a means to detect characteristic service regions

We start with a straightforward demonstration of applying spectral biclustering to identify service regions in instance $\mathcal{P}_7$ in Example E. This allows us to illustrate the open questions that arise when applying this method.

Example E continued The bipartite graph displayed in Figure 3.2a summarizes the allocation information from the 5 best solutions to $\mathcal{P}_7$ by connecting each customer to each

candidate to which it is allocated in at least one of the solutions. There is an edge between vertex $i \in I$ and vertex $j \in J$ if some demand of customer j is allocated to facility i in at least one of the solutions in $\mathcal{S}$, such that $\sum_{s \in \mathcal{S}} x_{ij}^s > 0$. The feature matrix $A \in \mathbb{R}^{8 \times 8}$ is constructed accordingly. Each entry denotes the sum of all fractions of demands of customer j served by facility i ($a_{ij} = \sum_{s \in \mathcal{S}} x_{ij}^s \in [0, |\mathcal{S}|]$). Figure 3.2b illustrates the resulting feature matrix as a heatmap.

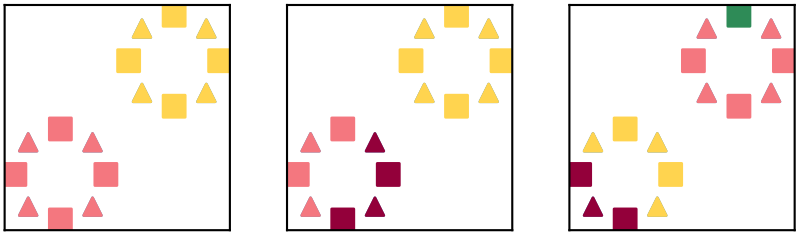


(a) Bipartite graph **(b)** A **(c)** Out-of-bicluster allocations **(d)** A re-ordered

Figure 3.2.: Biclustering of the 5 best alternative solutions ($\mathcal{P}_7$, Ex. E)

Spectral biclustering requires determining the number of biclusters the data is to be grouped into upfront. For instance $\mathcal{P}_7$, we know that there are two distinct service regions to be identified. In general, we can assume that there are, at most, as many biclusters as the maximum number of facilities operating in any solution. We use spectral biclustering to produce 2, 3, and 4 biclusters successively. For each of the resulting biclusterings $\mathcal{R}_r$, which group the set of customers and the set of candidates into r regions, we count the frequency with which a customer has been allocated to a facility from another bicluster across all the solutions in $\mathcal{S}$. Figure 3.2c illustrates how the number of out-of-bicluster-allocations increases with an increasing number of biclusters ($|\mathcal{R}|$). Since the bipartite graph has two connected components, it is 0 when $r = 1$ and $r = 2$. When the number of biclusters the graph is partitioned into increases further, the number of customers allocated to a facility outside its bicluster in at least one of the solutions in $\mathcal{S}$ increases. For $r = 3$, customers are served from non-bicluster facilities 4 times across all solutions. The number of out-of-bicluster allocations increases with an increasing number of biclusters. It only punishes the separation of a region that "should" not be separated but does not yield any punishment when two regions that could be separated are considered as one. In other words, it punishes inter-cluster similarity but does not punish intra-cluster dissimilarity.

One possibility to identify the most suitable biclustering for the data set is to set an upper bound for the number of out-of-bicluster allocations. One then chooses that biclustering that divides the customer-facility space into the maximum number of biclusters without exceeding this bound. If, in the present example, we set this bound to 0 as we know that there is a perfect (loss-free) separation of the facility-customer space, we obtain a biclustering into two biclusters as depicted in Figure 3.3a. Given this biclustering, we can reorder the rows and columns of the feature matrix A. The heatmap of the so-obtained matrix is presented in Figure 3.2d. It exhibits the block-diagonal structure we were aiming for.



(a) $r = 2$ **(b)** $r = 3$ **(c)** $r = 4$

Figure 3.3.: Biclusterings for different cluster numbers (Example E)

▲

Example E shows that one has several degrees of freedom when using spectral biclustering to identify persistent service regions in a set of solutions to a location-allocation instance. In the following, we explain how we handled the following questions:

1. How to derive the feature matrix $A_{|I|\times|J|}$ from the set of solutions $\mathcal{S}$?
2. How to assess the quality of a biclustering (separation into service regions)?
3. How to determine an appropriate number of biclusters (service regions)?

We answer these three questions sequentially in the following. We then present *RegClust*, an algorithm that determines persistent service regions among solutions to a particular CFLP instance. We conclude with a detailed illustration of the procedure based on the instances from Example A.

For a consistent notation, let $\mathcal{R}_r := \{R_1, R_2, \ldots, R_r\}$ denote a collection of r regions that constitute a biclustering. The subsets of candidate facilities and customers assigned to region R_k are denoted by $I_k \subseteq I$ and $J_k \subseteq J$, $k \in 1, \ldots, r$, respectively.

3.2.1. Derivation of the feature matrix

In traditional biclustering applications, the feature matrix is the starting point of the analysis. In contrast, several options exist to aggregate the information in a (multi-)set of solutions $\mathcal{S}$ into a feature matrix A. The elements of this feature matrix constitute the edge weights in the bipartite graph. As of the capacity constraint in the CFLP (see Eq. (1.3)), all information on the opening status of individual candidates in any solution is implicitly contained in the matrix of the allocation decisions $x^s \in \mathbb{R}^{|I|\times|J|}$. This makes an explicit consideration of the vector y^s of binary location decisions obsolete. Consequently, the information of interest is contained in $|\mathcal{S}|$ $|I| \times |J|$-matrices, and the question is how the information in these matrices should be aggregated. Furthermore, the question arises whether or not additional information from the parameters of the problem instance, e.g., demand volume or profits should be included in A. We present five alternative options to derive A and evaluate the degree to which different A lead to different biclusterings.

Definition 3.5. *Let $\alpha : (\mathcal{S}, \mathcal{P}) \longrightarrow A$ denote an **aggregation function** that converts the information contained in the set of solutions $\mathcal{S}$ and the problem instance $\mathcal{P}$ into a $|I| \times |J|$-dimensional feature matrix A.*

α^{sum}	A^{sum}	$a_{ij} := \sum_{s \in \mathcal{S}} x_{ij}^s$	Sum of allocation decisions across all solutions in $\mathcal{S}$
$\alpha^{indicator}$	$A^{indicator}$	$a_{ij} := \mathbb{1}_{(\sum_{s \in \mathcal{S}} x_{ij}^s > 0)}$	Indicator whether customer j is allocated to facility i in at least one of the solutions in $\mathcal{S}$
α^{count}	A^{count}	$a_{ij} := \sum_{s \in \mathcal{S}} \lceil x_{ij}^s \rceil$	Number of solutions in $\mathcal{S}$ in which customer j is allocated to facility i
α^{demand}	A^{demand}	$a_{ij} := \sum_{s \in \mathcal{S}} x_{ij}^s D_j$	Total demand of customer j served by facility i across all solutions in $\mathcal{S}$
α^{profit}	A^{profit}	$a_{ij} := \sum_{s \in \mathcal{S}} x_{ij}^s D_j \, (r_j - c_{ij})$	Total profit generated by serving customer j from facility i across all solutions in $\mathcal{S}$

Table 3.1.: Aggregation functions α to derive a feature matrix A^α from a set of solutions $\mathcal{S}$

Different possibilities to define such an aggregation function are listed in Table 3.1. The functions α^{sum}, $\alpha^{indicator}$, and α^{count} are based solely on the decision variables and do not include any parameters of the associated instances. This implies that allocations that serve a small demand volume or generate only little profit are weighted equally to those serving a large demand volume or generating high profits. α^{demand} includes information only on the demand. α^{profit} includes information on the allocated demand volume and the unit profit, thus the total profit generated by individual allocation decisions.

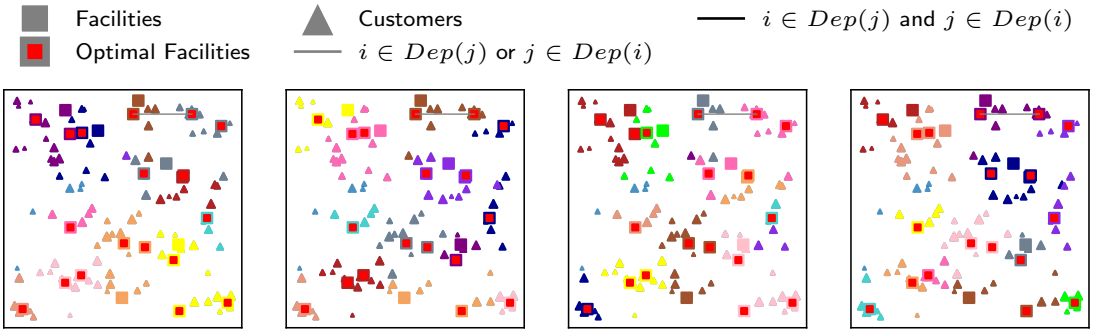
The rows of the feature matrix corresponding to candidate facilities not operating in any solution in $\mathcal{S}$ only have zero entries. Similarly, the columns corresponding to customers not allocated to any facility in any solution only have zero entries. Consequently, A may have neither full row nor column rank. No meaningful way exists to assign these candidates and customers to any region. Therefore, they are omitted from further considerations and assigned to a dummy region. Thus, the set of candidates and customers is reduced to $I^{\mathcal{S}} = \{i \in I | \sum_j a_{ij} > 0\}$ and $J^{\mathcal{S}} = \{j \in J | \sum_i a_{ij} > 0\}$, respectively. This leads to a feature matrix of reduced dimensionality.

Example A 3.1 (Identify service regions via biclustering) Consider instances $\mathcal{P}_1$-$\mathcal{P}_4$ from Example A and the sets of solutions $\mathcal{S}_{I \setminus i}$. Recall that $\mathcal{S}_{I \setminus i}$ is the set of optimal solutions to the modified instances in which exactly one facility operating in the optimal solution is removed from the set of candidates (see Eq. (2.11)). We focus on the problem instances $\mathcal{P}_1$ and $\mathcal{P}_2$ as representatives of instances composed primarily of independent and interdependent facilities, respectively. Experiments on $\mathcal{P}_3$ and $\mathcal{P}_4$ did not yield additional insights.

Based on $\mathcal{S}_{I \setminus i}$, we derive service regions via biclustering using different aggregation functions α and different target numbers of regions r for each instance. We then determine the Jaccard coefficients of different service regions produced with different aggregation functions.

Figure 3.4 depicts the service regions for $\mathcal{P}_1$ derived from the feature matrices $A^{indicator}$ and A^{profit} when separating the facility-customer space into $r = 12$ and $r = 13$ regions, respectively. From the experiments in Chapter 2, we know that for $\mathcal{P}_1$, removing several

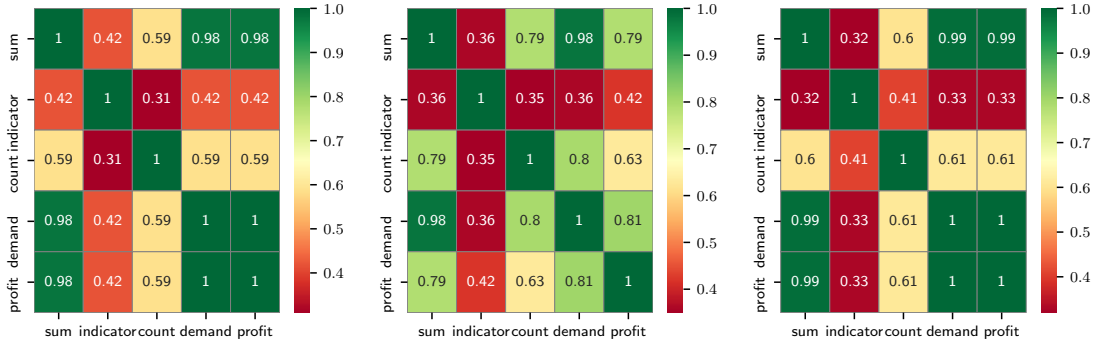
facilities had no effect on new facilities operating and resulted in other facilities taking over their customers. This is reflected throughout the produced service regions. Many regions contain only two facilities, the one operating in the optimal solution and the one operating upon its replacement.



(a) $A^{indicator}$, $r = 12$ **(b)** A^{profit}, $r = 12$ **(c)** $A^{indicator}$ $r = 13$ **(d)** A^{profit}, $r = 13$

Figure 3.4.: Service regions for different functions α and number of regions r ($\mathcal{P}_1$, Ex. A)

Figure 3.5 depicts the pairwise Jaccard coefficients for the service regions obtained with different aggregation functions when the target number of biclusters r is set to 11, 12, or 13 biclusters, respectively. The biclusterings obtained with the aggregation functions α^{sum}, α^{demand}, and α^{profit} are very similar. Their pairwise Jaccard coefficients are close to 1.0. Service regions produced with the aggregation functions $\alpha^{indicator}$ and α^{count} differ more significantly. The associated Jaccard coefficients are relatively low, with most being significantly below 0.5.

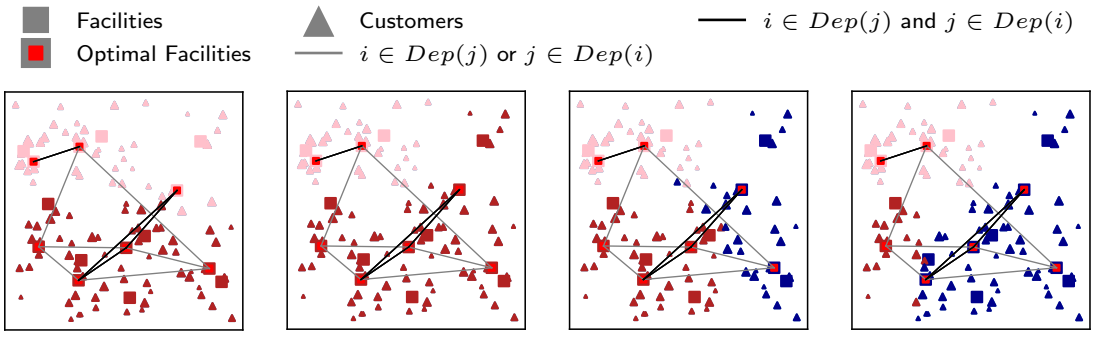


(a) $r = 11$ **(b)** $r = 12$ **(c)** $r = 13$

Figure 3.5.: JC for different functions α and number of regions r ($\mathcal{P}_1$, Ex. A)

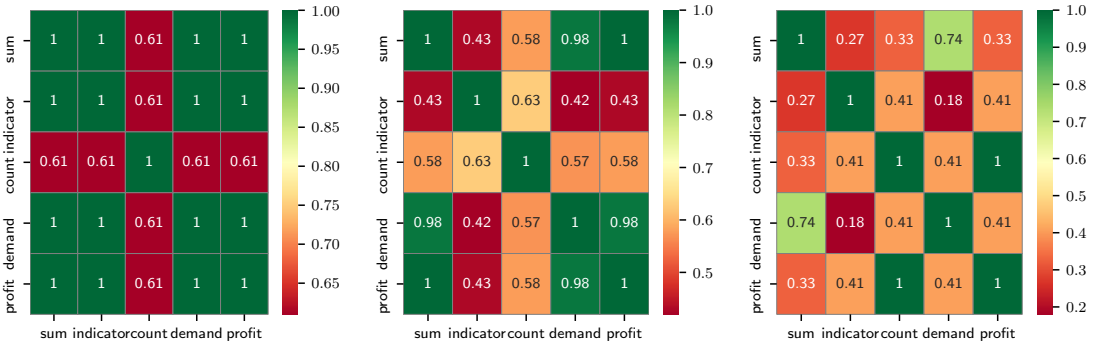
Figure 3.6 depicts the service regions for $\mathcal{P}_2$ derived from $A^{indicator}$, A^{count}, and A^{profit} for $r = 2$ and $r = 3$. Figure 3.7 depicts the associated matrices with the Jaccard coefficients. For $r = 2$, all aggregation functions except α^{count} produce identical service regions as indicated by Jaccard coefficients of 1.0. As Figure 3.6b illustrates, this consistent separation into two service regions distinguishes between the two subsets of interdependent facilities already identified in Chapter 2. Meanwhile, the service regions produced by α^{count} group facilities from the same interdependent subset into two different service regions (Figure 3.6a). For $r = 3$, A^{sum}, A^{demand}, and A^{profit} still produce very consistent

service regions with pairwise Jaccard coefficients close to or equal to 1.0 (Figure 3.7b). The resulting service regions still group together the previously identified subsets of interdependent location decisions (Figure 3.6d). For $r = 4$, we see that the service regions vary significantly. This indicates that no strong pattern inducing 4 different service regions exists in $\mathcal{S}_{I\setminus i}$ for $\mathcal{P}_2$.



(a) A^{count}, $r = 2$ **(b)** A^{profit}, $r = 2$ **(c)** $A^{indicator}$, $r = 3$ **(d)** A^{profit}, $r = 3$

Figure 3.6.: Service regions for different functions α and number of regions r ($\mathcal{P}_2$, Ex. A)



(a) $r = 2$ **(b)** $r = 3$ **(c)** $r = 4$

Figure 3.7.: JC for different functions α and number of regions r ($\mathcal{P}_2$, Ex. A)

▲

The aggregation functions α^{sum}, α^{demand}, and α^{profit} produce similar biclusterings and tend to preserve interdependence structures as determined in Chapter 2 better. However, the biclusterings obtained with different aggregation functions and different numbers of target regions differ significantly. A structured approach to evaluating the quality of service regions is needed.

3.2.2. Evaluation criteria

Many validation metrics for biclusterings exist in literature. Our application differs from problems in literature in that we are looking for coherent but not necessarily homogeneous substructures. We want to group together customers served jointly from the same set of interdependent facilities rather than customers that exhibit similar demand volumes or generate similar profits. This implies that we are not looking for high intra-cluster

similarity and inter-cluster dissimilarity. Instead, we are looking for high intra-cluster interaction and low inter-cluster interaction. We capture this idea in the following metrics.

3.2.2.1. Intra-region validity

To measure the intra-region validity of a particular region, we consider all facilities in that region and determine the proportion of customers allocated to any of these facilities that are not grouped within that region. This proportion represents the degree to which the opening decisions within that region depend on customers outside that region. We already counted this type of allocations in Example E and referred to them as out-of-bicluster allocation. We propose the following loss function that considers the "worst" intra-region validity across all regions evaluated by the proportion of the total edge weights in the bipartite graph that are incident with two nodes assigned to different regions.

Definition 3.6. *The **intra-regional loss** of a biclustering $\mathcal{R}$, $\ell_\alpha^{intra}(\mathcal{R})$, denotes the maximum proportion of out-of-region allocations across all regions such that*

$$\ell_\alpha^{intra}(\mathcal{R}) = \max_{k\in\{1,\dots,r\}} \frac{\sum_{i\in I_k}\sum_{j\in J^{\mathcal{S}}\setminus J_k} a_{ij}}{\sum_{i\in I_k}\sum_{j\in J^{\mathcal{S}}} a_{ij}}. \tag{3.9}$$

The value a_{ij} depends on the aggregation function α (see Table 3.1). Evidently, $\ell_\alpha^{intra}(\mathcal{R})$ will be 0 when $|\mathcal{R}| = r = 1$ and have a tendency to increase with increasing r.

3.2.2.2. Inter-region validity

With the intra-regional loss function, we focus on the coherence within individual regions. To evaluate the validity of the biclustering as a whole, we propose the following inter-regional loss function that sets the sum of the inter-regional edge weights in relation to the total edge weights in the bipartite graph.

Definition 3.7. *The **inter-regional loss** of a biclustering $\mathcal{R}$, $\ell_\alpha^{inter}(\mathcal{R})$, denotes the total proportion of out-of-region allocations*

$$\ell_\alpha^{inter}(\mathcal{R}) := \frac{\sum_{k=1}^{r}\sum_{j\in J_k}\sum_{i\in I^{\mathcal{S}}\setminus I_k} a_{ij}}{\sum_{j\in J^{\mathcal{S}}}\sum_{i\in I^{\mathcal{S}}} a_{ij}} = \frac{\sum_{k=1}^{r}\sum_{i\in I_k}\sum_{j\in J^{\mathcal{S}}\setminus J_k} a_{ij}}{\sum_{j\in J^{\mathcal{S}}}\sum_{i\in I^{\mathcal{S}}} a_{ij}}. \tag{3.10}$$

The loss function $\ell_\alpha^{inter}(\mathcal{R})$ yields the same result whether one considers the inter-regional edges as those edges connecting facilities to outside customers or customers to outside facilities. This is in contrast to the intra-regional similarity, which assigns edges to the region of the associated facility.

Obviously, $\ell_\alpha^{inter}(\mathcal{R}) = 0$ if $|\mathcal{R}| = r = 1$. When the bipartite graph induced by A^α has k connected components, then $\ell_\alpha^{inter}(\mathcal{R})$ will remain 0 for $|\mathcal{R}| = r \leq k$, $r \in \mathbb{N}$. For $r > k$, $\ell_\alpha^{inter}(\mathcal{R})$ will increase with increasing r. However, this increase is not necessarily monotonic as $\ell_\alpha^{inter}(\mathcal{R})$ only quantifies the inter-regional similarity, which is only part of the objective function of the spectral biclustering algorithm.

3.2.2.3. External region validation

External validation of the resulting biclustering is difficult as there is nothing like true division into service regions. However, once we split the set of candidates and customers into distinct regions based on a particular set of solutions $\mathcal{S}'$, we can evaluate the degree to which these regions reflect the partition implied by other sets $\mathcal{S}$.

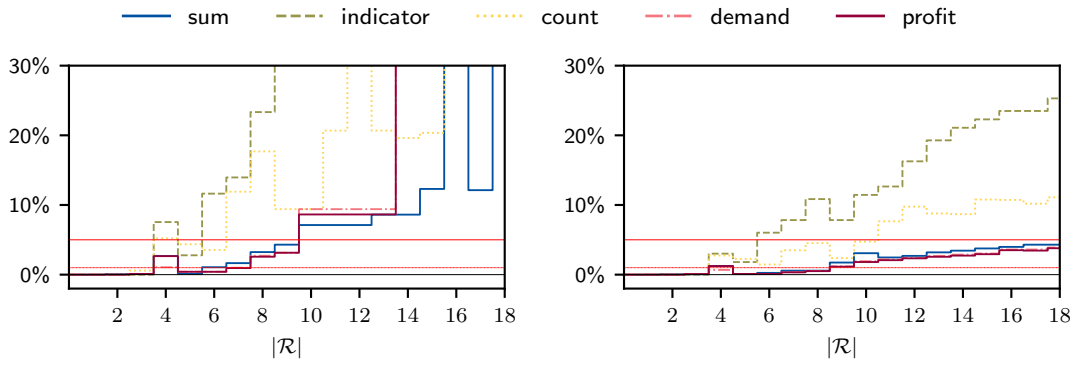
Definition 3.8. *Given a biclustering $\mathcal{R}$ and a set of solutions $\mathcal{S}$, the* ***external loss****, $\ell_\alpha^{external}(\mathcal{R}, \mathcal{S})$, evaluates the degree to which the regions in $\mathcal{R}$ reflect the implicit division into service regions in the solutions in $\mathcal{S}$. It is defined as the proportion of out-of-cluster allocations in $\mathcal{S}$ such that*

$$\ell_\alpha^{external}(\mathcal{R}, \mathcal{S}) = \frac{\sum_{r=1}^{|\mathcal{R}|} \sum_{i \in I_r} \sum_{j \in J_r} a_{ij}}{\sum_{i \in I} \sum_{j \in J} a_{ij}}. \tag{3.11}$$

Thereby, a_{ij} are the entries of the feature matrix A^α derived from $\mathcal{S}$ according to the aggregation procedure α. $\ell_\alpha^{external}(\mathcal{R}, \mathcal{S})$ determines the proportion of the weighted intra-regional allocations in $\mathcal{S}$.

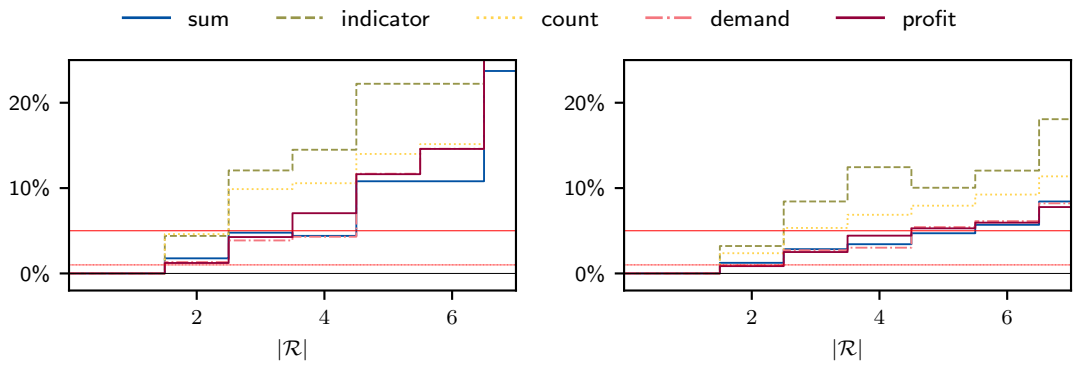
Example A 3.2 (Evaluation criteria for service regions) Figure 3.8 and Figure 3.9 depict the development of the internal validation metrics, ℓ_α^{inter} and ℓ_α^{intra}, for an increasing number of regions derived and evaluated with different aggregation functions for $\mathcal{P}_1$ and $\mathcal{P}_2$, respectively. The red horizontal lines indicate 5% or 1%, respectively. The aggregation functions $\alpha^{indicator}$ and α^{count} result in service regions whose validation metrics deteriorate with the lowest number of biclusters. Not only the produced biclusterings but also their validation metrics are nearly equal for aggregation functions α^{sum}, α^{demand}, and α^{profit}.

Furthermore, one can see that the intra-cluster loss derived from the "worst" cluster deteriorates much faster than the inter-cluster loss. For example, for $\mathcal{P}_2$ it illustrates that for more than 3 biclusters, the intra-cluster loss exceeds 5% for all aggregation functions, indicating no meaningful separation into more than three distinct regions.



(a) $\ell_\alpha^{intra}(\mathcal{R})$ **(b)** $\ell_\alpha^{inter}(\mathcal{R})$

Figure 3.8.: Development of validation metrics with an increasing number of regions determined based on different functions α ($\mathcal{P}_1$, Ex. A)



(a) $\ell_\alpha^{intra}(\mathcal{R})$ **(b)** $\ell_\alpha^{inter}(\mathcal{R})$

Figure 3.9.: Development of validation metrics with an increasing number of regions determined based on different functions α ($\mathcal{P}_2$, Ex. A)

▲

For both instances, the validation metrics deteriorate with an increasing number of regions. This deterioration, however, is not monotone. As there is no inherent measure of intra-cluster similarity, our measures all target low inter-cluster interaction. However, the underlying algorithm also aims for non-zero biclusterings of balanced cardinalities. This is necessary to avoid producing several biclusterings with only one entry, yet may produce biclusterings that perform differently regarding the presented metrics. This raises the question of how to determine the appropriate number of biclusterings.

3.2.3. Determining the number of service regions

There is no true number of service regions. Interdependent subsets of facilities often overlap, and the number of regions depends on the targeted level of coherence. The interaction between distinct regions will likely increase with an increasing number of regions. Thus, aiming for a target loss of 0 might prohibit one from identifying regions of strong but not perfect coherence.

We propose determining the number of regions by setting an upper bound on the loss value of $\theta \in [0, 1]$. Given a sequence of biclusterings $\mathcal{R}_k$ with $k \in \{2, \dots, r\}$ and $|\mathcal{R}_k| = k$, we choose the biclustering which contains the largest number of regions while maintaining $\ell_\alpha^{inter}(\mathcal{R})$ and $\ell_\alpha^{intra}(\mathcal{R})$ below θ.

We introduce the algorithm *RegClus*, which produces the associated biclustering, given a target loss θ. It is described in pseudo-code in Algorithm 2. It requires a set of solutions $\mathcal{S}$, the set of instance parameters $\mathcal{P}$, the aggregation function α mapping the information from $\mathcal{S}$ and $\mathcal{P}$ to a feature matrix A^α, and an upper bound on the loss θ. The algorithm chooses the biclustering with the largest number of regions whose corresponding loss values do not exceed θ.We refer to the resulting service regions as $\mathcal{R}(\mathcal{S}, \alpha, \theta)$.

Example A 3.3 (Regional clustering algorithm) We use *RegClus* to determine service regions instances for different α and $\theta = 0.05$ for sets of k best alternative solutions

Algorithm 2 *RegClus* - Regional Clustering Algorithm

Input: $\mathcal{S}$, $\mathcal{P}$, α, θ
Output: $\mathcal{R}$
1: $A \leftarrow \alpha(\mathcal{S}, \mathcal{P})$
2: $r^{max} \leftarrow \max_{s \in \mathcal{S}} \sum_{i \in I} y_i^s$
3: **for** $r \leftarrow 2$ to r^{max} **do**
4: $\quad \mathcal{R}_r \leftarrow Spectral\ Biclustering(A, r)$ ▷ see Algorithm 1
5: **end for**
6: $\mathcal{R} \leftarrow \arg\max_{\mathcal{R}_r, r \in \{2,\dots,r^{max}} \left\{ r \mid \ell_\alpha^{inter}(\mathcal{R}_r) \leq \theta, \ell_\alpha^{intra}(\mathcal{R}_r) \leq \theta \right\}$
7: **return** $\mathcal{R}$

and sets of solutions to instances with perturbed demand. In particular, we derive the service regions using *RegClus* from the set of 5 best alternative solutions $\mathcal{S}^5_{k-best}$. We then determine the number of regions $|\mathcal{R}|$ and evaluate the degree to which these service regions also reflect the allocations observable in the set of 20 best solutions $\mathcal{S}^{20}_{k-best}$ using $\ell_\alpha^{external}$. The results for instances $\mathcal{P}_1$-$\mathcal{P}_4$ from Example A are depicted in Table 3.2.

		α^{sum}	$\alpha^{indicator}$	α^{count}	α^{demand}	α^{profit}
$\mathcal{P}_1$	$\lvert\mathcal{R}\rvert$	12	6	9	12	12
	$\ell_\alpha^{external}\left(\mathcal{R}(\mathcal{S}^5_{k-best}, \alpha, 0.05), \mathcal{S}^{20}_{k-best}\right)$	0.96	0.75	0.95	0.96	0.96
$\mathcal{P}_2$	$\lvert\mathcal{R}\rvert$	2	1	1	2	2
	$\ell_\alpha^{external}\left(\mathcal{R}(\mathcal{S}^5_{k-best}, \alpha, 0.05), \mathcal{S}^{20}_{k-best}\right)$	0.85	0.62	0.87	0.85	0.86
$\mathcal{P}_3$	$\lvert\mathcal{R}\rvert$	7	3	3	11	13
	$\ell_\alpha^{external}\left(\mathcal{R}(\mathcal{S}^5_{k-best}, \alpha, 0.05), \mathcal{S}^{20}_{k-best}\right)$	0.95	0.93	0.95	0.95	0.95
$\mathcal{P}_4$	$\lvert\mathcal{R}\rvert$	1	2	1	2	2
	$\ell_\alpha^{external}\left(\mathcal{R}(\mathcal{S}^5_{k-best}, \alpha, 0.05), \mathcal{S}^{20}_{k-best}\right)$	0.94	0.72	0.92	0.93	0.93

Table 3.2.: Persistence of service regions in k best alternative solutions with different α ($\mathcal{P}_1$-$\mathcal{P}_4$, Ex. A)

The results show that for instances $\mathcal{P}_1$ and $\mathcal{P}_3$, which are mostly composed of independent facilities that serve small subsets of customers, *RegClus* identifies relatively many regions, particularly when the aggregation functions α^{sum}, α^{demand}, or α^{profit} are used. Based on the 5 best solutions and α^{profit}, 12 distinct regions are determined for $\mathcal{P}_1$, and 13 are determined for $\mathcal{P}_3$. These regions persist throughout the 20 best solutions as the external validation criterion is 0.96 and 0.95, respectively. For $\mathcal{P}_2$ and $\mathcal{P}_4$, at most two distinct regions are identified by *RegClus*. While for $\mathcal{P}_4$ and α^{profit}, the external validation criterion is high with 0.93, the separation of $\mathcal{P}_2$ into two distinct regions only yields an external validation of 0.86.

We perform a similar experiment considering sets of solutions derived for problem instances with perturbed demand. Every solution in this set $\mathcal{S}^p_{pert}$ is the optimal solution to a modified problem instance in which the demand of every customer has been perturbed by a defined perturbation level $p \in [0, 1]$ such that

$$D_j^{new} = D_j + (1 + p \cdot b - (1 - b) \cdot p), \qquad \forall j \in J, \tag{3.12}$$

with $b \sim B(0.5)$ a binary random variable following a Bernoulli distribution with parameter 0.5. Again, we derive sets of 20 such solutions for each instance. We then determine the

service regions based on a subset $\mathcal{S}'^{p}_{pert} \subset \mathcal{S}^{p}_{pert}$ of 5 randomly chosen solutions from this set and subsequently evaluate the degree to which these regions persist throughout the remaining solutions in that set. The results are displayed in Table 3.3.

		α^{sum}	$\alpha^{indicator}$	α^{count}	α^{demand}	α^{profit}
$\mathcal{P}_1$	$\lvert\mathcal{R}\rvert$	10	4	5	9	11
	$\ell_\alpha^{external}\left(\mathcal{R}(\mathcal{S}^{5\%}_{pert}, \alpha, 0.05), \mathcal{S}^{5\%}_{pert}\right)$	0.98	0.92	0.98	0.98	0.98
$\mathcal{P}_2$	$\lvert\mathcal{R}\rvert$	4	2	2	3	4
	$\ell_\alpha^{external}\left(\mathcal{R}(\mathcal{S}^{5\%}_{pert}, \alpha, 0.05), \mathcal{S}^{5\%}_{pert}\right)$	0.82	0.63	0.83	0.83	0.83
$\mathcal{P}_3$	$\lvert\mathcal{R}\rvert$	9	2	5	5	9
	$\ell_\alpha^{external}\left(\mathcal{R}(\mathcal{S}^{5\%}_{pert}, \alpha, 0.05), \mathcal{S}^{5\%}_{pert}\right)$	0.97	0.97	0.97	0.95	0.97
$\mathcal{P}_4$	$\lvert\mathcal{R}\rvert$	3	3	4	4	4
	$\ell_\alpha^{external}\left(\mathcal{R}(\mathcal{S}^{5\%}_{pert}, \alpha, 0.05), \mathcal{S}^{5\%}_{pert}\right)$	0.95	0.79	0.93	0.93	0.93

Table 3.3.: Persistence of service regions in solutions to instances with perturbed demand for different α ($\mathcal{P}_1$-$\mathcal{P}_4$, Ex. A)

Compared to the previous experiment based on sets of k best alternative solutions, we obtain more service regions for individual instances for instances $\mathcal{P}_2$ and $\mathcal{P}_4$. Furthermore, these regions have a higher validity in terms of $\ell_\alpha^{external}$. ▲

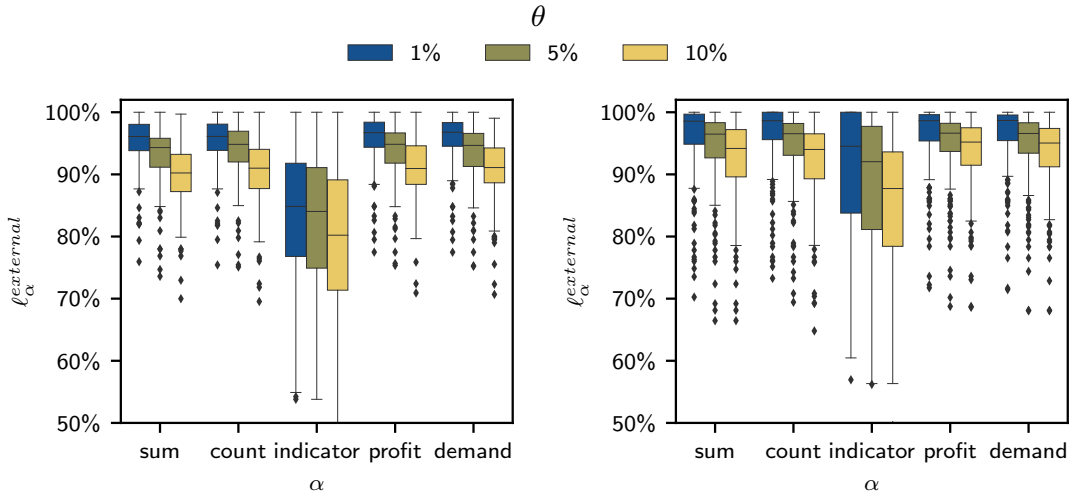
Example A 3.3 has shown that *RegClus* allows to derive regions from subsets of solutions that behold a relatively high validity throughout other, similar solutions. Which value of $\ell_\alpha^{external}(\mathcal{R}, \mathcal{S})$ is sufficient to validate the existence of the determined service regions depends on the purpose with which these regions are extracted. We will take a closer look at this in Chapter 4. Nevertheless, when there are regions of stronger coherence among multi-sets of solutions, *RegClus* is a suitable means to detect them.

3.3. Experimental validation: characteristic service regions in sets of solutions

We repeat the experiments from Example A 3.3 for all instances from data sets presented in Section 1.3. In particular, we generate sets of well-performing solutions $\mathcal{S}$ to each instance and use *RegClus* to derive service regions $\mathcal{R}$ based on a subset $\mathcal{S}' \subset \mathcal{S}$ of these solutions. We then evaluate the number of obtained regions and the degree to which the obtained regions reflect the service regions underlying solutions in the entire set $\mathcal{S}$ using $\ell_\alpha^{external}(\mathcal{R}, \mathcal{S})$. Again, we look at sets of k best alternative solutions and sets of solutions with slightly perturbed customer demands.

Figure 3.10 presents the distribution of the resulting external validation metric $\ell_\alpha^{external}(\mathcal{R}, \mathcal{S})$ for different aggregation functions α and different upper target levels of coherence θ. Figure 3.10a displays the results obtained for sets of k best alternative solutions. Figure 3.10b displays results for sets of solutions to instances with 5% demand perturbation. The average external validation and the average number of regions are displayed in Table 3.4. The following observations are noteworthy. Both α and θ have a significant effect. The larger θ, the larger the upper bound on the accepted internal validation metrics. Thus, it is unsurprising that with increasing θ, the average number of regions increases and the validity

decreases. Meanwhile, a difference between the validity of service regions obtained with different aggregation functions is barely noticeable in Figure 3.10. With the exception of $\alpha^{indicator}$, which produces the lowest validation metrics, all aggregation functions perform equally well. However, a look at the values in Table 3.4 shows that α^{profit} produces the largest number of regions with the highest external validation score on average. In general, for up to $\theta = 5\%$, the average validation scores exceed 90%, indicating a high persistence of service regions throughout the considered sets and a solid performance of *RegClus* in their identification.



(a) $\mathcal{S}^{20}_{k-best}$ **(b)** $\mathcal{S}^{5}_{k-best}$

Figure 3.10.: Distribution of $\ell_\alpha^{external}$ for different α and θ

θ	α	$\mathcal{S}^{20}_{k-best}$ avg. $\lvert\mathcal{R}\rvert$	$\mathcal{S}^{20}_{k-best}$ avg. $\ell_\alpha^{external}$	$\mathcal{S}^{5}_{k-best}$ avg. $\lvert\mathcal{R}\rvert$	$\mathcal{S}^{5}_{k-best}$ avg. $\ell_\alpha^{external}$
1%	sum	2.2	95.3	3.8	96.0
	count	2.0	95.4	3.0	96.5
	demand	2.4	95.7	3.8	96.4
	profit	2.4	95.8	3.9	96.5
5%	sum	3.5	93.2	5.9	94.3
	count	2.8	93.9	4.2	94.7
	demand	3.8	93.5	6.0	94.7
	profit	3.8	93.7	6.2	94.7
10%	sum	5.2	89.9	7.3	92.6
	count	4.0	90.4	5.4	92.3
	demand	5.2	90.8	7.4	93.4
	profit	5.3	91.0	7.6	93.5

Table 3.4.: Average number of regions and $\ell_\alpha^{external}$ for different α and θ

3.4. Conclusion

We use spectral biclustering to identify persistent characteristic decision patterns in the form of coherent service regions in multi-sets of solutions to the CFLP. Thereby, the proposed procedure can be applied to other location-allocation problems straightforwardly. To the best of our knowledge, this is the first work exploring the identification of persistent

structures amongst sets of solutions that not only considers the location decisions in the form of a frequency count of individual facilities but includes information on the allocation decisions. An experiment on a large number of instances from location literature demonstrates that given a set of well-performing solutions to a particular instance or variants of this instance, it is possible to identify regions of strong coherence in the sense that a significant proportion of the allocations throughout similar solutions occur within that region.

Furthermore, to the best of our knowledge, this is the first approach using pattern recognition techniques to detect parts of the decision space that exhibit strong interdependencies. It is not surprising that customers are assigned to facilities that are close in the sense that the variable transport costs are low. This induces a regional structure of regions of customers that are served jointly by a subset of facilities. This notion of regionality has not been explored before.

Recall that while the detected regions are visibly coherent, in the presented visualizations of the instances and solutions, the procedure does not require any actual locations in terms of coordinates or even rely on a persistent metric from which transport costs are derived. We summarize the main results of the previous chapter as follows:

- Persistent service regions of a defined target level of coherence can be identified from arbitrary sets of solutions using pattern recognition techniques.
- Spectral biclustering simultaneously divides the subsets of facilities and customers into service regions based on a feature matrix that aggregates the allocation information in a given set of solutions.
- Distinct service regions of extreme coherence can be identified in all considered instances in sets of k best solutions or sets of optimal solutions to slightly perturbed instances. This indicates the persistence of these service regions as an abstract decision pattern induced by the problem's input data.

Up to this point, service regions are derived from several optimal or near-optimal solutions to the original instance or modifications thereof. It is an ex-post analysis tool that allows for additional insights after an optimal or several well-performing solutions have already been found. In Chapter 4, we take a closer look at whether these service regions can be detected ex-ante, that is before the problem is solved, either from the problem data or easy-to-obtain, potentially infeasible solutions.

4. Anticipating service regions

Up to this point, we derived service regions from sets of optimal and near-optimal solutions. Thereby, we reduced the insights these service regions yield as an ex-post analysis to gain contextual information on the individual decisions in an optimal solution. However, as service regions induce a separation of the facility-customer space into distinct regions with low interaction levels, service regions bear the potential to facilitate the search for an optimal solution in a divide-and-conquer approach. Furthermore, the question arises whether these service regions directly result from the spatial patterns underlying a problem instance that can be derived directly from the data. Thus, in the following chapter, we address the following research question.

RQ3: Can service regions be anticipated from the problem's input data?

As service regions are a concept newly introduced in this thesis, no related work other than those linking data and decisions presented in Section 2.1 exists. In Section 4.1, we present characteristics that summarize properties of the problem instance and relate them to the level of interdependence in the optimal solution. While the effect of marginal changes in individual characteristics can be shown, the level of interdependence on unseen instances cannot be anticipated due to many largely overlying influencing factors. In Section 4.2, we then turn to another approach to derive well-performing service regions without knowledge of the actual well-performing solutions. More precisely, we derive service regions from sets of solutions to the linearly relaxed problem. These are easier to obtain but still yield insights into the implied spatial relationships between candidates and customers. We summarize the main findings in Section 4.3.

4.1. Anticipating service regions from instance characteristics

Service regions are a decision pattern persisting in well-performing solutions to a particular instance of the CFLP that implicitly separate the facility-customer space. Existing works examining the decision space of CFLP instances suggest that underlying spatial patterns induce characteristics of the problems' solutions. The parameters of the CFLP include spatial patterns only implicitly in the variable costs or profits between customers and facilities. The degree to which the variation in these parameters is responsible for specific characteristics of well-performing solutions also depends on their relative importance compared to the facilities' fixed costs and capacities.

In Subsection 4.1.1, we present characteristics describing critical properties of the problem instance. We compute these properties for the data sets from Section 1.3 and show that these instances are heterogeneous in various regards. In Subsection 4.1.2, we discuss how, all other conditions being equal, changes in these properties can be expected to affect the size of service regions and, thereby, the interdependence of individual facilities. We use instances $\mathcal{P}_1$ and $\mathcal{P}_2$ from Example A to validate our hypotheses. In Subsection 4.1.3, we transfer previous findings to the instances from Section 1.3.

4.1.1. Characteristic properties of problem instances for the CFLP

We introduce several characteristics that summarize the properties of problem instances to the CFLP. We subsequently determine these characteristics for the instances described in Section 1.3.

4.1.1.1. Parameter ratios

Parameters in the CFLP are measured in two different units: monetary units for costs and profits and quantities for capacities, demands, or transport volumes. The absolute values of individual parameters are less distinctive than the ratios between parameters measured in the same unit. Thus, to distinguish different problem instances, one can compare the ratio between the total capacity and demand in the instance [Klose and Görtz, 2007].

Definition 4.1. *The **tightness** sets the sum of all capacity in relation to the sum of all demand in the problem instance such that*

$$tightness := \frac{\sum_i Q_i}{\sum_j D_j}. \tag{4.1}$$

The inverse of the tightness can serve as a proxy for the fraction of candidate facilities that need to operate to have sufficient capacity to serve all demands.

The ratio between fixed and variable profits affects the degree to which a facility needs to be utilized to be profitable and, therefore, also impacts the number of candidate facilities operating in the optimal solution.

Definition 4.2. *The **profit ratio** sets the average variable unit profit, UVP, minus the average variable unit transportation costs, UVC, the average net unit profit, in relation to the average unit fixed costs, UFC, such that*

$$profitratio := \frac{UVP - UVC}{UFC}, \text{ with} \tag{4.2}$$

$$UFC := \frac{\sum_i F_i}{\sum_i Q_i}, \tag{4.3}$$

$$UVC := \frac{\sum_{i,j} D_j c_{ij}}{|I| \cdot \sum_j D_j}, \text{ and} \tag{4.4}$$

$$UVP := \frac{\sum_j D_j r_j}{\sum_j D_j}. \tag{4.5}$$

The profit ratio can serve as a proxy indicating whether fixed costs are more relevant in the sense that an optimal or near-optimal solution will likely maximize the utilization of individual facilities or whether profits are more relevant in the sense that more, less utilized facilities will likely be opened closer to customers.

Figure 4.1 displays the distribution of the above ratios for the sets of instances described in Section 1.3. Figure 4.1a shows that all instances consider at least as many customers as

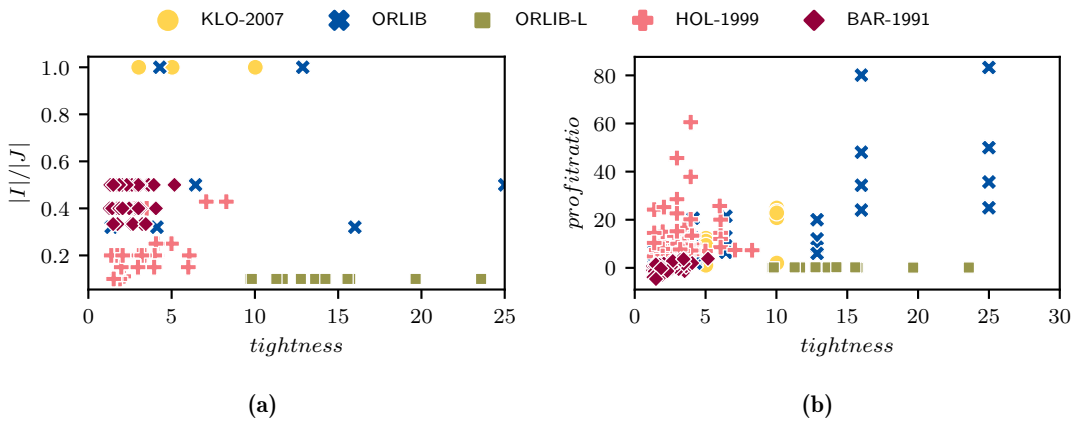


Figure 4.1.: Distribution of parameter ratios for considered test instances

candidate facilities. The tightness differs massively. For instances from the BAR-1991 and HOL-1999 data sets, the tightness is relatively low, indicating that a significant fraction of the candidates will likely operate in an optimal solution. The instances comprise tightness between 1.2 and 275.1, with 210 of the 237 considered instances exhibiting a tightness of less or equal to 10. Similarly, the profit ratio ranges from -4.6 to 83.3, indicating a huge difference in the relative importance of the fixed costs between instances.

4.1.1.2. Spatial patterns

Another interesting aspect that is not trivial to quantify is the spatial distribution of customers and candidates. Most considered instances were generated by randomly distributing candidates and customers on a unit square. It is reasonable to assume that these instances will exhibit different properties than those that exhibit spatial clusters similar to the densely populated east cost area in the ORLIB instances. However, to the best of our knowledge, a systematic evaluation of the spatial distribution of data points in facility location instances has not been done before.

We evaluate the spatial distribution of candidates and customers with the help of the distance of individual points to their nearest neighbor and a two-sided t-test, which tests the hypothesis of complete spatial randomness according to Clark and Evans [1954]. The test measures the degree to which the average distance of any point to its nearest neighbor differs from the expected average distance under spatial randomness, that is, a situation in which "any point has had the same chance of occurring on any sub-area as any other point". This is precisely the situation when coordinates are drawn from a uniform distribution.

Applied to the instances of the CFLP, we consider candidates and customers as data points. Generally, for CFLP instances, we do not have information on the actual location nor the distance between any two candidates or any two customers. Therefore, we use the coordinates we obtain with the help of multi-dimensional scaling (see Section 1.3) as a proxy for the location of candidates and customers in a hypothetical plane. These coordinates represent the entire information we have on their spatial distribution. Subsequently, we determine the Euclidean distances between all pairs of points and take the minimum distance per point as its estimated nearest neighbor distance. We denote the

average estimated nearest neighbor distance by D_o.

The expected distance to the nearest neighbor, D_E, in a randomly dispersed population of N points with specified density, ρ, is

$$D_E = \frac{1}{2\sqrt{\rho}} \tag{4.6}$$

with a standard deviation, σ_E, of

$$\sigma_E = \frac{0.26136}{\sqrt{N\rho}}. \tag{4.7}$$

The density is obtained as the number of points divided by the area. A detailed derivation of the formula developed initially by Hertz [1909] can be found in Clark and Evans [1954] and is based on the determination of the mean number of points in a given area with the help of the Poisson distribution.

Given D_o, D_E, and σ_E, it is possible to perform a two-sided t-test on the hypothesis that D_o equals D_E and thereby the null hypothesis that the mean distance to the nearest neighbor in the observed distribution cannot be distinguished from the mean distance under spatial randomness. This test is also referred to as a test for Complete Spatial Randomness and evaluates the test statistic

$$CSR = \frac{D_o - D_E}{\sigma_E} \tag{4.8}$$

based on the quantiles of the standard normal distribution. This implies that CSR-values whose absolute value exceeds 1.96 (2.58) allow one to reject the null hypothesis of spatial randomness with a confidence level of 95% (99%).

In particular, the test allows one to gain information on which spatial distribution one observes instead of spatial randomness. If the test statistic is negative, then the observed mean distance to the nearest neighbor is smaller than expected under randomness, indicating clusters of points at certain sub-spaces. If the test statistic is positive, the distance is larger than expected, indicating a more evenly dispersed distribution of points.

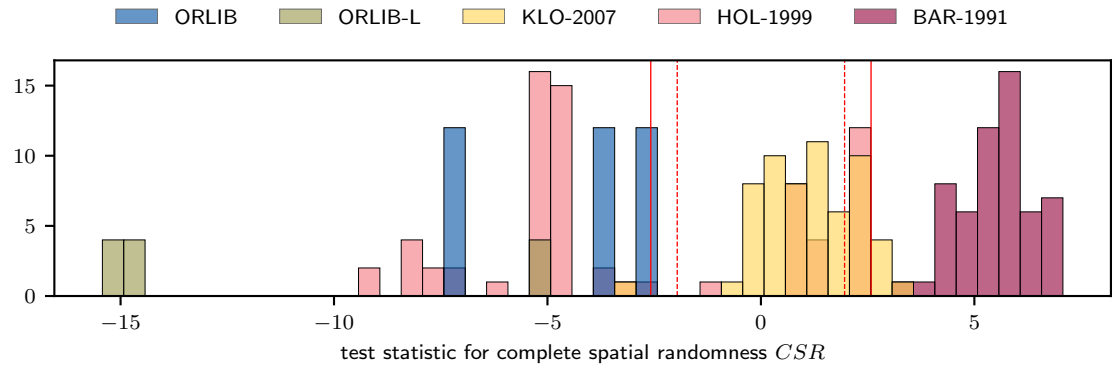


Figure 4.2.: Distribution of test statistics for test for complete spatial randomness for considered test instances

Figure 4.2 depicts the distribution of the test statistics of the instances in the considered

sets. Nearest neighbor distances are derived from the estimated coordinates in the plane. The size of the area is derived from the minimum and maximum estimated x- and y-coordinates of all points. We observe the spatial distribution of points in the individual instances, which range from clustered to evenly dispersed. The null hypothesis cannot be rejected for most KLO-2007 instances generated uniformly randomly. The instances from the ORLIB data set, which was derived from real-world data based in the United States, predominantly exhibit a clustered spatial structure. Meanwhile, the relatively symmetrical-looking data from the BAR-1991 instances (refer to Figure 1.7) are classified as evenly dispersed.

In conclusion, the instances from literature cover a wide range of characteristics regarding their size, spatial distribution, their demand-capacity ratio, as well as their profit-fixed costs ratio. However, even though the considered indicators only represent a few highly aggregated characteristics, the 206 instances do not contain an exhaustive set of all combinations of characteristics. Nor do they explicitly represent extreme cases. Therefore, to derive insights into the relationship between individual characteristics and the underlying service regions, we systematically vary the data of the two representative problem instances $\mathcal{P}_1$ and $\mathcal{P}_2$ from Example A.

4.1.2. Characteristic indicators of large service regions

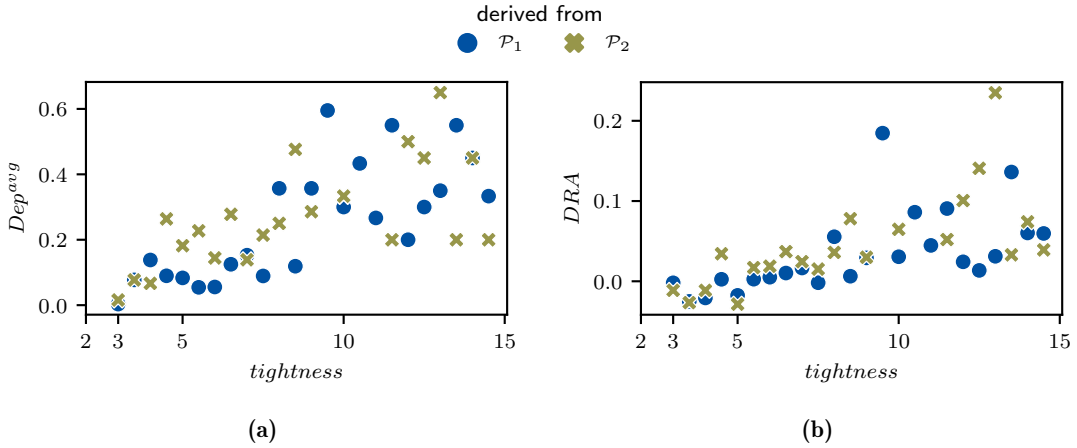
Service regions comprise subsets of customers and facilities in the sense that customers in that region will likely be served by a facility in that region in well-performing solutions. We observed that facilities serving customers in the same service region often interdepend. Thus, instances whose underlying service regions are large tend to have a high average dependence density (see Def. 2.13) and a high level of average demand reallocation (see Def. 2.22). Therefore, in the subsequent section, we first focus on whether we can link properties of the input data to these two measures that indicate large service regions.

In the following, we systematically vary characteristic properties of instances $\mathcal{P}_1$ and $\mathcal{P}_2$ from Example A. We rationalize how these changes affect the potential subset of customers a facility would serve if it was the sole operating facility.

Capacities / Tightness

Increasing the capacity of individual instances allows for serving more demand and, thereby, more customers from a single facility. Consequently, the potential subsets of customers served by individual candidates are more likely to overlap. Overlapping potential subsets of customers can be expected to be more likely to induce interdependencies between the respective facilities.

Example A 4.1 (Effect of increasing tightness on interdependence relationships) We successively increase the capacity of all facilities in problem instances $\mathcal{P}_1$ and $\mathcal{P}_2$ from Example A by multiplying the capacity of each candidate by a constant factor. Subsequently, we determine the resulting tightness in the associated instance and round it to the nearest integer.



(a) (b)

Figure 4.3.: Increasing tightness versus the level of interdependence ($\mathcal{P}_1$-$\mathcal{P}_2$, Ex. A)

Figure 4.3 depicts the distribution of indicators for larger service regions in the optimal solutions for the resulting problem instances indicated by the average dependence density, Dep^{avg}, (Figure 4.3a), and the demand reallocation, DRA (Figure 4.3b). For both instances, these indicators have a clear tendency to increase with increasing capacities. ▲

Fixed costs / Profit ratio

All other conditions being equal, increasing the fixed costs of individual facilities leads to a higher utilization of individual facilities. This means that to operate, facilities must serve more customers. Their associated subsets of customers expand and are more likely to overlap with those of other facilities. This leads us to expect an increasing level of interdependence between facilities operating in the optimal solutions to the resulting problem instances.

Example A 4.2 (Effect of increasing the profit ratio on interdependence relationships) We successively increase the fixed costs of all facilities in problem instances $\mathcal{P}_1$ and $\mathcal{P}_2$ by multiplying the fixed costs of each candidate by a constant factor, which, in turn, lead to decreasing profit ratios. The distribution of Dep^{avg} and the DRA in the optimal solutions of the resulting problem instances are depicted in Figure 4.4.

An increase in the profit ratio implies a decrease in the relevance of the fixed costs. This implies that opening more, less utilized facilities serving smaller subsets of customers is more profitable. According to the above rationale, this will result in fewer overlaps of the potential subset of customers served by different facilities and, hence, fewer interdependence relationships in the resulting solutions. This effect, however, is only visible for problem instances derived from $\mathcal{P}_2$, while for $\mathcal{P}_1$, profit ratios of less than two result in a decline of both, the Dep^{avg} and DRA. Due to the low capacities of individual facilities in $\mathcal{P}_1$, low profit ratios and the resulting high fixed costs result in only very few facilities being profitable in the optimal solutions. The values of the above indicators represent instances in which it is only profitable to operate less or equal to 2 facilities, which per se allows for little interaction.

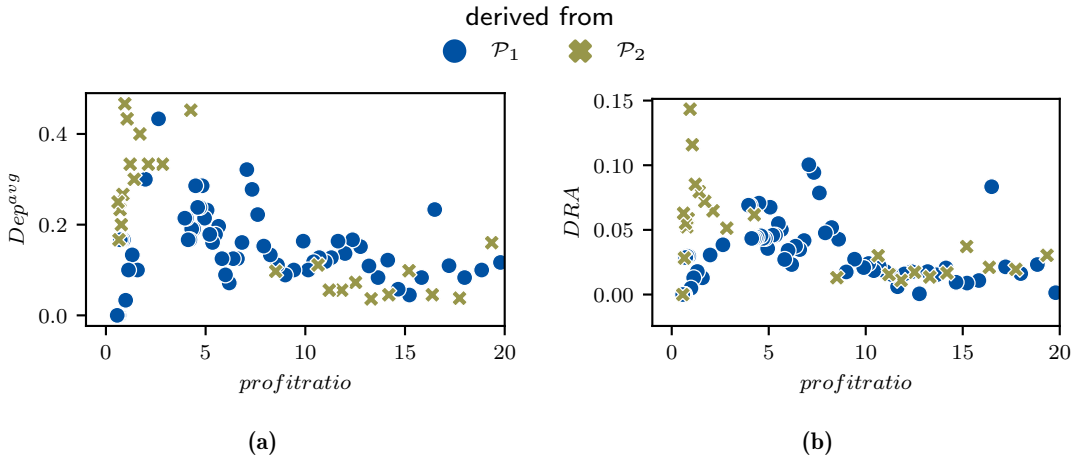


Figure 4.4.: Increasing profit ratio versus the level of interdependence ($\mathcal{P}_1$-$\mathcal{P}_2$, Ex. A)

▲

Spatial patterns

It is intuitive to assume that the spatial patterns underlying the problem instance affect the degree to which combinatorial dynamics affect the optimal solution. In particular, one expects independently operating facilities in clustered instances that exhibit a strong regional grouping and instances with highly combinatorial decision patterns to exhibit a high level of symmetry as is, e.g., the case in an evenly distribution spatial pattern. Linking the occurrence of combinatorial decision patterns to the existence of implied service regions already indicates that the size of individual clusters compared to the capacity of individual facilities will affect this relationship.

Example A 4.3 (Effect of spatial patterns on interdependence relationships) We systematically alter the spatial distribution of customers and candidates of instances $\mathcal{P}_1$ and $\mathcal{P}_2$ and subsequently regenerate variable transport costs and unit profits based on the new Euclidean distances according to the original generation procedure.

Figure 4.5 illustrates the systematic transformation from an evenly distributed spatial pattern (Figure 4.5a) to a random pattern (Figure 4.5f) to a clustered pattern (Figure 4.5i). To obtain these patterns, we gradually mix the original, random pattern with either predetermined evenly distributed coordinates or predetermined clustered coordinates representing 5 clusters. The mixing is performed by a linear composition of α times the old coordinates and $(1 - \alpha)$ times the new coordinates. The parameter α is gradually increased from 0 to 1, increasing by 0.05 in each iteration. This leads to a total of 82 instances to be considered in the following experiment. After determining the coordinates, we determine the resulting spatial pattern with the CSR test statistic. According to the test statistic and a confidence level of 1%, we obtained 40 clustered, 35 randomly distributed, and 7 evenly distributed problem instances.

Figure 4.6 shows the distribution of Dep^{avg} and DRA for instances with different spatial patterns. The results demonstrate that facilities in the optimal solution are more likely to

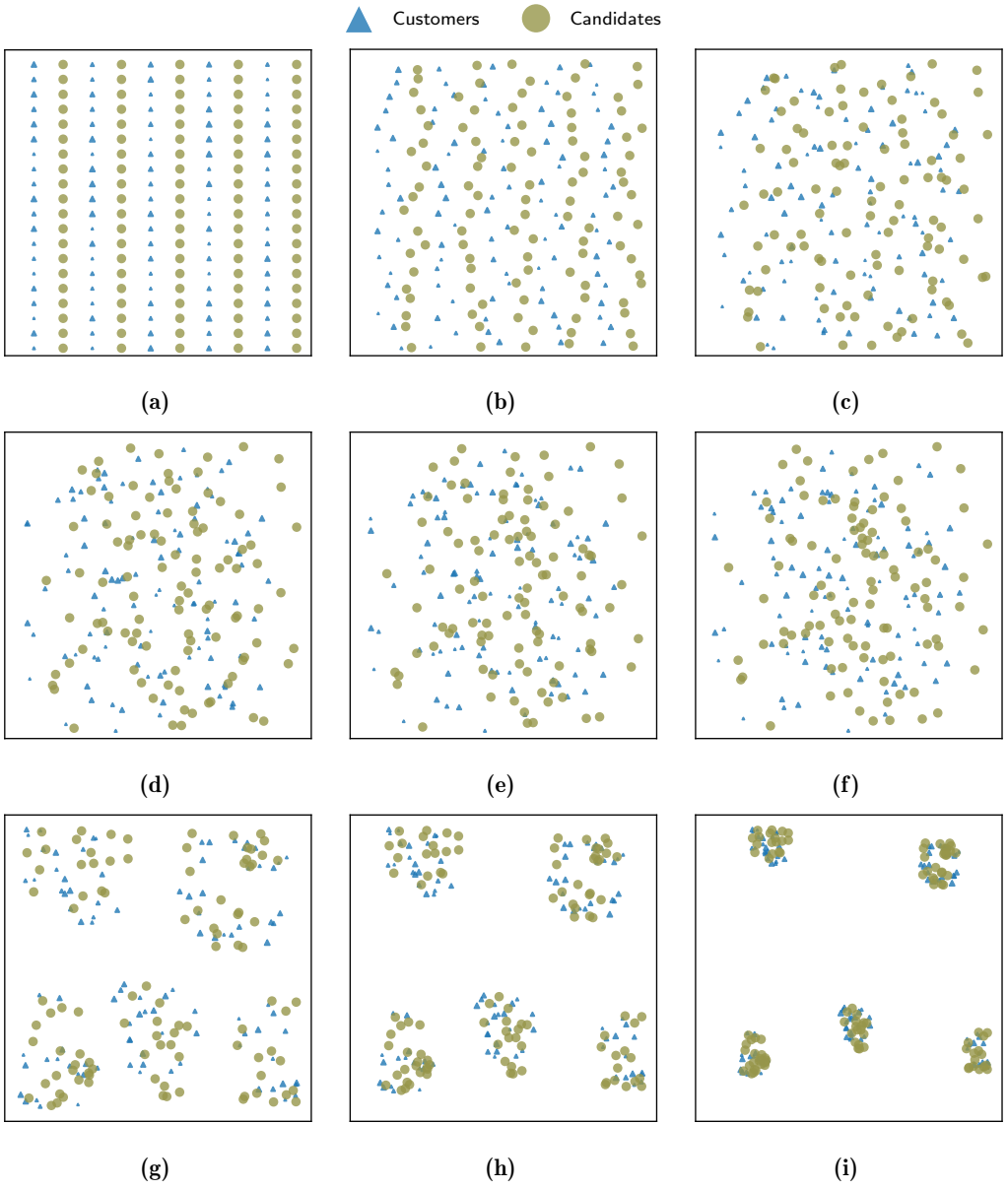
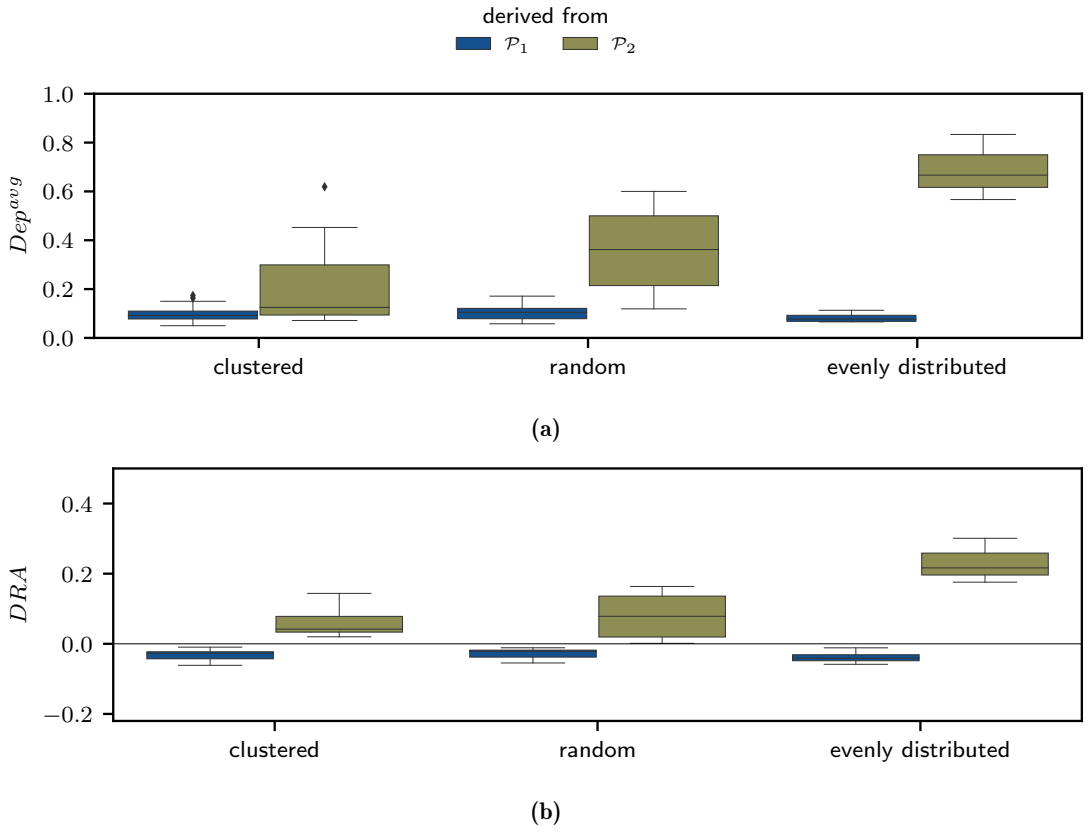


Figure 4.5.: Systematic variation of the spatial pattern ($\mathcal{P}_1$, Ex. A)



(a)

(b)

Figure 4.6.: Spatial pattern and tightness versus the level of interdependence ($\mathcal{P}_1$-$\mathcal{P}_2$, Ex. A)

interdepend, and thus, the solution exhibits larger service regions, when we have an evenly distributed spatial pattern. At the same time, however, we also see that this observation only holds for instances derived from $\mathcal{P}_2$, instances with higher per facility capacities. For instances derived from $\mathcal{P}_1$, the effect of spatial patterns is marginal as the small per-facility capacity leads to the service regions being extremely localized. This leads facilities to be mostly independent regardless of the underlying spatial pattern. ▲

All other conditions being equal, a large capacity-demand ratio, a low profit ratio, and evenly distributed spatial patterns contribute to large service regions implied by high levels of interdependence between facilities operating in the optimal solution. However, we also see that these effects overlap. For example, an instance exhibiting an evenly distributed spatial pattern but a low capacity-demand ratio might still exhibit low levels of interdependence in the optimal solution. In the following, we evaluate whether absolute values for these attributes observed on arbitrary instances suffice to anticipate dependence relationships between individual candidates and, thereby, the size of the underlying service regions.

4.1.3. Experimental validation: linking service regions to instance data

We evaluate whether it is possible to anticipate the level of interdependence of the facilities operating in the optimal solution. Figure 4.7 depicts the distribution of Dep^{avg} for different spatial patterns and varying tightness (Figure 4.7a) and profit ratio (Figure 4.7b), respectively. It illustrates that none of these characteristics alone, or combined with the

underlying spatial patterns, yields sufficient insights into the underlying problem structure to anticipate how combinatorial dynamics will affect the optimal solution. Similar observations have been made for the effect on DRA. In consequence, either combinations of these characteristics lead to overlying effects or relevant characteristics not considered at this point are necessary to anticipate the size of underlying service regions. However, given the fact that these patterns are only inherent in the solution to a particular problem instance and do not necessarily constitute themselves in actual spatial patterns in the form of clusters or other spatial zones, it is likely that they only emerge in combination with constraints and objective of the associated mathematical program. Thus, some information on the structure of well-performing solutions must be provided to identify these patterns.

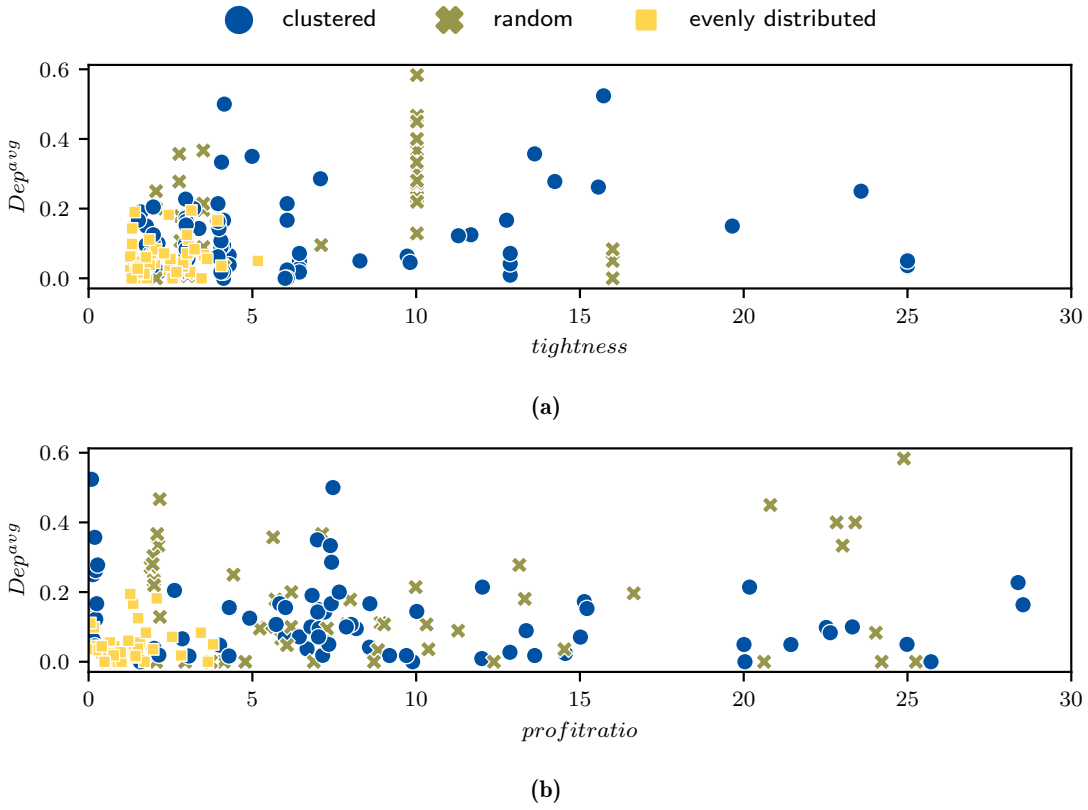


Figure 4.7.: Relationship between instance characteristics and level of interdependence for considered test instances

4.2. Anticipating service regions from integer-infeasible solutions

The experiments in the previous section have shown that overlapping effects of spatial patterns, demand-capacity ratios, and the ratio between fixed costs and variable profits make it challenging to anticipate even the level of interdependence between the facilities operating in an optimal solution. These overlapping effects imply that service regions are not a direct consequence of the spatial patterns underlying the problem instance and, hence, cannot be derived from just these patterns.

Using the aggregated matrices of allocation decisions to well-performing solutions bears the advantage that the effect of capacity, demand, required utilization to offset fixed costs,

etc., are already implicitly included in the allocation decisions. These decisions combine all information from the problem's input data with the objective and constraints of the mathematical program. Thus, in the following, we explore whether service regions that characterize well-performing and, in particular, the optimal solution to a particular problem instance can also be derived from solutions that are easier to obtain. In particular, we look at sets of solutions to linearly relaxed version of the CFLP.

4.2.1. Service regions from sets of linearly relaxed solutions

Service regions group together those customers that are served jointly by facilities in their region. They manifest the spatial relationships between facilities and customers. Service regions appear coherent in the hypothetical, two-dimensional plane derived by MDS. Well-performing solutions will always allocate customers to facilities close to them. How close is close enough, however, depends not only on the spatial relationships but also on, e.g., capacity-demand ratios and demand and capacity distribution. The allocation matrices of a solution to the CFLP comprise all this information and indicate which facility is "close enough" to which customers to serve them profitably. It is reasonable to assume that this information on the implied spatial relationships cannot only be derived from well-performing solutions but also integer-infeasible solutions to the linearly relaxed CFLP. The latter solutions are significantly easier to obtain, particularly for large problem instances. We test this hypothesis with the following experiment.

We consider three sets of solutions to the linearly relaxed CFLP for a given instance. The first set $\mathcal{S}_{opt}^{LR}$ contains only one solution, the optimal solution to the linearly relaxed problem. The second set of solutions, $\mathcal{S}_{I\setminus i}^{LR}$, is derived in a similar manner to $\mathcal{S}_{I\setminus i}$ (refer to Eq. (2.11)). It is composed of the set of solutions obtained when facilities operating in the optimal solution to the linearly relaxed problem are iteratively removed, and the problem is resolved. Thus, let $I^{\star^{LR}}$ denote the index set of facilities for whom $y_i > 0$ in the optimal solution to the linearly relaxed problem, and $s_{I\setminus i}^{\star^{LR}}$ denote the optimal solution to the linearly relaxed problem in which facility i has been removed from the set of candidates. Then, we obtain

$$\mathcal{S}_{I\setminus i}^{LR} := \left\{ s_{I\setminus i}^{\star^{LR}}, i \in I^{\star^{LR}} \right\}. \tag{4.9}$$

The sets $\mathcal{S}_{opt}^{LR}$ and $\mathcal{S}_{I\setminus i}^{LR}$ only assign facilities to service regions that serve customers in at least of the solutions. Thus, we consider a third set, $\mathcal{S}_{I}^{LR}$, of exactly $|I|$ solutions. These are obtained by iteratively fixing $y_i = 1$ for each candidate facility and solving the relaxed CFLP. Let $s_i^{\star^{LR}}$ denote the optimal solution to the linearly relaxed CFLP in which y_i has been fixed to 1, so that we obtain

$$\mathcal{S}_{I}^{LR} := \left\{ s_i^{\star^{LR}}, i \in I \right\}. \tag{4.10}$$

Notice that $\mathcal{S}_{I}^{LR}$ is the only set that is guaranteed to contain some spatial information on every candidate.

We derive service regions from each of the three different sets of linearly relaxed solutions

using *RegClus* (Algorithm 2 from Section 3.2). We then evaluate the degree to which the resulting service regions characterize allocation patterns in well-performing or even optimal solutions to the original problem. Therefore, we derive sets of k best alternative solutions $\mathcal{S}^k_{k-best}$ and determine $\ell^{external}_\alpha(\mathcal{R}(\mathcal{S}^{LR}_\cdot, \alpha, \theta), \mathcal{S}^k_{k-best})$ (see Def. 3.8). The external validation criterion measures the degree to which a given set of regions is representative of the allocation decisions in the solutions of another set of solutions. Values close to 1.0 indicate that the service regions obtained from sets of linearly relaxed solutions persist throughout well-performing and optimal solutions.

Example A 4.4 We derive the $\mathcal{S}^{LR}_{opt}$, $\mathcal{S}^{LR}_{I\setminus i}$, and $\mathcal{S}^{LR}_{I}$ for instances $\mathcal{P}_1$ and $\mathcal{P}_2$ from Example A. Based on these sets, we determine service regions using the aggregation functions $\alpha^{indicator}$ and α^{profit}, and threshold levels θ of 0.01 and 0.05. Table 4.1 depicts the number of service regions *RegClus* produces for the different inputs.

$\mathcal{S}^{LR}_\cdot$	$\mathcal{S}^{LR}_{opt}$				$\mathcal{S}^{LR}_{I\setminus i}$				$\mathcal{S}^{LR}_{I}$			
α	$\alpha^{indicator}$		α^{profit}		$\alpha^{indicator}$		α^{profit}		$\alpha^{indicator}$		α^{profit}	
θ	0.01	0.05	0.01	0.05	0.01	0.05	0.01	0.05	0.01	0.05	0.01	0.05
$\mathcal{P}_1$	17	17	17	17	1	2	4	11	1	2	4	13
$\mathcal{P}_2$	7	7	7	7	1	4	4	7	1	1	1	5

Table 4.1.: Number of regions in linearly relaxed solutions ($\mathcal{P}_1$-$\mathcal{P}_2$, Ex. A)

For $\mathcal{S}^{LR}_{opt}$, one service region is produced for every facility serving some customers in the optimal solution to the linearly relaxed problem. When service regions are derived from $\mathcal{S}^{LR}_{I\setminus i}$ or $\mathcal{S}^{LR}_{I}$ the number of service regions the facility-customer space is partitioned into decreases significantly. Particularly, for $\mathcal{S}^{LR}_{I}$ and $\mathcal{P}_2$ only for α^{profit} and $\theta = 0.05$, more than one service region can be distinguished.

We derive sets of k best alternative solutions for $\mathcal{P}_1$ and $\mathcal{P}_2$, for $k \in \{1, 5, 20\}$. If k equals 1, we only consider the optimal solution to the CFLP. The results for the external validation criterion for the different combinations of service regions and sets are displayed in Table 4.2.

	$\mathcal{S}^{LR}_\cdot$	$\mathcal{S}^{LR}_{opt}$				$\mathcal{S}^{LR}_{I\setminus i}$				$\mathcal{S}^{LR}_{I}$			
	α	$\alpha^{indicator}$		α^{profit}		$\alpha^{indicator}$		α^{profit}		$\alpha^{indicator}$		α^{profit}	
	k θ	0.01	0.05	0.01	0.05	0.01	0.05	0.01	0.05	0.01	0.05	0.01	0.05
$\mathcal{P}_1$	1	0.63	0.63	0.69	0.69	1.00	0.99	0.96	0.91	1.00	0.95	0.91	0.86
	5	0.57	0.57	0.69	0.69	1.00	0.97	0.96	0.92	1.00	0.96	0.91	0.86
	20	0.46	0.46	0.68	0.68	1.00	0.95	0.95	0.91	1.00	0.95	0.92	0.86
$\mathcal{P}_2$	1	0.44	0.44	0.38	0.38	1.00	0.77	0.76	0.63	1.00	1.00	1.00	0.67
	5	0.32	0.32	0.38	0.38	1.00	0.72	0.76	0.62	1.00	1.00	1.00	0.70
	20	0.31	0.31	0.38	0.38	1.00	0.67	0.71	0.62	1.00	1.00	1.00	0.69

Table 4.2.: External loss $\ell^{external}_\alpha(\mathcal{R}(\mathcal{S}^{LR}_\cdot, \alpha, \theta), \mathcal{S}^k_{k-best})$ ($\mathcal{P}_1$-$\mathcal{P}_2$, Ex. A)

For $\mathcal{P}_1$, integer-infeasible solutions are well-suited to determine service regions that characterize well-performing and optimal solutions. The external validation metric exceeds 0.9 for all service regions derived from $\mathcal{S}^{LR}_{I\setminus i}$ and the majority of service regions derived from

$\mathcal{S}_I^{LR}$. For $\mathcal{P}_2$, those configurations that produce more than one service region perform significantly worse, with $\ell_\alpha^{external}$-values between 0.6 and 0.8.

Throughout both instances, we observe that the best results are obtained with the set of solutions $\mathcal{S}_{I\setminus i}^{LR}$. The aggregation function $\alpha^{indicator}$ combined with a threshold of 1.0 is too restrictive, and only a single service region is produced. Best results throughout all configurations and instances are obtained using $\alpha^{indicator}$ and a threshold of 0.05. Similarly, good results that perform slightly better on sets of k best solutions for $\mathcal{P}_2$ are obtained with α^{profit} and a threshold of 0.01. ▲

The results from Example A 4.4 show that service regions derived from integer infeasible solutions bear the potential to identify persistent decision patterns that characterize well-performing and even optimal solutions. In the following, we validate these initial findings on a broader set of instances.

4.2.2. Experimental validation: linking service regions to infeasible solutions

We repeat the experiment on the instances from the data sets described in Section 1.3. We restrict considerations to the set of linearly relaxed solutions $\mathcal{S}_{I\setminus i}^{LR}$ which performed best in Example A 4.4. Table 4.3 displays the average number of regions obtained from $\mathcal{S}_{I\setminus i}^{LR}$ for different combinations of α and θ. Furthermore, it shows the average external loss determined for the optimal solution ($k = 1$ in the set of k best solutions). Results are grouped according to problem size and average dependence density.

		$\alpha^{indicator}$, $\theta = 0.01$		$\alpha^{indicator}$, $\theta = 0.05$		α^{profit}, $\theta = 0.01$	
$\lvert I\rvert$	Dep^{avg}	avg. $\lvert\mathcal{R}\rvert$	avg. $\ell_\alpha^{external}$	avg. $\lvert\mathcal{R}\rvert$	avg. $\ell_\alpha^{external}$	avg. $\lvert\mathcal{R}\rvert$	avg. $\ell_\alpha^{external}$
$[0, 25)$	$[0, 0.1)$	1.2	1.00	1.2	0.99	1.8	0.98
	≥ 0.1	1.0	1.00	1.2	0.99	1.3	0.99
$[25, 50)$	$[0, 0.1)$	1.1	1.00	1.4	0.98	1.7	0.96
	≥ 0.1	1.0	1.00	1.3	0.99	1.2	0.99
$[50, 100)$	$[0, 0.1)$	1.9	0.96	2.8	0.91	4.7	0.89
	≥ 0.1	2.8	0.90	3.1	0.86	2.6	0.87
$[100, 200)$	$[0, 0.1)$	2.7	0.97	3.7	0.93	6.3	0.79
	≥ 0.1	1.1	0.98	2.6	0.87	5.5	0.79
≥ 200	$[0, 0.1)$	1.4	0.98	2.8	0.92	2.8	0.91
	≥ 0.1	1.4	0.98	3.6	0.87	6.6	0.61

Table 4.3.: Average number of regions and average external loss for the optimal solution

The results show that the service regions derived from linearly relaxed solutions represent the allocation pattern in the optimal solution. The external loss is relatively high throughout and often exceeds 0.95. We conclude from the considered parameter configurations that when considering integer-infeasible solutions, the aggregation function $\alpha^{indicator}$ performs better than an aggregation function that emphasizes the actual allocation volume, i.e., α^{profit}. While for $\theta = 0.01$, few service regions can be distinguished. Instead, $\theta = 0.05$ yields a good trade-off between the number of service regions produced and the representation of these service regions in the optimal solution.

4.3. Conclusion

We assess whether service regions persisting throughout optimal and near-optimal solutions to an instance can be derived either purely from the problem's input data before solving the CFLP even once or from sets of solutions that are potentially easier to obtain, e.g., sets of solutions to the linear relaxation. For the first, we present concise measures comprising the characterizing properties of the multi-dimensional data vector constituting the CFLP instance. We relate these measures to the average dependence density and demand reallocation. These measure the level of interdependence between the facilities operating in the optimal solution. Increasing values imply larger service regions. However, while the marginal effect of individual parameter variations on the level of interdependence can be observed and explained, anticipating the level of interdependence in new instances could not be achieved with the considered methods. While it is intuitive to assume that, i.e., properties that characterize the spatial patterns underlying a problem instance can be related to the implied separation into service regions, other characteristics like the demand and capacity ratio also affect the size of these regions. This makes it challenging to anticipate the level of interdependence for unseen instances.

We evaluate the level to which service regions can be derived from sets of solutions that are easier to obtain than the optimal solution but still contain the added information on the objective and constraints that is missing when looking purely at the data. Therefore, we derive service regions from integer-infeasible solutions and evaluate whether they are representative of service regions identifiable in sets of k best alternative solutions. The results indicate that the so-obtained service regions perform well in grouping together facilities that interdepend in the optimal solution.

The main results of the previous section can be summarized as follows:

- Service regions cannot be derived purely from the problem's input data. In particular, they are not a direct consequence of the underlying spatial pattern. The latter is only one factor shaping these regions.
- It is possible to obtain service regions that characterize the allocation patterns in well-performing solutions from sets of integer-infeasible solutions.

We will evaluate the value of this newborn insight during our examination of the interplay between service regions and the performance of solution algorithms in Chapter 5.

5. Computational implications of service regions

The reported solve times for instances $\mathcal{P}_1$-$\mathcal{P}_4$ from Example A indicate that the size of the underlying service regions in a problem instance may be related to the difficulty this instance poses for state-of-the-art solvers. The service regions result from interdependencies between individual location decisions and, therefore, are a manifestation of the combinatorial element of the CFLP.

In the following chapter, we investigate how the implied separation into distinct service regions of the facility-customer space and the interdependence of location decisions serving the same region affect the performance of both heuristic and exact solution procedures. In particular, we look at the following two questions

RQ4: How do service regions affect the performance of exact and heuristic solution procedures for the CFLP?

RQ5: Can the acknowledgment of service regions improve the performance of exact solution algorithms?

In Section 5.1, we survey publications on exact and heuristic solution algorithms for the CFLP regarding whether or not the challenges these algorithms face on individual instances are further investigated. In particular, we are interested in whether these challenges are attributed to characteristics of the problem instance or the optimal solution.

In Section 5.2 and Section 5.3, we address RQ4. In Section 5.2, we evaluate the performance of two heuristics, ADD and Kernel Search, and relate their performance to the size of their underlying service regions. We use the instances from Example A to analyze the individual steps of the algorithm and determine at which point the interdependence of the optimal facilities serving customers in a service region jointly poses a challenge. We then evaluate the relationship between the size of the underlying service regions and several performance measures of CPLEX's MIP solver as an example of a state-of-the-art exact branching procedure. A detailed analysis of the search trees produced for the instances from Example A indicates that ignoring the implied spatial relationships leads to inefficient variable selection decisions.

Based on these newborn insights, we address RQ5 in Section 5.4. We use *RegClus* (Algorithm 2 introduced in Chapter 3) to derive service regions from integer-infeasible solutions early in the search tree of the branching procedure. We demonstrate that acknowledging these service regions in the subsequent search process can lead to more efficient variable selection and, thus, shorter solution times. We end with a brief outlook in Section 5.5.

5.1. Related work on the characteristics of challenging instances

Several well-performing exact and heuristic solutions procedures for the CFLP were proposed in the early 2000s. However, today's state-of-the-art solvers, e.g., CPLEX's MIP

solver, can solve instances of the CFLP that consist of several hundred facilities and customers in a reasonable amount of time, even on personal-use hardware. For strategic location decisions, decision-makers typically do not require solutions within seconds but can wait for a few minutes or even hours. Consequently, it is unsurprising that the number of publications on this topic decreased in recent years.

Although identifying an optimal solution for moderately sized CFLP instances is less challenging today than it was before, the properties that make some instances much harder to solve than others are still unknown. The notion of abstract decision patterns that persist throughout optimal and near-optimal solutions is a concept newly introduced in this thesis. Thus, their relationship to the performance of different algorithms is an unexplored issue. Meanwhile, understanding the difficulties that specific data and decision structures present can not only yield important insights when it comes to solving instances but also be valuable when solving extensions of the problem.

We review works presenting heuristic and exact solution algorithms for the CFLP and a variant, the Single-Source CFLP (SS-CFLP). We must point out that all works discussed in the following consider cost-minimizing formulations of the CFLP that require full demand satisfaction. The primary objective of the computational experiments in these works is to showcase the efficiency of the presented algorithms on benchmark instances. Few authors point out individual instances that pose a particular challenge to their algorithm. The distinguishing features that cause this challenge are rarely examined.

Table 5.1 presents an overview of the reviewed works. It provides information on the problem (either CFLP or Single-Source CFLP (SS-CFLP)), whether an exact or heuristic algorithm is presented, the benchmark instances that were used, and whether or not the authors made any remarks on the particular difficulty of specific instances. The overview highlights that comments on the difficulty of individual instances are rare.

Ahuja et al. [2004] present a very large-scale neighborhood search (VLNS) algorithm for the SS-CFLP. They test their procedure on benchmark instances from ORLIB, OLRIB-L, and HOL-1999, as well as undisclosed real-world instances based on an Italian cookie factory. While the authors explicitly address challenges solving instances $p29$-$p32$ from the HOL-1999 data sets, the reasons are not further explored. The instances are simply said to exhibit an inherent difficulty. Chia-Ho Chen [2008] combine a Lagrangian Heuristic and Ant Colony System (LH-ACS) to solve the SS-CFLP. They test their procedure on instances from HOL-1999 and ORLIB-L. As their procedure performs better than previous ones, differences among the performance on individual instances are not further discussed. Yang et al. [2012] present a Cut-and-Solve (C&S) based algorithm for the SS-CFLP. They test their procedure on the data sets HOL-1999 and BAR-1991 and some self-generated instances. The authors determine the ratio between fixed and variable costs as the distinguishing characteristic between different problem instances and generate instances in which either one or the other dominates. However, they do not link differences in the performance of their procedure to these characteristics.

As discussed in Section 2.2, one of the best-performing heuristics for the CFLP to date

Source	Model	Procedure		Data	Remarks on indiv. instances
Díaz and Fernández [2002]	SS-CFLP	B&P	(E)	BAR-1991	-
Ahuja et al. [2004]	SS-CFLP	VLNS	(H)	ORLIB ORLIB-L HOL-1999	✓
Klose and Görtz [2007]	CFLP	B&P	(E)	ORLIB-L KLO-2007	✓
Chia-Ho Chen [2008]	SS-CFLP	LH-ACS	(H)	BAR-1991 HOL-1999	
Avella et al. [2009]	CFLP	B&C&P	(E)	KLO-2007 AVE-2009	✓
Görtz and Klose [2012]	CFLP	B&B + LR	(E)	KLO-2007 AVE-2009	✓
Yang et al. [2012]	SS-CFLP	C&S	(E)	HOL-1999 BAR-1991	
Guastaroba and Speranza [2012]	CFLP	KS	(H)	ORLIB ORLIB-L AVE-2009	
Guastaroba and Speranza [2014]	SS-CFLP	KS	(H)	ORLIB ORLIB-L HOL-1999	

E: Exact, H: Heuristic

Table 5.1.: Related work on heuristic and exact procedures for the CFLP and SS-CFLP

is the Kernel Search (KS) algorithm as presented in Guastaroba and Speranza [2012]. After initially presenting the algorithm for the (multi-source) CFLP, two years later, the authors present results for the SS-CFLP in Guastaroba and Speranza [2014]. In both works, the authors perform extensive computational results on the instances from ORLIB, ORLIB-L, HOL-1999, and AVE-2009. However, the focus of these experiments is the performance of their KS with different parameter settings. Differences regarding the performance in individual instances are not discussed in detail. Díaz and Fernández [2002] propose a Branch-and-Price (B&P) procedure for the SS-CFLP and test their procedure on the BAR-1991 benchmark instances. The results are competitive with the state-of-the-art solution methods. However, they already indicate extreme variation in the difficulty posed by individual instances. For example, for instances $p42$-$p49$, obtained with the same generation procedure and containing 30 facilities and 75 customers each, the number of nodes explored during CPLEX's branch-and-cut solver varies between 408 and more than 50,000. However, these differences are not further investigated.

Klose and Görtz [2007] present another B&P procedure for the CFLP and evaluate their procedure on the ORLIB-L instances, and the self-generated instances according to the generation procedure presented in Cornuejols et al. [1991]. The KLO-2007 data set used throughout this work is derived according to the same procedure with the parameter settings from Klose and Görtz [2007]. The authors vary the capacity demand ratio, the tightness (see Def. 4.1), between 3, 5, and 10. The authors observe significant differences in the difficulty of instances from the two considered sets. In particular, instances from the ORLIB-L data set of size 1000×100 ($|J| \times |I|$) can be solved in a magnitude five times faster than the instances of the generator of size 500×100. The authors suggest that a reason for this is that the fixed costs to capacity ratio in the generated instances reflect

economies of scale, while in the ORLIB-L instances, all facilities have the same capacity with varying fixed costs. This makes some facilities more advantageous than others, which in turn makes them easier to identify as optimal. Furthermore, the authors observe that the tightness has a profound effect on the resulting difficulty of the problem instance. In particular, their procedure is outperformed by CPLEX's MIP solver on instances with a high tightness ratio. They attribute this to the fact that, amongst others, these instances are characterized by an extremely large number of solutions whose objective function value is merely as small as the objective function value of an optimal solution. The authors conclude that a method that requires only a relatively "small" effort per node will then likely outperform a method that requires a large effort per node in the hope of reducing the number of nodes to be processed further.

Avella et al. [2009] propose a Branch-and-Cut-and-Price algorithm that is based on a reformulation of the CFLP as a fixed charge network flow problem. The latter yields novel opportunities to generate cutting planes. The authors test their procedure on instances generated according to the same procedure as KLO-2007. They acknowledge that tightness values of 2 or 3 lead to easy-to-solve instances. Therefore, they generate instances with tightness values up to 20 and increase the number of candidates and customers to 1000. The authors observe that their cut generation procedure is beneficial for very large instances with a tightness ratio greater than 5. For instances generated with tightness values of 5 and lower, the value of the LP relaxation is already relatively close to the optimal solution. We look at this in Section 5.3.

Görtz and Klose [2012] present a Branch-and-Bound (B&B) method based on Lagrangean Relaxation (LR) and subgradient optimization. The authors use the instances from ORLIB-L, KLO-2007, and AVE-2009. They observe that the fixed costs significantly affect the difficulty of an instance. In particular, some instances from AVE-2009 exhibit smaller fixed costs, resulting in more facilities being open in the optimal solution, resulting in less restrictive capacity restrictions. According to the authors, this makes instances with high tightness ratios relatively easy to solve. Also, instances that exhibit higher transportation costs (5 times larger than in the original generator) are determined to be relatively easy to solve. Notice that both characteristics, low fixed costs and high transportation costs, lead to the same effect: a high relevance of the transportation costs, ergo a need to be close to customers and hence a decreased relevance of the capacity restrictions as more facilities operate in an optimal solution.

In the following, we take a novel approach to investigating the immense variation of the difficulty of individual problem instances of the same size. In particular, we investigate the degree to which this difficulty is related to the size of the implied service regions. Facilities serving customers in the same service regions interdepend. Consequently, solutions based on large service regions exhibit more interdependencies between individual decisions. We show that these interdependencies pose a challenge to well-performing heuristics (Section 5.2) and exact branching procedures (Section 5.3). In Section 5.4, we explore how acknowledging these interdependencies can be used to speed up exact branching procedures. To the best of our knowledge, this is the first work that explores this direction.

5.2. Implications for heuristics

Several well-performing heuristics for the CFLP implicitly or explicitly exploit that the optimal solutions to many instances are composed predominantly of independent facilities (see Def. 2.11). These facilities are part of the optimal solution because they serve customers in a relatively small service region optimally, independently of the allocations in other parts of the facility-customer space. In consequence, algorithms that operate in a forward-myopic manner are likely to perform well. Iteratively adding locally optimal candidates to a solution is successful when the interdependencies between the individual facilities play a minor role in the optimal solution. At the same time, one can expect a weaker performance of greedy algorithms in solutions predominantly composed of interdependent facilities, facilities that serve larger service regions optimally only in combination. This is particularly true when the (individually) most profitable facility servicing a service region is not part of the optimal solution, i.e., not part of the optimal interdependent subset of facilities serving that region. The forward-myopic procedure of the greedy algorithm will start by picking the most profitable facility and fail to anticipate its connections with other facilities in that region.

We verify this assumption by looking at the performance of a specific greedy procedure. We then evaluate the performance of KS, one of the best-performing heuristics for the CFLP.

5.2.1. Effect on basic procedures: Greedy (ADD)

Jacobsen [1983] present the *ADD* procedure for the CFLP as a generalization of a greedy procedure presented by Kuehn and Hamburger [1963] for the uncapacitated p-median problem. *ADD* starts from no facilities operating and, in each iteration, "greedily" adds that facility to the set of operating facilities, which yields the greatest improvement of the objective value. Contrasting the profit-maximizing model considered throughout this work, Jacobsen [1983] present and test their procedure on a cost-minimizing CFLP, which requires full demand satisfaction. We outline an adapted version of the *ADD* procedure in Algorithm 3. In its basic form, *ADD* requires solving one transportation problem for every closed candidate in every iteration. To speed up the solution of the latter, Jacobsen [1983] present methods relying on upper and lower bounds for these transportation problems rather than their exact solutions. In Algorithm 3, we use a greedy approach to allocate customers to facilities.

Example A 5.1 (Performance of *ADD* on service regions of different size) Consider instances $\mathcal{P}_1$-$\mathcal{P}_4$ from Example A. $\mathcal{P}_1$ and $\mathcal{P}_3$ are two instances whose optimal solutions are primarily composed of independent facilities implying relatively small service regions. $\mathcal{P}_2$ and $\mathcal{P}_4$ exhibit strong combinatorial dynamics with larger subsets of interdependent facilities serving larger underlying service regions. Table 5.2 displays the objective value obtained with *ADD*, z^{ADD}, the optimal objective value, $z^\star$, and their relative gap, gap_{ADD}. Let I^{ADD} denote the index set of facilities operating in the solution produced by *ADD*. Table 5.2 furthermore displays the proportion of facilities in I^{ADD} that also operate in

Algorithm 3 *ADD* for the profit-maximizing CFLP

Input: $\mathcal{P}$
Output: z^{ADD}

$z^{ADD} \leftarrow 0$, $I^{open} \leftarrow \emptyset$, $I^{closed} \leftarrow I$, $done \leftarrow$ `FALSE`, ▷ Initialize.
while not *done* **do** ▷ Iteratively select and open facilities.
  $z^{currentbest} \leftarrow 0$, $i^{currentbest} \leftarrow$ None
  for $i' \in I^{closed}$ **do** ▷ Evaluate all remaining facilities.
    $z^{i'} \leftarrow -\sum_{i \in I^{open}} F_i + F_{i'}$
    $\eta_{ij} \leftarrow \begin{cases} r_j - c_{ij} & \text{, if } i \in I^{open} \cup \{i'\}, \\ 0 & \text{, otherwise.} \end{cases} \forall i, j$
    $Q'_i \leftarrow \begin{cases} Q_i & \text{, if } i \in I^{open} \cup \{i'\}, \\ 0 & \text{, otherwise.} \end{cases} \forall i$
    $D'_j \leftarrow D_j \ \forall j$
    while $\min\{\sum_i Q'_i, \sum_j D'_j\} > 0$ **do** ▷ Determine allocation and objective
      $i'', j \leftarrow \arg\min_{i,j} \eta_{ij}$ ▷ value when adding i' to the solution.
      $volume \leftarrow \min\{Q'_{i''}, D'_j\}$
      $z^{i'} \leftarrow z^{i'} + \eta_{i''j} \cdot volume$
      $Q'_{i''} \leftarrow \max\{Q_{i''} - volume, 0\}$
      $D'_j \leftarrow \max\{D'_j - volume, 0\}$
      if $\min\{Q'_{i''}, D'_j\} = 0$ **then**
        $\eta_{i''j} \leftarrow 0$
      end if
    end while
    if $z^{i'} > z^{currentbest}$ **then** ▷ Evaluate which i' yields the biggest improvement.
      $z^{currentbest} \leftarrow z^{i'}$ and $i^{currentbest} \leftarrow i'$
    end if
  end for
  if $z^{currentbest} > z^{ADD}$ **then** ▷ If the obj. value can be improved, add $i^{currentbest}$
    $z^{ADD} \leftarrow z^{best}$, $I^{open} \leftarrow I^{open} \cup (i)$, $I^{closed} \leftarrow I^{closed} \setminus i$ ▷ to the solution.
  else
    $done =$ `TRUE` ▷ Terminate otherwise.
  end if
end while
return z^{ADD}

the optimal solution as well as the proportion that is part of the set of replacements, *Rep* (refer to Subsection 2.3.2). There is a clear distinction between the performance of *ADD* on instances $\mathcal{P}_1$ and $\mathcal{P}_3$, and the instances $\mathcal{P}_2$ and $\mathcal{P}_4$. For $\mathcal{P}_1$ and $\mathcal{P}_3$, whose optimal solutions are primarily composed of independent facilities, the solution produced by *ADD* equals the optimal solution. Meanwhile, for $\mathcal{P}_2$ and $\mathcal{P}_4$, whose optimal solutions are primarily composed of interdependent facilities implying larger service regions, gap_{ADD} is 0.5% and 0.6%, respectively. While that appears rather small, it must be pointed out that the present instances have many solutions with an objective value within less than 0.5% of the optimal objective value. Therefore, the results indicate a significant deviation from the optimal policy. This is confirmed when looking at the percentage of facilities operating in the optimal solution that are identified by *ADD*. The location decisions differ significantly, with only 42.9% and 57.1% being part of $I^\star$. For instance $\mathcal{P}_4$, 28.6% of the facilities in I^{ADD} are part of *Rep*. The results confirm that the interdependence of facilities makes them difficult to detect heuristically in a forward-myopic manner.

	z^{ADD}	$z^\star$	gap_{ADD} (%)	$\lvert(I^{ADD} \cap I^\star)\rvert/\lvert I^{ADD}\rvert$ (%)	$\lvert(I^{ADD} \cap Rep)\rvert/\lvert I^{ADD}\rvert$ (%)
$\mathcal{P}_1$	2373.1	2373.1	0.0	100.0	0.0
$\mathcal{P}_2$	19696.2	19803.9	0.5	42.9	0.0
$\mathcal{P}_3$	6525.0	6525.0	0.0	100.0	0.0
$\mathcal{P}_4$	37936.1	38177.0	0.6	57.1	28.6

Table 5.2.: Performance indicators for *ADD* ($\mathcal{P}_1$-$\mathcal{P}_4$, Ex. A)

Next, we look at the individual steps of *ADD* when finding a solution to $\mathcal{P}_4$. Table 5.3 shows the three candidate facilities with the largest values objective value in each iteration, denoted z^1, z^2, and z^3. Independent facilities like facility 128, facility 66, and facility 107 are chosen early in iterations 1, 2, and 4. In iteration 5, facility 10, part of a strongly interdependent subset with facility 127 (iteration 7) and facility 155 (iteration 8), is chosen. The subset facility 10 is part of is not optimal. Instead, facility 10 is part of *Rep*. However, after facility 10 has been identified, none of the other optimal facilities in the corresponding optimal subset was found.

iteration	1	2	3	4	5	6	7
$i : z^i = z^1$	**128**	**66**	**119**	**107**	*10*	144	*127*
$i : z^i = z^2$	**66**	**119**	**107**	*10*	*127*	*127*	*165*
$i : z^i = z^3$	**119**	**107**	*10*	*127*	**155**	*165*	*155*
iteration	8	9	10	11	12	13	14
$i : z^i = z^1$	*155*	**154**	**78**	*125*	137	**162**	**135**
$i : z^i = z^2$	**154**	**78**	*38*	184	20	46	98
$i : z^i = z^3$	**78**	*38*	157	83	24	*152*	46

bold: $\in I^\star$, *italic:* $\in$ *Rep*

Table 5.3.: Top three candidate facilities per iteration in *ADD* ($\mathcal{P}_4$, Ex. A)

▲

The analysis confirms our hypothesis regarding the performance of *ADD* when identifying interdependent subsets of facilities serving larger subsets of customers jointly. The forward-

myopic procedure fails to anticipate interdependencies between facilities and is prone to detect suboptimal subsets. This is confirmed for general instances when looking at a larger database in Subsection 5.2.3.

5.2.2. Effect on advanced procedures: Kernel Search

Algorithm 4 *Kernel Search (KS)* for the profit-maximizing CFLP

Input: $\mathcal{P}$, m
Output: z^{KS}

1: Solve LP relaxation of CFLP.
2: Sort facilities in non-decreasing order based on the demand they serve in the LP relaxation.
3: Build initial kernel, I^K, with I^K the first m facilities from the sorted list.
4: Build the sequence of buckets $\{I^h\}_{h=1,\dots,NB}$ from the $|I| - m$ facilities. Each bucket contains the next m facilities from the sorted list of facilities, such that $NB = \left\lceil \frac{|I|}{m} \right\rceil - 1$.
5: Solve the CFLP considering only the facilities from I^K and obtain z^{KS} as the currently best objective value.
6: **for** $h \in 1, \dots, NB$ **do**
7: Solve CFLP considering facilities in I^K and current bucket I^h.
8: Add the following constraints:
1. $\sum_{i \in I^h} y_i \geq 1$: At least of facility from the bucket must operate.
2. $\sum_i \sum_j x_{ij} D_j (r_j - c_{ij}) - \sum_i F_i y_i > z^{KS}$: The objective value must improve.
9: Obtain $z^\star$ as the optimal solution to this problem.
10: **if** CFLP feasible **then**
11: $z^{KS} \leftarrow z^\star$, $I^K \leftarrow I^K \cup \{i \in I^h | y_i = 1\}$
12: **end if**
13: **end for**
14: **return** z^{KS}

KS is a meta-heuristic that is particularly effective for the CFLP. Algorithm 4 describes KS for the profit-maximizing CFLP. Based on information from the LP relaxation, the binary decision variables are divided into buckets. The bucket containing the most promising subset of decision variables is denoted as the kernel. Then, in an iterative procedure, the decision variables in the kernel are merged with those of individual buckets, and the resulting (small) mixed-integer program is solved. The facilities operating in these solutions are added to the kernel. If a facility does not operate for a given number of iterations, it is removed from the kernel. Besides the problem instance $\mathcal{P}$, the algorithm requires the size of the initial kernel and the buckets m as input parameters. If, at some point, facilities are removed from the kernel if they have not operated in a given number of iterations, this number is another input required. For a detailed description of the algorithm, we refer to Guastaroba and Speranza [2012]. We deviate from their description in the following ways:

- In each iteration and each kernel, we consider the entire set of customers, thus all allocation variables, instead of only the most promising customers for each candidate facility in the kernel.
- As the instances considered are relatively small, we explore the entire list of buckets and do not delete facilities from the kernel at any point.

We choose these deviations to be less dependent on the specific parameter settings of the algorithm, thereby making our results more generalizable. The authors set m to the number of candidates with a positive value in the LP relaxation. This parameter choice is rather generous, as it makes the likelihood of including all optimal facilities already in the initial kernel very high. Instead, we set $m_1 = tightness^{-1} \cdot |I|$, approximating the minimum number of facilities necessary to fulfill all demands. We compare this setting to setting $m_2 = tightness^{-1} \cdot |I| \cdot 0.5$, which significantly decreases the size of the initial kernel and each bucket, making it more likely that interdependent subsets of facilities are assigned to different buckets.

We anticipate the following challenge regarding the detection of interdependent subsets of facilities constituting large service regions. Suppose that, in the initial kernel, a non-optimal facility that is part of the second-best subset of facilities serving a particular service region is included and opened in the solution in the first iteration of the algorithm. Individually, this facility is the most profitable facility in that service region. However, it is not part of the optimal subset of facilities but only part of the set of replacements. This facility is only identified as non-optimal if the entire subset of facilities that optimally serve this service region is evaluated jointly. This would require all facilities in one service region to be in the same bucket. However, as no notion of implied or actual spatial relationship is included in the formation of the buckets, there is no guarantee that such an evaluation will occur at any point during the algorithm.

Example A 5.2 (Performance of *KS* on service regions of different size) Table 5.4 displays the results obtained with *KS* for instances $\mathcal{P}_1$-$\mathcal{P}_4$. The results are compared to the optimal solution.

	m_1			m_2		
	gap_{KS} (%)	$overlap(I^\star, I^K)$	$overlap(I^\star, I^{KS})$	gap_{KS} (%)	$overlap(I^\star, I^K)$	$overlap(I^\star, I^{KS})$
$\mathcal{P}_1$	0.00	0.94	1.00	0.00	1.00	1.00
$\mathcal{P}_2$	0.03	0.57	0.57	0.21	0.60	0.50
$\mathcal{P}_3$	0.00	0.76	1.00	0.00	0.76	1.00
$\mathcal{P}_4$	0.00	0.83	1.00	0.00	0.60	1.00

$m_1 = tightness^{-1} \cdot |I|$, $m_2 = 0.5 m_1$

Table 5.4.: Performance indicators for *KS* ($\mathcal{P}_1$-$\mathcal{P}_4$, Ex. A)

KS performs extremely well on all four instances and finds the optimal solution to instances $\mathcal{P}_1$, $\mathcal{P}_3$, and $\mathcal{P}_4$. The relative objective value gap for $\mathcal{P}_2$ is negligible with 0.03% for m_1 and 0.21% for m_2. We determine the overlap coefficient (refer to Def. 2.7) between the set of facilities operating in the optimal solution, $I^\star$, set of facilities in the initial kernel, I^K, and the set of solutions operating in the solution found by *KS*, I^{KS}, respectively. Particularly for the smaller kernel size, m_2, the results indicate that while for $\mathcal{P}_1$ and $\mathcal{P}_3$ the initial kernel contains many facilities also operating in the optimal solution. For instances $\mathcal{P}_1$ and $\mathcal{P}_3$, the overlap coefficient is 1.0 in both cases as the set of facilities operating in the optimal solution is already in the initial kernel. In fact, for these instances, the initial kernel is composed exclusively of optimal facilities. Meanwhile, for $\mathcal{P}_2$ and $\mathcal{P}_4$, the initial

kernel contains several facilities that are not part of the optimal solution. In particular, when we decrease the size of the initial kernel, the overlap coefficient between $I^\star$ and I^K decreases significantly in both instances, indicating that several of the most "promising" facilities in these solutions are not optimal. While for $\mathcal{P}_4$, the optimal solution could be subsequently retrieved, for $\mathcal{P}_2$, another (well-performing) solution was found at the end in which candidates differed significantly from the candidates in the optimal solution. Therefore, while in the presence of multiple near-optimal solutions, *KS* yields very good results, it is still prone to end up with suboptimal subsets of facilities, particularly, when the optimal interdependent subset of facilities is never evaluated jointly as individual facilities are placed in different buckets. Again, we will validate this observation on a larger set of instances in the following. ▲

5.2.3. Experimental validation: heuristics and large service regions

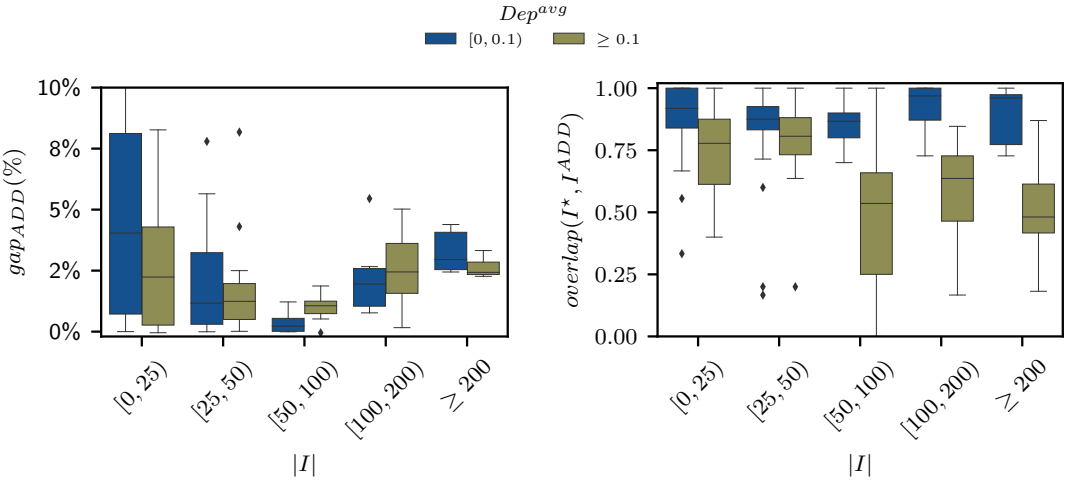
In the following, we validate the above findings on the problem instances from Section 1.3.

5.2.3.1. ADD

Example A 5.1 indicated that *ADD* performs significantly better on instances whose solutions primarily comprise independently operating facilities and exhibit little to no combinatorial dynamics. This is confirmed when looking at a larger set of problem instances from literature, in particular, the instances from Section 1.3. Figure 5.1 depicts the difference between the optimal solutions and the solution obtained with *ADD*.Therefore, we compare solutions in the objective value and the decision space. The observations are grouped according to whether the optimal solution exhibits primarily independently operating facilities ($Dep^{avg} < 0.1$) or whether it depicts interdependent facility subsets of varying size ($Dep^{avg} \geq 0.1$). The differences between these two groups of instances are significant. as indicated by the distribution of gap_{ADD} (%). As Figure 5.1a depicts, there are several other influencing factors besides the size of the underlying service regions. However, for problems with 0 to 199 candidates, *ADD* performs better on facilities with primarily independently operating facilities. Interestingly, the anticipated effect that *ADD* performs better on instances with smaller service regions and, therefore, less combinatorial dynamics ($Dep^{avg} < 0.1$) can only be observed for instances with more than 50 candidate facilities. Consequently, the interdependence of individual facilities is not the only factor affecting the performance of greedy procedures. Figure 5.1b depicts the distribution of the overlap coefficients between the optimal solutions and the solutions obtained with *ADD*. Throughout all instances, we see that the facilities operating in the *ADD* solution differ more significantly from those operating in the optimal solution for instances exhibiting a high interdependence. Consequently, the high observed gaps for instances with a smaller number of facilities may be attributed to the fact that individual facilities are likely to have a higher impact on the optimal objective value with a potentially higher variation.

5.2.3.2. Kernel Search

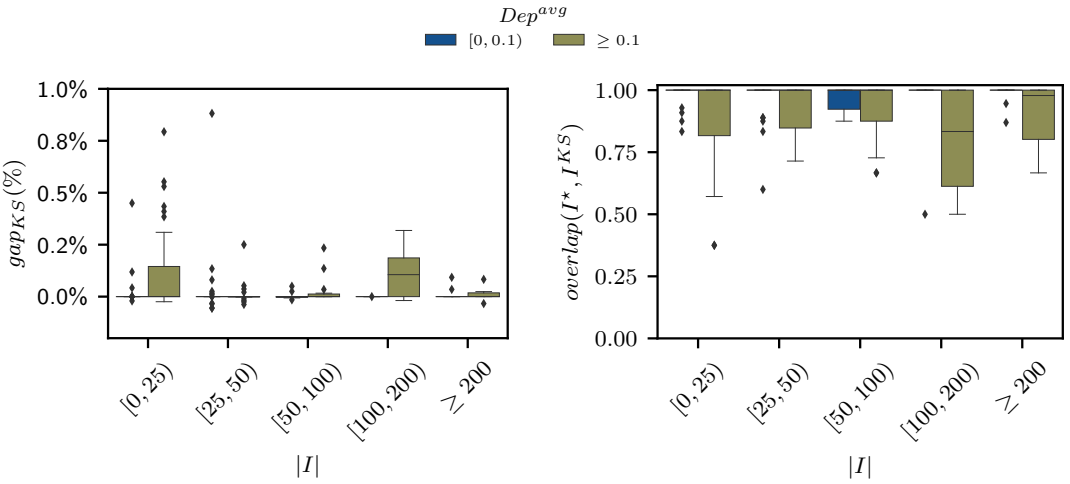
Figure 5.2 depicts the difference between the optimal solutions and the solutions obtained with *KS* in both the objective value and the decision space. In general, *KS* performs



(a) **(b)**

Figure 5.1.: Difference between optimal solution and solutions obtained with *ADD*

significantly better than *ADD*. When the size of the initial kernel (and each bucket) equals the inverse tightness times the number of candidates (m_1), the optimal solution is found for over 90% of the instances. However, as Figure 5.2 depicts, in the presence of large underlying service regions, *KS* may fail to detect the optimal subsets of facilities, and the resulting set of facilities operating in the heuristic solution may differ significantly from the set of facilities operating in the optimal solution. This difference becomes more significant when the size of the kernel is smaller, e.g., the case when $m = m_2$, and interdependent subsets of facilities may not be evaluated jointly.



(a) m_2 **(b)** m_2

Figure 5.2.: Difference between optimal solution and solutions obtained with *KS* with $m_2 = 0.5m_1$

We evaluated the degree to which interdependent subsets of facilities in an optimal solution to a facility location instance affect the performance of heuristics. Both, the detailed analysis of instances from Example A in Example A 5.1 and Example A 5.2 as well as the experiments on sets of instances from Section 1.3, support our hypothesis that interdependent facilities are more difficult to detect heuristically than independent facilities. We will

evaluate the effect on exact branching procedures in the following.

5.3. Implications for exact branching procedures

Rather than a specific algorithm, branching procedures, or, more specifically, branch-and-bound methods, describe a family of algorithms that share a common core solution procedure that allows producing exact solutions to NP-hard problems. The procedures "implicitly enumerate all possible solutions to the problem under consideration by storing partial solutions called subproblems in a tree structure. Unexplored nodes in the tree generate children by partitioning the solution space into smaller regions that can be solved recursively (i.e., branching), and rules are used to prune off regions of the search space that are provably suboptimal (i.e., bounding). Once the entire tree has been explored, the best solution found in the search is returned.[Morrison et al., 2016]" Three components of the procedure have a significant effect on the performance: the search strategy, i.e., the order in which subproblems in the tree are explored; the branching strategy, i.e., how the solution space is partitioned to produce new subproblems in the tree, and the pruning rules, i.e., rules that prevent exploration of suboptimal regions of the tree. Most recent developments focus on developing better pruning strategies, e.g., cutting planes or column-generation procedures.

CPLEX's MIP solver implements a Branch-and-Cut (B&C) procedure in which each component of the branching procedure is enhanced by a bundle of methods. At each node of the search tree, CPLEX solves the LP relaxation, attempts to generate cutting planes to cut off the current solution, invokes a heuristic to try to find an integer feasible solution in the neighborhood of the current relaxed solution, selects a branching variable and finally places two nodes that result from branching up or down on the branching variable back into the tree. For each of these steps, CPLEX includes several automated methods [IBM Corporation, 2021b]. As the solver is proprietary, not all strategies used throughout the solution procedure are publicly available.

Example A 5.3 (Performance of CPLEX's MIP solver) Example A demonstrated that the solve times CPLEX's MIP solver takes to determine optimal solutions to problem instances of the same size may differ significantly. Instances $\mathcal{P}_1$ and $\mathcal{P}_2$ both comprise 100 candidates and customers. Yet, the solver takes 0.19 seconds to find an optimal solution to $\mathcal{P}_1$, while it takes 7.03 seconds for $\mathcal{P}_2$. For instances $\mathcal{P}_3$ and $\mathcal{P}_4$ that contain 200 candidates and customers each, that difference is even bigger with solve times of 1.63 seconds and 54.99 seconds, respectively. This difference regarding the performance of the solver is not only visible in the solve time but also in other statistics describing the B&C procedure as displayed in Table 5.5. The depth of the search tree, the number of nodes explored, and the number of incumbents that were processed differ significantly and explain the differences in the reported solution times. For example, while only eight nodes need to be explored to obtain an optimal solution to $\mathcal{P}_3$, 2326 nodes are explored in the solution process of $\mathcal{P}_4$.

We also report the solutions to the incumbents found throughout the search procedure and determine which percentage of the facilities operating in any incumbent solution is

	Dep^{avg}	size ($\|I\| \times \|J\|$)	time (s)	tree depth	# nodes	# incumb.	$\in I^\star$ (%)	$\in I^\star \cup Rep$ (%)
$\mathcal{P}_1$	0.00	100×100	0.19	0	0	1	100	100
$\mathcal{P}_2$	0.50	100×100	7.03	73	534	8	50	86
$\mathcal{P}_3$	0.00	200×200	1.63	5	8	3	100	100
$\mathcal{P}_4$	0.45	200×200	54.99	87	2326	10	67	94

Table 5.5.: Solve statistics of CPLEX's MIP solver ($\mathcal{P}_1$-$\mathcal{P}_4$, Ex. A)

part of the optimal solution ($\in I^\star$), or either part of the optimal solution or the replacements ($\in I^\star \cup Rep$). For P_1 and P_3, the incumbent solutions contain exclusively facilities operating in the optimal solution. Thus, as already found by Avella et al. [2009], the LP relaxation is close to the optimal solution. In contrast, for instances $\mathcal{P}_2$ and $\mathcal{P}_4$, only 50% and 67% of the facilities from the incumbents also operate in the optimal solution. If one includes the set of facilities in the sets of replacements, these percentages increase to 86% and 94%, respectively. This emphasizes that fractional nodes corresponding to interdependent facilities lead to sub-optimal branching decisions more often and, consequently, explore more sub-optimal solutions. ▲

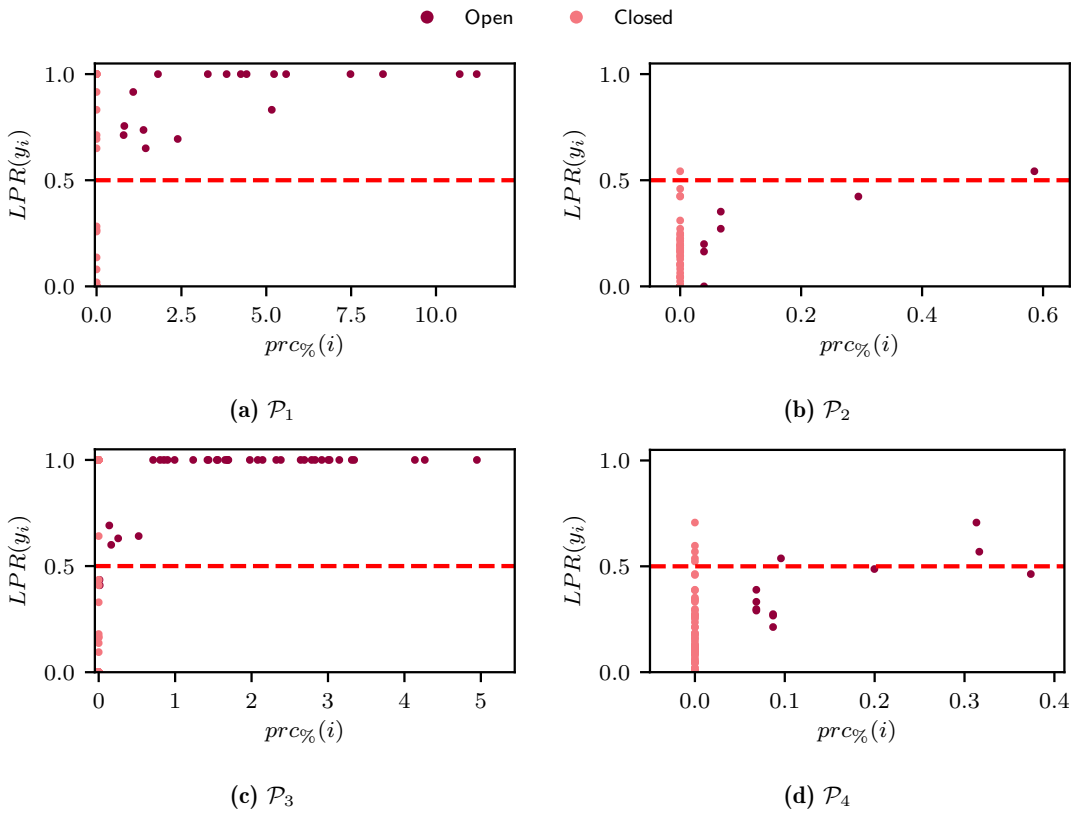
In the following, we explore how the size of the underlying service regions in a given instance, and thus the implied interdependencies between facilities, affects the performance of the branching process. In particular, we look at the effect of the variable selection strategy. Our central hypothesis is that when facilities interdepend, it is advantageous to explore all interdependent facilities in a service region shortly after another. In particular, in Subsection 5.3.1, we determine the degree to which interdependence relationships affect the distribution of values in the LP relaxation. In Subsection 5.3.2, we examine the effect these interdependence relationships have on the performance of different variable selection strategies. Again, we evaluate all hypotheses on the instances from Example A. We validate all findings on instances from Section 1.3 in Subsection 5.3.3.

5.3.1. Effect of service regions in the linear programming relaxation

Avella et al. [2009] observe that for instances generated according to the same procedure as $\mathcal{P}_1$ and $\mathcal{P}_3$ from Example A, the values of the decision variables in the LP relaxation are very close to the optimal solution. State-of-the-art MIP solvers such as CPLEX use a bundle of methods to determine the order in which nodes are branched. However, high values of decision variables in the LP relaxation are a standard criterion for identifying decision variables on which to branch early on. When a facility i that is part of an interdependent subset of suboptimal facilities $Rep(i')$, $i \neq i'$, is branched on early in the search process, the search tree is likely to detect the sub-optimal solution containing the facilities in $Rep(i')$ before the optimal solution operating the facilities in $Dep(i')$. In consequence, the associated branch is not fathomed quickly. As the solver is unaware of these interdependencies the search procedure is prolonged.

Example A 5.4 (Values in the LP relaxation) Figure 5.3 depicts the distribution of the optimal values for the location decisions in the linear programming relaxation ($LPR(y_i)$)

relative to their pseudo-reduced costs (see Def. 2.6). The color indicates whether or not the facilities are open or closed in the optimal solution.



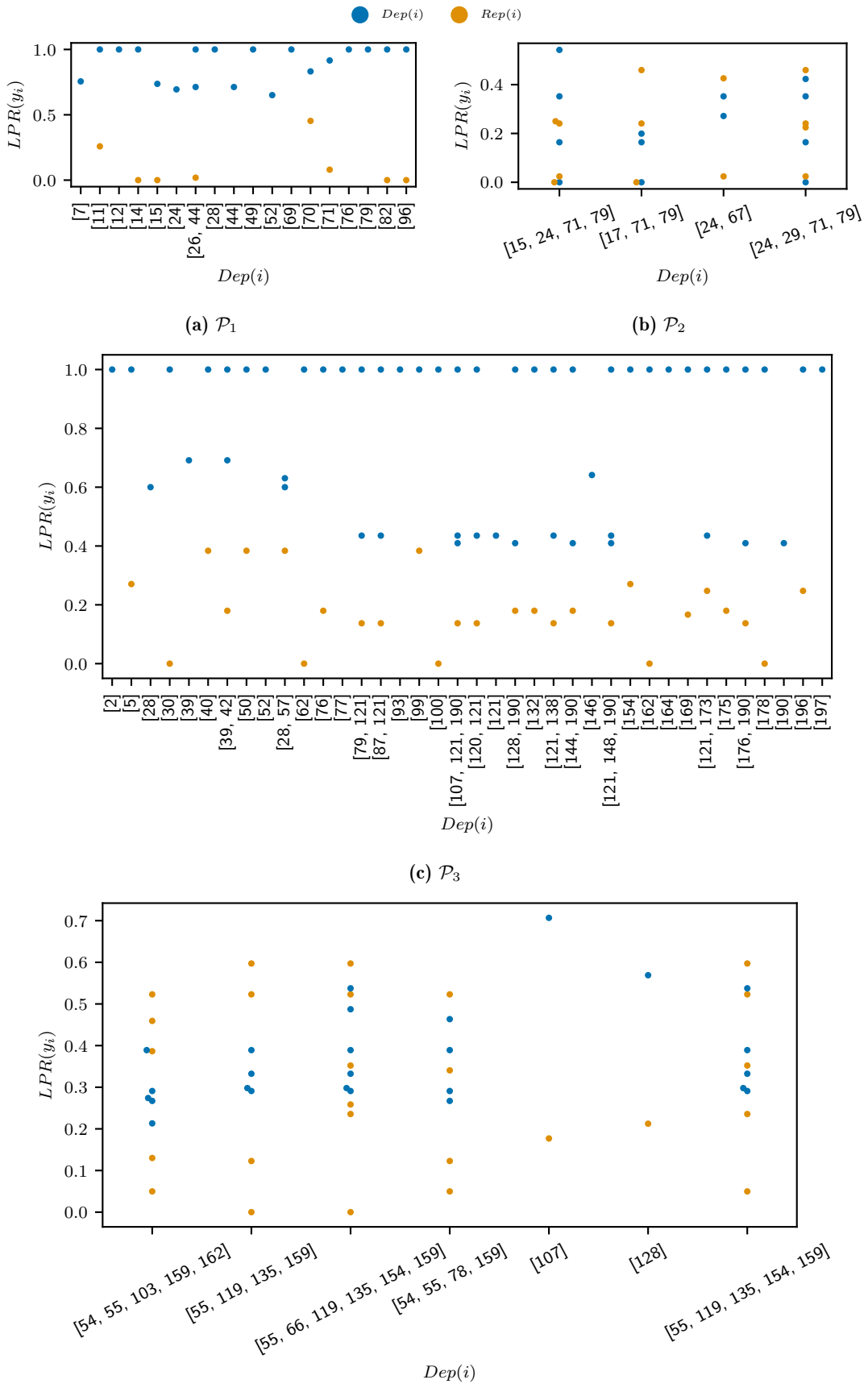
(a) $\mathcal{P}_1$ **(b)** $\mathcal{P}_2$

(c) $\mathcal{P}_3$ **(d)** $\mathcal{P}_4$

Figure 5.3.: LPR-values versus $prc_{\%}$-values ($\mathcal{P}_1$-$\mathcal{P}_4$, Ex. A)

There are significant differences between instances $\mathcal{P}_1$ and $\mathcal{P}_3$ versus $\mathcal{P}_2$ and $\mathcal{P}_4$. In $\mathcal{P}_1$ and $\mathcal{P}_3$, the LPR-values of all open facilities exceed 0.5, with the majority equaling 1.0 in both instances. Thus, the solution to the linear programming relaxation is indeed very close to the optimal solution. For $\mathcal{P}_2$ and $\mathcal{P}_4$, this is totally different. The absolute LPR-values of facilities operating in the optimal solution are relatively low, mostly below 0.5, and in particular, not higher than those of facilities not operating in the optimal solution.

Figure 5.4 sets the LPR-values in relation to whether or not the associated facilities are in the sets $Dep(i)$ or $Rep(i)$. One can see very clearly that for instances $\mathcal{P}_1$ and $\mathcal{P}_3$, the LPR-value of the facilities operating in the optimal solution ($i \in Dep$) are significantly higher than the LPR-value of the facilities operating in the corresponding subsets of replacement facilities ($Rep(i)$). In contrast, for the interdependent subsets of facilities operating in the optimal solutions of $\mathcal{P}_2$ and $\mathcal{P}_4$, for several interdependent subsets $Dep(i)$, the LPR-values of facilities operating in the replacement $Rep(i)$ are higher for at least some facilities in $Rep(i)$. ▲



(a) $\mathcal{P}_1$ **(b)** $\mathcal{P}_2$

(c) $\mathcal{P}_3$

(d) $\mathcal{P}_4$

Figure 5.4.: LPR-values of location decisions for all $Dep(i)$ and $Rep(i)$ ($\mathcal{P}_1$-$\mathcal{P}_4$, Ex. A)

Example A 5.4 supports the idea that interdependent subsets of optimal facilities, as induced by large customer service regions, are difficult to detect for standard branching procedures as oftentimes the LPR-values of facilities that are operating in an interdependent set of replacements are likely to be branched on early on. The branch that operates these facilities is likely only pruned once the entire set of facilities operating in the optimal solution and the set of replacements has been branched on so that a direct comparison is possible. This may occur much later during the search process. To further validate this hypothesis, we take a closer look at the incumbents produced throughout the search process for $\mathcal{P}_2$ and $\mathcal{P}_4$.

Example A 5.5 (Dependence relationships of facilities operating in incumbent solutions) To validate the above considerations, we present the solutions at individual incumbents of the search trees created during solving $\mathcal{P}_2$ and $\mathcal{P}_4$. The subsets of facilities operating in the associated solutions and whether they can later be found in $I^\star$ or Rep are depicted in Table 5.6 and Table 5.7. The node ID represents their rank in the search process. The nodes are ordered sequentially in the order of their creation, not in the order of their exploration. The incumbent in the last row corresponds to the optimal solution. We see two different effects. In the search process for $\mathcal{P}_2$, one part of the search tree with the strongly interdependent subset $\{17, 71, 79\}$ was explored in parallel to another part that explored solutions operating its replacing subset $\{63, 66, 97\}$. For $\mathcal{P}_4$, facility 10 ($\in Rep$)

Node ID	$\in I^\star$	$\in Rep$	other
3	15	-	-
8	15, 29	66	-
11	15, 29, 67	97	-
14	15, 24, 29	66,97	-
17	15, 24, 29	83	5, 59
377	15, 17, 29, 67	21, 66, 97	-
787	15, 17,24, 29, 67	63, 66, 97	-
917	15, 17, 24, 29, 67, 71, 79	-	-

Table 5.6.: Facilities operating in the incumbent solutions ($\mathcal{P}_2$, Ex. A)

ID	$\in I^\star$	$\in Rep$	other
3	107, 128, 154	10, 155	-
6	78, 107, 128, 154	10, 147	-
9	66, 78, 107, 128, 154, 162	10	-
12	66, 78, 103, 107, 128, 154	10, 155	-
13	66, 78, 103, 107, 119, 128, 154, 159, 162	10, 195	98
22	54, 66, 78, 103, 107, 128, 135, 154, 162	10, 152, 155	
131	66,78, 103, 107, 119, 128, 135, 162	10, 147, 152	-
2059	54, 55, 66, 78, 103,107, 128, 154, 162	10, 152	-
66	55,66,78,103,107,119,128,135, 154, 162	10, 152	
70	54, 55,66,78,103,107,119,128,135, 154, 159, 162		

Table 5.7.: Facilities operating in the incumbent solutions ($\mathcal{P}_4$, Ex. A)

is branched on early and operates in the first incumbent solution. Facility 10 is part of the set of replacements of 5 interdependent subsets of facilities. Only when the resulting

branch has been fully explored does CPLEX's search process devotes attention to the other branching direction and finds the optimal solution that does not contain facility 10. This confirms the effect that we anticipated previously. Meanwhile, the two independent facilities, 107 and 128, were easy to detect. They operate in every incumbent solution. These two facilities also had the highest value in the LP relaxation at the root node. Facility 107 had a value of 0.71, and facility 128 had a value of 0.57. ▲

We conclude that combinatorial dynamics in the form of interdependent service regions lead to the fact that the values of the decision variables in the LP relaxation are a less reliable indicator for the optimal branching direction of individual variables. Consequently, it is more likely to happen that suboptimal facilities are operating in early incumbents. If these facilities are part of an interdependent subset of facilities, the entire subset must often be explored before the branch can be fathomed. In consequence, the size of the search tree, particularly its depth, increases, leading to longer solve times.

5.3.2. Effect on the performance of different search strategies

As stated before, CPLEX's MIP solver is a proprietary software that implements a bundle of methods. In particular, it has several ready-to-use options implemented for each step of the branching procedure, e.g., the search strategy (which node to explore next) or the variable selection strategy (which variable to branch on given a fractional solution in a node). For the latter, common state-of-the-art methods like branching on the variable with minimum or maximum infeasibility, the variable with the highest pseudo-reduced costs, the highest estimated pseudo-shadow prices, or strong branching, variable selection based on partially solving a number of subproblems, are pre-implemented and can be selected by the user. However, CPLEX's default variable section strategy is undisclosed and selects "the best rule based on the problem and its progress [IBM Corporation, 2021a]." To keep the procedure proprietary, the user can't read out information on the branch of the current node (e.g., parent nodes) unless it is a user-defined, self-implemented branching strategy.

We use CPLEX's callback classes to implement state-of-the-art variable selection strategies ourselves. This allows us to attach labels to nodes and to retrieve the search tree from a particular solve procedure. We use CPLEX's `BranchCallback` class, which, given the fractional solution at a node, allows the user to create two custom branches. Creating the branches according to a customized procedure implies a custom implementation of the variable selection strategy. We implement the minimum feasibility strategy, a strategy that branches on that supposed-to-be-binary variable that is closest to 0 or 1, maximum feasibility, a strategy that branches on that supposed-to-be-binary variable that is closest to 0.5, and pseudo-cost branching, a strategy that branches on that supposed-to-be-binary variable with the highest estimated pseudo-shadow price. Thereby, CPLEX has a pre-implemented function, `get_pseudo_costs()`, that allows to obtain the current estimated pseudo-shadow price at every node.

By backtracking the search tree, we can get information on how often certain binary variables (location decisions) were branched on and in which order. This, in turn, allows

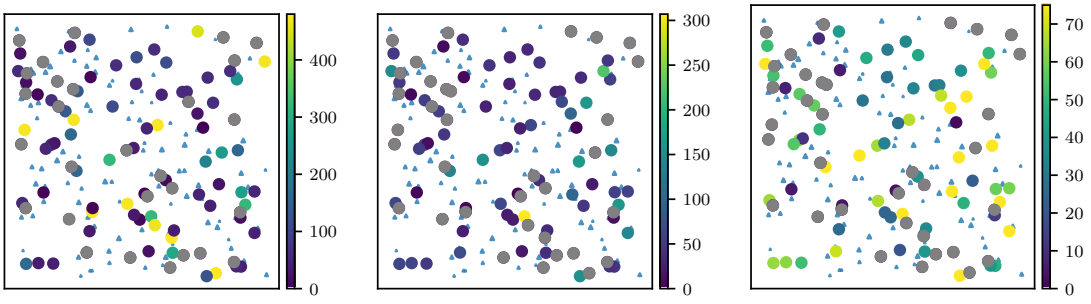
backtracking whether our hypothesis that the ignorance of the interdependence of facilities serving the same service region is a key impediment throughout the search procedure.

Example A 5.6 (Effect of different variable selection strategies) Table 5.8 displays the solve time, the maximum tree depth, as well as the number of explored nodes when solving instances $\mathcal{P}_1$-$\mathcal{P}_4$ with different variable selection strategies. The significant discrepancy between the solve times and the size of the search tree between instances of the same sizes is persistent throughout all search procedures. While CPLEX's default procedure is the best-performing strategy overall, it is slightly outperformed by the max. infeasibility strategy on $\mathcal{P}_4$.

	CPLEX default			min. infeasibility			max. infeasibility			max. pseudo costs		
	time (s)	nodes	depth	time (s)	nodes	depth	time (s)	nodes	depth	time (s)	nodes	depth
$\mathcal{P}_1$	0.2	0	0	0.2	0	0	0.2	0	0	0.2	0	0
$\mathcal{P}_2$	7.1	534	73	14.6	980	64	9.8	524	66	10.8	656	64
$\mathcal{P}_3$	1.6	8	5	1.9	12	5	1.8	8	4	1.8	6	4
$\mathcal{P}_4$	56.0	2326	144	244.2	6574	131	53.5	903	87	100.4	2114	86

Table 5.8.: Performance of different variable selection strategies ($\mathcal{P}_1$-$\mathcal{P}_4$, Ex. A)

We want to know whether the hypothesis that the ignorance of spatial relationships and the resulting interdependencies leads to candidates belonging to the same service region being branched on and searched at very different levels of the search tree. However, the order in which variables are branched on throughout the search process depends not only on the branch variable selection but also on the node selection strategy. Per default, CPLEX is using a best-bound strategy. However, a depth-first strategy is also pre-implemented. Figure 5.5 illustrates the order in which candidate facilities are branched on. The color code indicates the iteration that the candidate facility was branched on for the first time.



(a) best bound **(b)** depth first, all iterations **(c)** depth first, early iterations

Figure 5.5.: Iteration in which variables are branched on first for different search strategies and variable selection strategies ($\mathcal{P}_2$, Ex. A)

Figure 5.5a and Figure 5.5b illustrate the iteration with which candidates are branched on first for the best bound and the depth-first strategy, respectively. Individual candidates are branched multiple times, and some candidates are branched on rather late in the search procedure, e.g., at iteration 300, the color-coding does not reveal the sequence in the early

iterations. Therefore, in Figure 5.5c, the sequence up to iteration 75 is illustrated. The figure clearly shows that the variable-branching and node selection strategy applied by the procedure lead to candidates at different spatial regions of the facility customer space being branched on sequentially. In particular, in every spatial region of the facility-customer space, some facilities are branched on very late in the search process. If the associated variable is essential to rendering a potentially explored subset of facilities sub-optimal, the associated branch will be pruned extremely late. ▲

This supports our hypothesis that the sub-optimality of certain location decisions revealed only in combination with the branching decisions of other candidates in the same service region is detected late in the search process when the spatial relationships of these candidates are not acknowledged. Service regions detected by *RegClus* can be a means to make this implied spatial information available during the search process.

5.3.3. Experimental validation: implications for branching procedures

We hypothesize that the interdependence of individual candidates induced by larger service regions significantly affects the performance of branching procedures, in particular, the performance of CPLEX's MIP solver. In the following, we validate the observations from Example A 5.4 and Example A 5.6 on the instances from Section 1.3. In particular, Table 5.9 depicts the Spearman correlation coefficients between the average dependence density, Dep^{avg}, (see Def. 2.13) and several statistics indicating the complexity of the search process in CPLEX's B&C procedure grouped by problem size. In particular, the table depicts correlation with the solve time in seconds, the depth of the search tree, and the percentage of facilities operating in any incumbent solution operating in the optimal solution $\in I^\star$ (%), or either the optimal solution or any replacement $\in I^\star \cup Rep$ (%). The data shows that within each size group, there is a significant correlation between the solve time as well as the depth of the search tree and the presence of dependence relationships between individual facilities indicated by Dep^{avg}. While an increasing average dependence density negatively affects both the percentage of optimal solutions and their replacements examined in the search process, a linear relationship between Dep^{avg} and $\in I^\star$ (%) or $\in I^\star \cup Rep$ (%) cannot be established for individual size groups. However, the low negative correlation across all size groups indicates a negative impact. Thus, with increasing interdependence relationships between individual facilities, more solutions that operate facilities that neither operate in the optimal solution nor its replacements are examined.

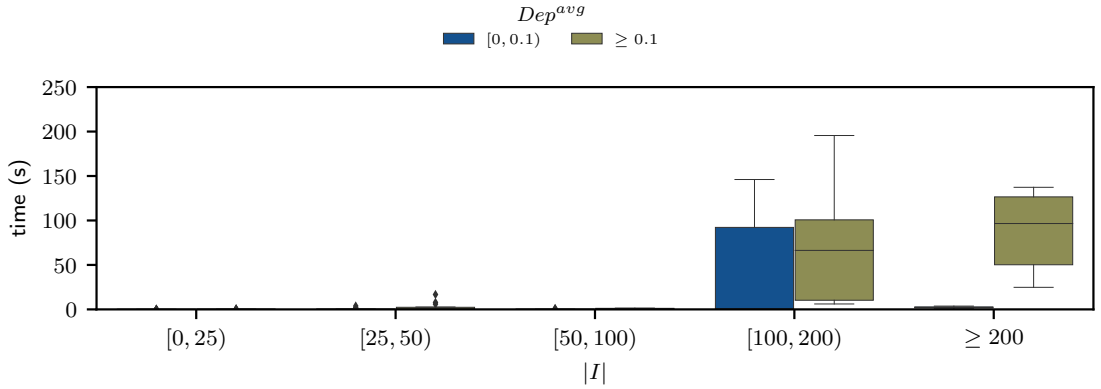
The correlation coefficient of 0.49*** between the total solve time and Dep^{avg} on any instance is significant, even when instances are not grouped by size, but all instances are considered. However, if we determine the correlation coefficient across all instances between the number of candidate facilities $|I|$ and the solve time, we obtain 0.83***. Even though the problem size has a more significant impact on the solve time, the interdependence relationships in the optimal solution significantly affect the performance of the branching procedure.

$\|I\|$	time (s)	tree depth	$\in I^\star$ (%)	$\in I^\star \cup Rep$ (%)
$[0, 25)$	0.76***	0.59***	-0.59***	-0.35*
$[25, 50)$	0.71***	0.54***	-0.22	-0.01
$[50, 100)$	0.84***	0.7***	-0.34	-0.28
$[100, 200)$	0.35*	0.63***	-0.51**	-0.35
≥ 200	0.86***	0.95***	-0.51	-0.46
> 0	0.49***	0.55***	-0.38***	-0.26***

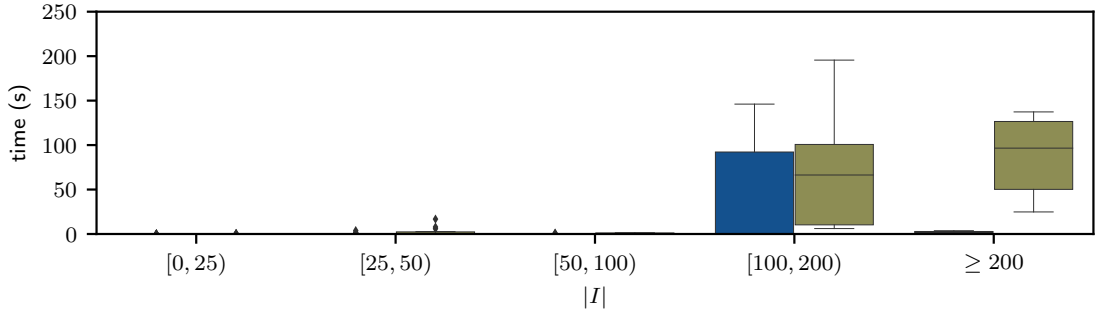
$^*p < 0.05$; $^{**}p < 0.01$;$^{***}p < 0.001$

Table 5.9.: Spearman's correlation coefficient between average dependence density (Dep^{avg}) and statistics of CPLEX's B&C procedure

Figure 5.6 depicts the distribution of solve times observed for different variable selection strategies grouped by whether or not the solution has previously been identified to exhibit strong interdependencies ($Dep^{avg} \geq 0.1$). For both explored search strategies, maximum infeasibility, and maximum pseudo costs, we observe significant differences regarding the solve times. The profound effect of the existence of interdependencies on the size of the search tree is further supported by Table 5.10 and Table 5.11 that respectively display the average depth of the search tree and the average number of nodes explored. They show that for all variable selection strategies, the existence of interdependent subsets of facilities significantly affects the size of the search tree.



(a) max. infeasibility



(b) max. pseudo costs

Figure 5.6.: Distribution of solve times for different variable selection strategies

The exploration of CPLEX's search process made it evident that the sequential branching on individual nodes corresponding to individual facilities inherent to any branching

		tree depth					
		max. infeasibility		max. pseudo costs		min. infeasibility	
$\|I\|$	Dep^{avg}	$[0, 0.1)$	≥ 0.1	$[0, 0.1)$	≥ 0.1	$[0, 0.1)$	≥ 0.1
$[50, 100)$		2.8	16.1	3.4	21.4	3.2	24.3
$[100, 200)$		6.1	48.2	15.1	55.6	18.0	58.1
≥ 200		5.8	84.0	8.6	89.0	10.6	103.0

Table 5.10.: Average depth of the search tree for different variable selection strategies

		# nodes					
		max. infeasibility		max. pseudo costs		min. infeasibility	
$\|I\|$	Dep^{avg}	$[0, 0.1)$	≥ 0.1	$[0, 0.1)$	≥ 0.1	$[0, 0.1)$	≥ 0.1
$[50, 100)$		13.2	98.8	17.8	87.4	23.5	120.9
$[100, 200)$		68.9	667.1	81.9	629.4	124.6	839.4
≥ 200		26.6	964.8	20.8	1926.8	78.0	4034.8

Table 5.11.: Average number of nodes explored for different variable selection strategies

procedure is very effective in identifying independent facilities (see Def. 2.9). However, when facilities in the optimal solution strongly interdepend (see Def. 2.10), it is difficult to prune parts of the tree that explore alternative interdependent subsets. We demonstrated that the state-of-the-art variable and node selection strategies fail to acknowledge the interdependencies of individual decisions, which, in discrete location problems, are induced by the underlying service regions. In the following, we explore how the information on the implied customer regions can help to improve the branching process.

5.4. Pattern-based regional branching

The experiments in Section 5.3 suggest that ignoring the interdependencies of facilities serving the same service region can lead to inefficient branching decisions. Rather than identifying the optimal set of facilities to serve a certain region, the branching procedure jumps from service region to service region, ignoring spatial relationships implied by the profit matrix. This procedure works fine when service regions are small, and facilities are independent. In this case, just as in greedy procedures, successively branching on individual variables of different regions will be just as efficient as branching according to the implied regional pattern. However, when facilities interdepend, it is likely that branches can only be pruned when all promising facilities belonging to the same service region have been branched on. When this happens at very different levels of the tree, this leads to large search trees. In consequence, the search procedure takes longer.

In the following, we explore how information on the interdependence of location decisions can enhance the branching procedure. In particular, we identify service regions from fractional solutions early in the search tree and then branch the variables of facilities serving different regions sequentially. The idea is that by branching on variables from the same region one after another, optimal interdependent subsets can be identified earlier, leading to lower tree depth and, thus, faster solve times. We use the *RegClus* (Algorithm 2 introduced in Section 3.2) to identify these service regions. This, in particular, implies that

no information on the underlying spatial structure in the form of coordinates or transport costs adhering to a specific metric is necessary.

We describe the regional variable selection strategy and perform preliminary experiments on the instances from Example A in Subsection 5.4.1. In Subsection 5.4.2, we test the proposed branching strategy on the instances from Section 1.3.

5.4.1. Regional variable selection

We first describe the proposed regional variable selection strategy which assumes that information on the underlying service regions is readily available. We subsequently discuss how to obtain these service regions from fractional solutions early in the search tree, which is the more challenging question in this context.

Given a set of regions $\mathcal{R}$ and a fractional solution $\hat{s} = (\hat{z}, \hat{y}, \hat{x})$ at a node, we choose the next variable to branch on based on the procedure described in Algorithm 5. We first choose a region k to concentrate on. Of all regions that still contain facilities whose associated decision variables are fractional, we choose the smallest one, that is, the one with the fewest facilities. Then, among all facilities in the region k whose associated decision variable is fractional, we use the maximum infeasibility strategy. This means we choose that facility whose value is closest to 0.5.

Algorithm 5 Regional variable selection

Input: $\mathcal{R}$, $\hat{y}$
Output: i
1: $k \leftarrow \arg\min_{k=1,\dots,|\mathcal{R}|} |I_k| s.t. \sum_{i \in I_k} \hat{y}_i$ is fractional ▷ Choose target region.
2: $i \leftarrow \arg\min_{i \in I_k} |\hat{y}_i - 0.5|$ ▷ Choose facility within that region.
3: **return** i

Algorithm 5 is simple and straightforward. The challenging task is not to determine a branching variable but to obtain the set of regions $\mathcal{R}$ in the first place. We determine $\mathcal{R}$ with *Regclus*. *Regclus* requires three inputs: a set of solutions $\mathcal{S}$, an aggregation function α, and a target level of coherence of the produced regions θ. For each of these inputs, several choices present themselves.

The first question concerns the solutions that should be considered when invoking *Regclus*. When considering only the solution to the LP relaxation, it is likely that, unless customers are served by more than one facility, each fractional facility will be assigned its own service region, and interdependence relationships remain undetected. Information from more than one solution is needed to detect the dependence relationships that form distinctive service regions. One option is to take the solutions to the first n nodes of the search tree. This way, some branching decisions will have enforced some overlap between the subsets of customers served by individual facilities. Another idea is to take only the first n "good" solutions, e.g., the first n incumbents or the first n solutions at a specified depth of the search tree. In the latter case, good means that the associated branch could not be pruned up to a certain level.

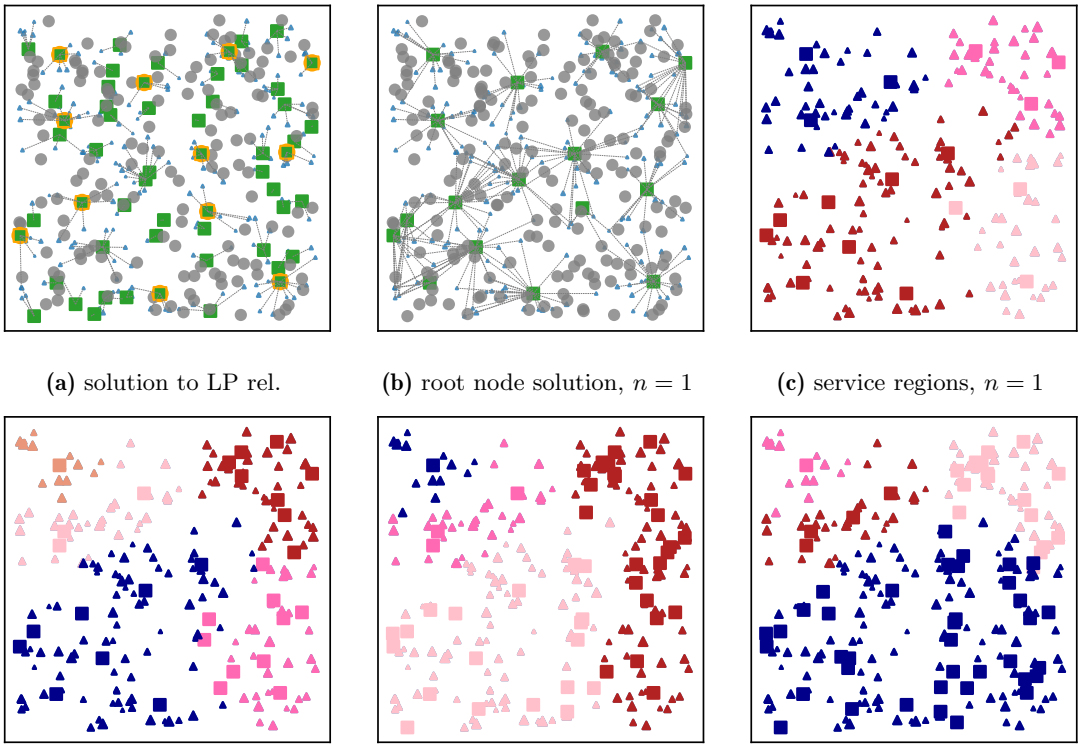
For α and θ, we essentially have all options presented in Section 3.2 to choose from. The aggregation function α determines the weight with which individual allocation decisions in any of the solutions in $\mathcal{S}$ are considered in the biclustering underlying *Regclus*. While more comprehensive functions like, e.g., α^{profit}, that weigh each allocation by their associated value in the objective value seem like a sensible choice, it might also deter the algorithm from low-profit allocations that do not contribute much to the optimal objective value but essentially distinguish the optimal from the sub-optimal solution. On the contrary, an aggregation function like $\alpha^{indicator}$ that assigns each allocation the same weight, not even considering whether it occurred in many or just a single solution, might give too much emphasis on irrelevant allocations potentially enforced by suboptimal branching constraints. Similarly, a suitable target level of coherence θ is difficult to derive up front. A larger value of θ allows for more interaction between regions and will thus produce a finer separation into potentially smaller regions.

Lastly, another question is whether it is worth updating $\mathcal{R}$ throughout the branching procedure or whether it suffices to produce one set of regions initially and work with that until the end. The advantage of updating $\mathcal{R}$ could be that more and better information on the underlying decision patterns reveal themselves throughout the search process, in particular, if only solutions that have proven themselves as advantageous (e.g., incumbents or solutions at relatively low levels of the search tree that could not yet be pruned) are considered. Furthermore, only facilities and customers that have an associated non-zero allocation in any of the considered solutions are assigned to a service region by *Regclus*. Thus, deriving regions from a relatively small set of solutions at the beginning of the tree might result in some relevant facilities not even being included in any of the service regions. However, even though *Regclus* is a very efficient procedure, each time it is invoked adds to the total time of the branching procedure. Furthermore, regional substructures are likely to be identified early on.

We perform preliminary experiments on the instances from Example A to determine sensible choices for the above parameters.

Example A 5.7 (Pattern-based regional branching) In the subsequent experiment, we focus on instances $\mathcal{P}_2$ and $\mathcal{P}_4$, which exhibit larger service regions consisting of more than one facility and that require exploring a search tree of considerable size. Figure 5.7a depicts the solution to the LP relaxation of $\mathcal{P}_4$. Each facility serves only a very small number of customers, and few customers are served by more than one facility. Using solely this solution as an input for *RegClus* will produce many extremely small regions that do not yield additional insights. Figure 5.7b displays the solution at the first node CPLEX's MIP solver performs a branching operation on, the root node. It differs from the solution to the LP relaxation in the sense that CPLEX has already applied several cuts and various other enhancements. One can see that fewer facilities operate in total, and more customers are served by more than one facility. Throughout all our experiments, we observe that in many of the solutions at the root node, several customers are served by more than one facility, even though capacities are not necessarily scarce. Figure 5.7c illustrates the regions produced by *RegClus* when taking only the solution at the root node and setting

$\alpha =$*count* and $\theta = 0.1$. Figure 5.7d-Figure 5.7f display the regions obtained when taking the solutions of the first 5, the first 10, or even the first 20 nodes as an input for *RegClus*. One can see that while different pairs of groups are joined for different sets of solutions, several service regions remain constant across all presented divisions.



(a) solution to LP rel. **(b)** root node solution, $n = 1$ **(c)** service regions, $n = 1$

(d) service regions, $n = 5$ **(e)** service regions, $n = 10$ **(f)** service regions, $n = 20$

Figure 5.7.: Bipartite graph and service regions for diff. n (α^{count}, $\theta = 0.1$, $\mathcal{P}_4$, Ex. A)

Given that the produced service regions change only minorly, we provide preliminary experiments for $\mathcal{P}_2$ and $\mathcal{P}_4$ for service regions obtained from the solutions at the root node ($n = 1$), and the first $n = 5$ nodes respectively. In our initial experiment, we consider all different functions α and vary θ between 1%, 5%, and 10%. As our variable selection strategy only affects the search process as long as variables associated with facilities assigned to a region are fractional, the rest of the selection process is performed independently of any service regions. In particular, we compare our performance to the performance of the search when using either CPLEX's default variable selection procedure or maximum infeasibility throughout.

Table 5.12 displays the solve time in seconds (including the time it took to invoke *RegClus*), the depth of the search tree as well as the number of nodes explored for different input parameters to *RegClus*. When figures are printed in bold, they outperform CPLEX's default search. The results indicate the potential of including information on the interdependencies between individual location decisions in the search process. For $\mathcal{P}_4$, the branching procedure could be shortened in most considered parameter settings. And while the settings significantly affect the performance and must be investigated further, the potential improvement is enormous. For $n = 1$, $\alpha = \alpha^{count}$, and $\theta = 0.1$, the total time to find an

$\mathcal{S}$	α	θ	$\mathcal{P}_2$ time (s)	$\mathcal{P}_2$ nodes	$\mathcal{P}_2$ depth	$\mathcal{P}_4$ time (s)	$\mathcal{P}_4$ nodes	$\mathcal{P}_4$ depth
1	sum	0.01	8.8	882	73	**46.6**	2547	**135**
		0.05	**6.3**	645	73	**34.9**	**1481**	**134**
		0.1	**6.5**	572	113	**51.3**	2328	**142**
	indicator	0.01	8.2	882	73	**45.3**	**2096**	**141**
		0.05	7.9	882	73	**45.3**	**2315**	**142**
		0.1	**6.7**	645	73	**34.8**	**1418**	**134**
	count	0.01	8.8	882	73	**43.0**	**2096**	**141**
		0.05	8.0	882	73	**38.5**	**1852**	**95**
		0.1	**6.6**	645	73	62.1	2941	**125**
	demand	0.01	**6.5**	657	73	**43.1**	**1894**	**138**
		0.05	**5.8**	**505**	**71**	**48.9**	**2275**	**134**
		0.1	**6.7**	633	74	**55.8**	2463	**116**
	profit	0.01	**6.5**	657	73	**43.6**	**1894**	**138**
		0.05	7.2	724	74	**49.0**	**2275**	**135**
		0.1	7.6	724	74	56.3	2463	116
5	sum	0.01	8.3	711	75	**35.3**	**1175**	**135**
		0.05	**6.2**	534	**72**	**54.8**	**2284**	**127**
		0.1	**6.8**	**514**	73	76.9	3160	**142**
	indicator	0.01	7.5	711	75	**46.5**	**2147**	**132**
		0.05	7.5	711	75	60.5	**2230**	**130**
		0.1	**6.2**	**482**	**70**	69.0	2833	**138**
	count	0.01	8.1	711	75	**39.4**	**1384**	**96**
		0.05	7.5	711	75	**41.0**	**1519**	**109**
		0.1	8.0	711	75	**54.5**	**2284**	**127**
	demand	0.01	8.2	949	73	**41.4**	**1498**	**103**
		0.05	7.3	669	72	**55.8**	**2103**	**136**
		0.1	8.7	767	75	57.4	**2284**	**127**
	profit	0.01	7.0	610	**72**	**44.0**	**1498**	**103**
		0.05	8.5	767	75	59.3	**2103**	**136**
		0.1	9.2	767	75	**53.0**	**2284**	**127**
CPLEX default			7.0	534	73	56.0	2326	144

Table 5.12.: Performance of regional variable selection; non-regional variables are selected according to CPLEX's default procedure ($\mathcal{P}_2$, $\mathcal{P}_4$ Ex. A)

optimal solution could be reduced by 38% from 56.0 to 34.8 seconds. Thereby, the number of nodes explored in the search process could be reduced from 2326 to 1418 (39%). In 17 out of the 30 considered parameter settings, the solve time of $\mathcal{P}_4$ could be reduced by more than 10%. For $\mathcal{P}_2$, which was significantly faster to solve in the first place, the reduction in the solve times was less significant but still visible for several configurations.

			$\mathcal{P}_2$			$\mathcal{P}_4$		
$\mathcal{S}$	α	θ	time (s)	nodes	depth	time (s)	nodes	depth
1	sum	0.01	10.8	714	**59**	67.2	1427	107
		0.05	**7.1**	**427**	**65**	95.1	2072	115
		0.1	**7.5**	**377**	**63**	81.3	1861	116
	indicator	0.01	9.8	714	**59**	**42.0**	**788**	**78**
		0.05	9.8	714	**59**	**52.7**	1026	100
		0.1	**7.4**	**427**	**65**	93.5	2033	115
	count	0.01	10.4	714	**59**	**42.3**	**788**	**78**
		0.05	10.0	714	**59**	**51.0**	934	97
		0.1	**7.6**	**427**	**65**	94.0	2005	116
	demand	0.01	**7.4**	**487**	73	86.8	1854	120
		0.05	**6.8**	**378**	**54**	71.5	1428	112
		0.1	11.3	755	**56**	**48.0**	**882**	83
	profit	0.01	**7.6**	**487**	**54**	87.2	1854	120
		0.05	**7.1**	**452**	**60**	71.7	1428	112
		0.1	**7.3**	**452**	**60**	**48.8**	**882**	83
5	sum	0.01	**7.8**	**451**	**64**	75.0	1449	107
		0.05	**6.9**	**436**	**64**	84.0	1868	114
		0.1	**7.2**	**436**	**64**	61.7	1476	93
	indicator	0.01	**7.4**	**474**	66	86.2	1973	107
		0.05	**7.2**	**474**	66	86.0	1973	107
		0.1	**7.9**	**474**	66	**44.3**	924	**82**
	count	0.01	10.2	713	66	86.2	1973	107
		0.05	10.9	713	66	75.2	1449	107
		0.1	**8.0**	711	75	56.5	1254	89
	demand	0.01	**7.5**	**451**	**64**	52.3	921	84
		0.05	**7.4**	**377**	**62**	**50.0**	996	90
		0.1	**8.0**	**377**	**62**	72.0	1653	95
	profit	0.01	**7.7**	**451**	**64**	**50.9**	921	84
		0.05	**7.4**	**359**	**57**	101.1	2011	110
		0.1	9.8	**359**	**57**	94.6	2216	129
max. infeasibility			9.7	524	66	52.0	903	87

Table 5.13.: Performance of regional variable selection; non-regional variables are selected based on max. infeasibility ($\mathcal{P}_2$, $\mathcal{P}_4$ Ex. A)

Table 5.13 displays similar results. This time, the non-regional facilities were selected based on the maximum infeasibility rule. The positive effect of considering individual regions is much less significant, in particular, for $\mathcal{P}_4$. However, it must be pointed out that for this instance, the maximum infeasibility rule without any regions performs exceptionally well. Therefore, whether or not it is significantly less valuable to consider regional variable selection with a maximum infeasibility search rule must be tested on other instances. ▲

Example A 5.7 shows the potential of the region-based variable selection rule. In the following, we perform tests on instances from Section 1.3, to evaluate different parameter settings and test the effectiveness.

5.4.2. Experimental validation: pattern-based regional branching

In the following, we evaluate the performance of the regional variable selection scheme on the instances from Section 1.3. Thereby, we restrict ourselves to those 126 instances that require at least one variable selection decision in CPLEX's state-of-the-art solver and cannot be solved directly at the root node via various pre-processing schemes.

We are particularly interested in the following three questions:

1. Is regional branching variable selection an effective means to speed up the solution process of a branching procedure?
2. How relevant is the choice of input parameters α, θ, and n for this effectiveness?
3. Is the effectiveness of regional variable selection related to properties of the problem instance?

Table 5.14 displays the average solve time in seconds, the average solve time without the time it took to invoke *RegClus*, the average number of nodes explored, the average depth of the search tree when using the regional branch variable selection scheme on those variables assigned to a service region, and CPLEX's proprietary default branch variable selection scheme procedure otherwise. The results are compared with the average values obtained when solving the problem solely with CPLEX's proprietary branch variable selection scheme. The results are displayed for different input parameters of *RegClus*, particularly different aggregation functions α, different target levels of coherence θ, and different n, the number of infeasible solutions from the start of the search tree used to determine the set of regions.

The results show that, on average, regional variable selection does not improve CPLEX's default branch variable selection procedure. Instead, the solve time increases by an average of 12.4%, including the take it takes to invoke *RegClus*, and by an average of 8.3% to perform solely the branching procedure. The number of nodes processed increases by an average of 21.3%. Solely, the depth of the search tree decreases by an average of -5.7%. However, the regional variable selection scheme was invoked and (potentially) altered the variable selection decisions in a significant proportion of the processed nodes. In each of these decisions, the inner-regional variable selection choice was performed based on maximum infeasibility branching, a variable selection scheme inferior to CPLEX's default procedure. Consequently, to evaluate the effect of regional branch variable selection, a more accurate assessment of the impact of the regional component is necessary. Therefore, in the following, we compare the performance with and without regional branch variable selection when all variable selection decisions are based on maximum infeasibility branching.

The results for maximum infeasibility branching are displayed in Table 5.15. They show an entirely different picture and confirm the effectiveness of regional branch variable selection in improving the performance of state-of-the-art branching procedures. Independently of the choice of θ and α, the total solving time could be reduced compared to using solely maximum infeasibility branching by an average of 8.6%, and an average of 11.6% when not including the time it takes to invoke *RegClus*. This is surprising, as both the number

θ	α	n	time (s)	time without *RegClus* (s)	nodes	depth
0.01	sum	1	25.0	24.4	480.8	**18.2**
		5	25.5	24.8	471.6	**18.4**
	indicator	1	24.0	23.4	477.0	**18.5**
		5	25.2	24.4	466.8	**18.0**
	count	1	23.2	22.6	475.8	**18.4**
		5	23.5	22.7	468.9	**17.8**
	demand	1	26.7	25.9	484.2	**18.7**
		5	25.9	24.5	462.6	**17.9**
	profit	1	23.9	22.9	482.7	**18.6**
		5	24.3	22.1	462.2	**18.4**
0.05	sum	1	23.7	23.2	414.7	**18.1**
		5	23.5	22.8	407.6	**18.4**
	indicator	1	24.0	23.4	494.3	**19.2**
		5	25.6	24.9	479.0	**17.8**
	count	1	24.0	23.4	490.1	**18.9**
		5	23.7	22.9	470.4	**17.6**
	demand	1	25.0	24.3	485.4	**18.3**
		5	24.1	22.7	405.1	**17.8**
	profit	1	24.5	23.6	477.1	**18.4**
		5	24.6	22.4	422.8	**18.2**
0.1	sum	1	24.6	24.1	437.9	**18.9**
		5	23.8	23.0	430.2	**19.0**
	indicator	1	24.3	23.7	470.1	**17.8**
		5	25.6	24.8	475.4	**18.0**
	count	1	24.6	24.0	486.0	**17.9**
		5	24.3	23.5	463.3	**18.2**
	demand	1	25.1	24.4	494.8	**19.0**
		5	24.0	22.5	423.7	**18.8**
	profit	1	24.4	23.6	437.9	**18.7**
		5	25.2	23.0	423.9	**18.7**
CPLEX default			21.8	-	380.3	19.4

Table 5.14.: Average performance of regional variable selection; non-regional variables are selected according to CPLEX's default procedure

of nodes processed and the depth of the search tree increase slightly. Notice, however, that not all nodes need the same time to solve, and thus the reduced solve times indicate that the procedure effectively reduces the number of "difficult" nodes.

θ	α	n	time (s)	time without *RegClus* (s)	nodes	depth
0.01	sum	1	**27.3**	**26.8**	452.4	15.2
		5	**28.1**	**27.3**	465.8	15.5
	indicator	1	**26.2**	**25.6**	456.7	15.2
		5	**28.1**	**27.3**	467.6	15.6
	count	1	**26.0**	**25.4**	455.0	15.2
		5	**27.1**	**26.3**	466.6	15.5
	demand	1	**28.4**	**27.7**	451.2	14.8
		5	**28.1**	**26.6**	451.1	14.7
	profit	1	**28.6**	**27.7**	455.1	15.3
		5	**29.5**	**27.1**	446.6	**14.4**
0.05	sum	1	**27.1**	**26.5**	461.3	15.5
		5	**28.1**	27.4	470.3	15.9
	indicator	1	**26.5**	**25.9**	463.9	15.2
		5	**28.0**	**27.3**	467.6	15.7
	count	1	**26.3**	**25.7**	462.1	15.2
		5	**27.4**	**26.6**	464.3	15.6
	demand	1	**27.0**	**26.3**	454.9	15.9
		5	**28.1**	**26.7**	459.4	14.9
	profit	1	**27.0**	**26.1**	452.6	15.3
		5	**29.6**	**27.5**	465.0	14.9
0.1	sum	1	**26.9**	**26.3**	473.4	16.0
		5	**27.5**	**26.8**	472.7	16.0
	indicator	1	**27.1**	**26.5**	466.4	15.8
		5	**26.7**	**26.0**	452.5	15.1
	count	1	**27.1**	**26.5**	466.9	15.7
		5	**27.0**	**26.3**	460.1	15.5
	demand	1	**27.1**	**26.4**	465.7	15.7
		5	**28.4**	**27.0**	471.0	15.3
	profit	1	**26.5**	**25.6**	455.4	15.6
		5	**29.7**	**27.5**	474.6	15.4
max. infeasibility			30.1	-	444.2	14.6

Table 5.15.: Average performance of regional variable selection; non-regional variables are selected based on max. infeasibility

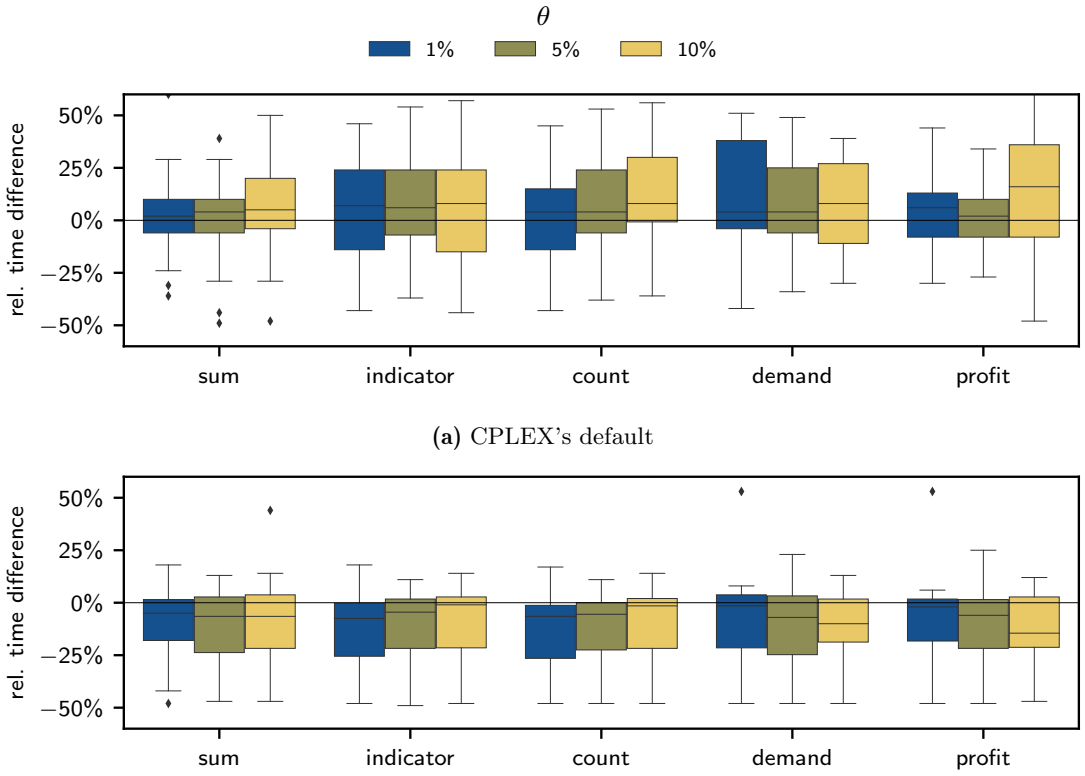
In both of the above tables, one can see that increasing the number of nodes included in $\mathcal{S}$ from $n = 1$ to $n = 5$ almost consistently resulted in an increase of solve times, both with and without including the time for *RegClus*. We observe similar trends when increasing the number of considered nodes even further to 10, 15, or 20 nodes. We conclude that including more nodes of possibly severely suboptimal solutions in *RegClus* is ineffective. For future research, we suggest including only promising solutions, e.g., new incumbents.

In the following, we evaluate the performance of the procedure on individual instances and take a closer look at the impact of different θ and α. As many of the considered instances solve extremely fast, in the following, we reduce all considerations on the 22 problem instances that take at least 5 seconds to solve or require at least 10 branching operations with CPLEX's default procedure. The solve times of the instances presented in Section 1.3 vary drastically and a concise depiction of differences requires different

scales. Therefore, for both CPLEX default branching and branching solely with maximum infeasibility branching, we determine the relative time difference between the time taken with regional branch variable selection ($time_A$) and the time taken without regional branch variable selection ($time_B$) such that

$$\text{rel. time difference} = \frac{time_A - time_B}{time_B}. \tag{5.1}$$

Figure 5.8 depicts the distribution of the relative time differences for different α and different θ. As already observable from the average values in Table 5.14 and Table 5.15, the solve time increases significantly for the majority of instances when using CPLEX's default variable selection scheme. Yet, it must be pointed out that in individual instances, the average solve time decreases significantly by up to 50% through regional variable selection. However, the solve time decreases consistently compared to maximum infeasibility branching. Thereby, a clear dominance of a certain choice of neither α nor θ is observable.



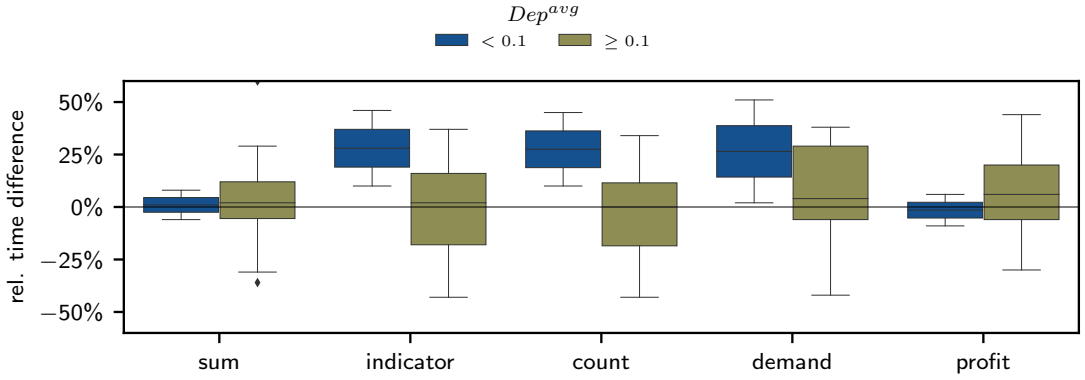
(a) CPLEX's default

(b) max. infeasibility

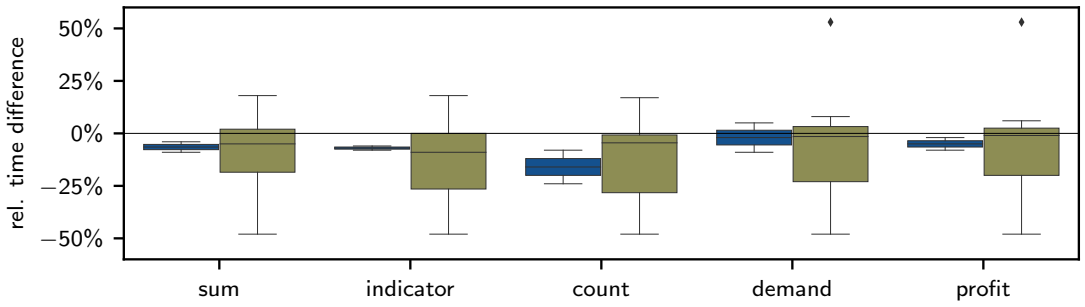
Figure 5.8.: Distribution of relative differences in solving time with and without regional branch variable selection for different θ and α, $n = 1$

Figure 5.9 depicts the distribution of the relative time differences for different α, grouped according to the previously determined average dependency values. Here, we observe a significant difference, particularly when looking at the results for CPLEX's default variable selection procedure. In particular, the solving process is slowed for instances exhibiting little to no interdependencies as indicated by $Dep^{avg} < 0.1$. Meanwhile, it is improved and slowed down in equal proportions on instances exhibiting high interdependencies,

indicating larger service regions. The results for maximum infeasibility branch variable selection draw a similar picture. The reductions in the solving time are more significant for instances exhibiting high interdependence structures.



(a) CPLEX's default



(b) max. infeasibility

Figure 5.9.: Distribution of relative differences in solving time with and without regional branch variable selection for different α and interdependence levels ($\theta = 0.01$, $n = 1$)

The results are very promising and indicate that explicitly considering regional information in the form of stronger and weaker coherence of certain areas of the facility-customer space can significantly reduce the solving time required by B&B procedures. While the scheme does not beat CPLEX's default variable selection strategy, it would be interesting to see whether an improvement could be observed when regional branching and that proprietary strategy are combined. This default strategy could then determine the next branching variable within a determined region. The regional variable selection allows for significant improvement in the solve times when maximum feasibility branching is deployed.

5.5. Conclusion

We evaluate the degree to which larger service regions affect the difficulty of individual problem instances. Larger in this context means that these regions are optimally served by more than one facility, leading to interdependencies between these location decisions. Up to this point, characteristics of instances that proved difficult to solve or decision patterns that were particularly hard to identify were rarely examined. We explored the effect the interdependence relationships of individual facilities have on both heuristic and exact methods. A particular challenge is imposed by the fact that in both, heuristic methods

and exact branching procedures, location decisions are evaluated sequentially. However, the interdependence relationships mean that facilities may not be the best choice if they serve customers alone but prove to be superior to other candidates only in combination.

The latter is a general attribute of solutions to a mixed-binary mathematical problem in which the binary part of the problem exhibits a combinatorial dynamic. The interdependencies in discrete location problems are often restricted to specific areas of the facility-customer space. The latter, thereby, does not necessarily refer to the actual locations but the implied positions given the variable unit net profit combined with the outreach of individual facilities given the tightness and profit ratio. We show that these regions can already be detected from infeasible solutions early during the search process and that their explicit consideration throughout the branching procedure can significantly increase the effectiveness of branching decisions.

We summarize the main results of the previous chapter as follows:

- Large service regions significantly affect the performance of exact and heuristic solution procedures as interdependent subsets of facilities are particularly difficult to detect. Hence, solution algorithms should be tested regarding their ability to detect these interdependent facilities and potentially include specific subroutines for this purpose.
- Service regions can already be identified from integer infeasible solutions. Using knowledge of the implied separation of the facility-customer space to search individual service regions sequentially can significantly improve the performance of branching procedures.

To the best of our knowledge, this is the first time the challenges combinatorial decision patterns pose for heuristic and exact procedures to discrete location instances have been discussed in detail. Furthermore, this is the first work that uses pattern recognition on early, potentially infeasible regions in the search tree of a branching procedure to detect interdependence relationships between individual decisions. At this point, the detected relationships were only used to improve the variable selection process. Exploring the degree to which the knowledge of these implied spatial relationships can be used to enhance different components of branching procedures can be a valuable path to explore in future research. For example, Yildiz et al. [2022] suggests the effectiveness of multi-variable branching. This means that rather than branching on an individual variable and creating two nodes that cut off its fractional value, one considers several variables in that branching decision. The authors suggest creating two nodes that cut off the fractional value of a sum of variables. This raises the question of which variables should be branched on together. In their work on decomposition branching Yildiz et al. [2022] restricts all considerations to works that imply a decomposable problem structure. Our pattern detection-based regional clustering algorithm allows for identifying a decomposition of the facility-customer space implied purely from the problem data. This is an exciting path to explore further as it would allow decomposing larger instances independently of the mathematical problem formulation but purely from characteristics implied by the data.

6. Service regions and increasing demands

Location decisions oftentimes involve high investments and are taken on a strategic level. Hence, models that explicitly consider time have received considerable attention [Melo et al., 2006, Alumur et al., 2012a, Jena et al., 2015, Cortinhal et al., 2015]. Location decisions are made over the course of a planning horizon, which Dunke et al. [2018] define as "the time-frame corresponding to available data (meaningful/trustworthy information) or the time span defined by a decision-maker for having the system fully operational and/or appropriately adapted to the circumstances". One can distinguish between continuous- and discrete-time models depending on whether or not the moments for performing changes are determined endogenously as part of the decisions or exogenously given. The vast majority of works on facility location over time assume a discrete-time model. This means that the planning horizon is divided into a discrete set of periods at the beginning of which decisions can be taken [Nickel and Saldanha da Gama, 2019]. Usually, it is assumed that these periods are of equal length and correspond to planning units such as years, quarters, or months. In the subsequent chapter, we evaluate the effects of moving away from equal-length periods to more general discretizations of the planning horizon, which requires the following definitions.

Definition 6.1. *A **breakpoint**, τ^t, denotes a point in time at which decisions can be taken.*

Definition 6.2. *A **discretization**, T, of a planning horizon that comprises a defined (time) interval, $[\tau^{min}, \tau^{max}]$, is a sequence of breakpoints such that $T := \{\tau^1, \tau^2, \dots, \tau^t, \dots\}$ with $t \in \mathbb{N}$, $\tau^{min} = \tau^1$ and $\tau^t < \tau^{t+1} < \tau^{max}$.*

As illustrated in Figure 6.1, the sequence of breakpoints induces a set of periods. Each breakpoint τ^t denotes the starting point of a period t that spans the interval $[\tau^t, \tau^{t+1})$ for $t \in 1, \dots, |T|-1$, and $[\tau^t, \tau^{max}]$ for $t = |T|$, respectively. The index set of a discretization, $\mathcal{I}(T) := \{1, 2, \dots, |T|\}$, denotes the set of periods a planning horizon is divided into. It corresponds to the indices of the consecutive periods t. The starting point of period t is the breakpoint τ^t.

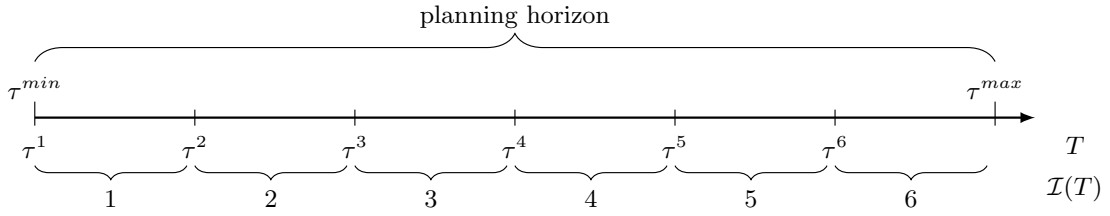


Figure 6.1.: Periods induced by a discretization T

When the planning horizon is divided into discrete periods, these periods must be linked to one another, or else the problem will fall into $|T|$ separate subproblems that optimize each period individually [Nickel and Saldanha da Gama, 2019]. The Multi-period Capacitated Facility Location Problem (with phase-in constraints) (MP-CFLP) links periods by

ensuring that once a facility has been established, it operates until the end of the planning horizon [Roodman and Schwarz, 1977]. It is a frequently used extension of the static CFLP for situations where suppliers face increasing demands. The objective of the MP-CFLP is to maximize the total profit obtained over the course of the planning horizon, the sum of the total fixed costs, and the variable profits. The binary location decision y_{it} is 1 if facility i operates in period t and 0 otherwise. The non-negative allocation decisions x_{ijt} denote the fraction of the demand of customer j served from facility i in period t. The operation of a facility for the duration of a period has a time-dependent fixed cost f_{it}. The allocation of customers to facilities incurs variable transportation costs c_{ij} and generates a variable profit r_j per unit of the customer demand d_{jt}. In each period, customers cannot be served more than their original demand (6.2), and the facility capacities q_{it} per period must not be exceeded (6.3). The location decisions of different periods are linked by the phase-in constraint (6.4), which ensures that once a facility operates, it must continue to operate until the end of the planning horizon.

MP-CFLP

$$\max \sum_t \Big[\sum_i \sum_j \quad (r_{jt} - c_{ij}) x_{ijt} d_{jt} - \sum_i f_{it} y_{it} \Big] \tag{6.1}$$

$$\text{s.t.} \sum_i x_{ijt} \quad \leq 1 \qquad j \in J,\ t \in \mathcal{I}(T) \tag{6.2}$$

$$\sum_j d_{jt} x_{ijt} \quad \leq q_{it} y_{it} \qquad i \in I,\ t \in \mathcal{I}(T) \tag{6.3}$$

$$y_{it-1} \quad \leq y_{it} \qquad i \in I,\ t \in \mathcal{I}(T) \setminus 1 \tag{6.4}$$

$$y_{it} \quad \in \{0,1\} \qquad i \in I,\ t \in \mathcal{I}(T) \tag{6.5}$$

$$x_{ijt} \quad \geq 0 \qquad i \in I,\ j \in J,\ t \in \mathcal{I}(T). \tag{6.6}$$

Switching from a static to a multi-period model increases the flexibility of decision-makers as they may adapt their location-allocation policy to varying conditions, e.g., changing customer demands. However, the price to pay is the complexity of the problem, as multiple periods make the problem increasingly difficult to solve due to the higher number of variables and constraints. Alumur et al. [2012a] quantify the gain a decision-maker receives in return for the additional computational complexity when shifting from a static to a multi-period model.

Definition 6.3. *The **Value of the Multi-period Solution**, $VMPS$, denotes the relative arithmetic difference between the optimal objective value of the multi-period problem, $z^\star_{MP}$, and the optimal objective value of its static counterpart, $z^\star_{SC}$,*

$$VMPS := \frac{z^\star_{MP} - z^\star_{SC}}{z^\star_{MP}}. \tag{6.7}$$

Example B 6.1 (Value of the Multi-period Solution (VMPS)) Consider problem instances $\mathcal{P}^4_1$ and $\mathcal{P}^4_2$ from Example B. Table 6.1 depicts the optimal objective value and the number of facilities operating in the optimal solution for the CFLP and the corresponding 4-period

MP-CFLP. Furthermore, it displays the fraction of the total customer demand served over the course of the planning horizon, further denoted by $\beta(s)$. Contrasting one's intuition, for both instances, the objective value of the multi-period model is less than that of the static model. An explanation can be derived straight from the mathematical programming formulations of the CFLP and the MP-CFLP. In the static CFLP, the capacity constraint (1.3) aggregates and thereby relaxes constraint (6.3) such that

$$(1.3) \Leftrightarrow \sum_t \sum_j d_{jt} x_{ij} \leq \sum_t q_{it} y_i \qquad \forall i \in I \nLeftrightarrow (6.3). \tag{6.8}$$

For this reason, three more facilities, 21 rather than 18, are operating in the final period of the multi-period solution to $\mathcal{P}_1^4$ compared to the static solution. Meanwhile, over the course of the planning horizon, 5% less of the total demand is being served. This results in a negative VMPS of -4.0%. For $\mathcal{P}_2^4$, one additional facility is operating in the final period of the multi-period solution. Furthermore, in the multi-period solution, 100% of the demand (0.2% more than in the static problem) are served. The VMPS for $\mathcal{P}_2^4$ is -0.4%.

	CFLP			MP-CFLP				
	$z^\star_{SC}$	$\lvert I^\star_{SC}\rvert$	$\beta(s)$	$z^\star_{MP}$	$\lvert I^\star_{MP}\rvert$	$\beta(s)$	$VMPS$	$overlap(I^\star_{SC}, I^\star_{MP})$
$\mathcal{P}_1$	2373.1	18	93%	2281.8	21	88%	-4.0%	1.0
$\mathcal{P}_2$	19803.9	7	99.8%	19709.4	8	100%	-0.4%	0.7

Table 6.1.: Performance of CFLP and 4-period MP-CFLP ($\mathcal{P}_1^4$-$\mathcal{P}_2^4$, Ex. B)

Recall that the optimal solution to instance $\mathcal{P}_1$ was predominantly comprised of independent facilities. The set of facilities operating in the solution to the multi-period model is a superset of the facilities operating in the solution to the static model. The overlap coefficient of the respective index sets is 1.0. The optimal solution to instance $\mathcal{P}_2$ predominantly comprises interdependent facilities. The operating facilities in the solution to the multi-period model differ significantly from those operating in the static solution as the overlap coefficient is only 0.7, implying that facilities operating in the optimal solution to the static instance are no longer optimal in the solution to the multi-period instances. While the possibility of the latter is long-known, the results from Chapter 2 and Chapter 3 raise the question of whether the temporal stability of individual location decisions, as well as the VMPS depend on the underlying decision patterns. ▲

A decrease in the objective value despite more accurate information being available and used in the multi-period setting is counter-intuitive. Yet, it is a direct consequence of the more accurate description of the temporal availability of capacity and occurrence of customer demands over the course of the planning horizon. Still, moving to a multi-period model not only increases the flexibility of the decision-maker but also increases the restrictions regarding when to use limited resources. This is a rarely, if at all, discussed trade-off in multi-period modeling. As each period itself can be seen as an aggregation of several smaller periods, this trade-off is not only invoked when shifting from a static to a multi-period model but whenever the number of periods the planning horizon is divided into changes. Furthermore, decision patterns in the form of service regions already

observable in the static counterpart seem to impact the potential benefit of a multi-period approach. We summarize the above aspects in the following research question:

RQ6: When is it worth explicitly considering time in the CFLP?

Several parts of the upcoming chapter have been published in Bakker and Nickel [2024].

6.1. Related work on the value(s) of the multi-period solution

Literature on facility location explicitly considering multiple periods is constantly increasing, particularly in the context of supply chain management [Jena et al., 2015, Cortinhal et al., 2015, Correia et al., 2018, Correia and Melo, 2021]. Yet, only a few authors quantify the benefit of their approach or discuss how they determine the moments in time for decision-making. In the following, we review existing work on the value of the multi-period solution. After Alumur et al. [2012b] introduced the idea, Nickel and Saldanha da Gama [2019] provide the first formal definition of the VMPS as the relative arithmetic difference between the optimal objective value of the multi-period problem and the optimal objective value of its static counterpart (compare Eq. (6.7)). But just as there is no single multi-period model to a static CFLP, the static counterpart is also not defined uniquely. While in Example B 6.1 the static model aggregates periods by summing up the parameters, another option is to choose reference values, e.g., the maximum value of each parameter over the planning horizon. Nickel and Saldanha da Gama [2019] distinguish between the weak and the strong VMPS. The weak VMPS is obtained when comparing the optimal solution of a multi-period model to the optimal solution of a static counterpart that is based on some sort of data aggregation. The strong VMPS is obtained when comparing the optimal solution to a multi-period model to that of a static counterpart for which no data aggregation has been performed and whose solution would consequently also be feasible in the multi-period context.

In the works that did examine the VMPS, the distinction between strong and weak VMPS was not made. Alumur et al. [2012b] compare the solution of a 5-period model to the solution obtained when first solving a 1-period model with the averaged parameter values and then fixing the resulting optimal location decisions in the 5-period model. On average, they observe a VMPS of 2%. Aras et al. [2015] apply the same procedure to a model for locating recycling sites and report VMPS between 7.62% and 26.75%. Tari and Alumur [2014] derive the static counterpart from the parameters' maxima and then again fix the resulting decisions in the multi-period model. Depending on the problem instance, they report values between 0 and 33%. Kchaou Boujelben et al. [2016] present a multi-period model in which only the allocation decisions are allowed to be altered over the planning horizon and location decisions are static. The authors report a VMPS between 0.5% and 1% and conclude that as computation times increase significantly when shifting to a multi-period model, dynamic customer assignments do not yield sufficient benefit in their case. Marković et al. [2017] compute the VMPS in the context of an evasive flow capturing model. The authors repeatedly solve the static counterpart over the course of the planning horizon based on the aggregated (summed) parameters that comprise a

subset of the periods. The authors observe that the VMPS increases with a steeper surge in expected flow intensities per period, leading to the assumption that stronger changes in the data may lead to greater values of the VMPS. Marín et al. [2018] report the average VMPS observed over multiple instances of a stochastic coverage location problem. The authors report the VMPS in absolute terms but observe that it is approximately five times as high as the added benefit obtained from deploying a stochastic instead of a deterministic approach as measured by the value of the stochastic solution.

All comparisons are made between multi-period and static models. The number of periods or the granularity of planning decisions is not investigated. To the best of our knowledge, the first and only authors to acknowledge that the number of periods a planning horizon is divided into might significantly alter the quality and structure of the resulting solution are Albareda-Sambola et al. [2009] in the context of multi-period incremental service facility location problem. The authors do not use the VMPS but make a qualitative statement and propose an ex-post what-if analysis to determine the number of periods that leads to an equilibrium between costs and quality of service.

The review of literature shows that, up to this point, RQ5 is insufficiently addressed. There is little insight into the conditions under which a multi-period modeling approach is valuable. The reported values for the VMPS differ massively and range between 0 to 33%. This can partially be attributed to the choice of different static counterparts. We discuss their implications regarding the interpretation of the VMPS in Section 6.2. Nevertheless, even with consistent static counterparts, the VMPS still depends on the underlying problem instance. While it is intuitive to assume that the degree of change in the time-dependent parameters of the instance is a good indicator of the degree to which a decision-maker may benefit from a multi-period model, examples can be found where this relationship cannot be observed. Therefore, in Section 6.3, we discuss the implications of discretizing the planning horizon into distinct periods in a restricted problem setting. In Section 6.4, we use these insights to anticipate the effect of multi-period modeling based on the characteristics of the problem data. We summarize our main findings in Section 6.5.

6.2. Static counterparts and their corresponding VMPS

A unique definition of a static counterpart to a multi-period model does not exist. Different static counterparts may have different optimal solutions with different optimal objective values, leading to different VMPS. In the following, we discuss several static counterparts and their corresponding VMPS, as well as their interpretation in terms of quantifying the aforementioned trade-off between added flexibility in decision-making and tighter constraints regarding the utilization of restricted resources. We illustrate all static counterparts at the MP-CFLP, a basic example of a multi-period, resource-constrained location model.

6.2.1. The value of added flexibility

Nickel and Saldanha da Gama [2019] define the VMPS derived from a static counterpart in which no aggregation has been performed on the data as a strong VMPS, further denoted

by $VMPS_s$. The fact that data is not being aggregated implies that any feasible solution, particularly any optimal solution, to the static counterpart will also be a feasible solution of the multi-period model. A straightforward way to derive a static counterpart that fulfills this criterion is to add constraints to the multi-period model that ensure that decision variables take on the same value during all periods of the planning horizon.

For the MP-CFLP, we restrict this idea to the strategic location decisions. Preventing the allocation decisions from fully exploiting available capacities in each period would lead to very bad optimal objective values for the static counterpart while reflecting a situation of little interest to a decision-maker. Thus, to obtain the static counterpart corresponding to the $VMPS_s$ we add the following constraint to the MP-CFLP

$$y_{it} = y_{it-1} \qquad \forall i \in I,\ t \in \mathcal{I}(T) \setminus 1. \tag{6.9}$$

The resulting static counterpart suffers from the aforementioned time-induced restrictions on the utilization of the facilities' capacities. At the same time, it cannot adapt location decisions to changing parameters. Consequently, the relative difference between the optimal objective value of this static counterpart and the optimal objective value of the MP-CFLP perfectly captures the benefit resulting from added flexibility. Both models share the same objective value. The set of feasible solutions to the MP-CFLP, $M_{MP} := \{(x, y) | (6.2) - (6.6)\}$, is a super-set of the set of feasible solutions of the described static counterpart, $M_{SC-s} := \{(x, y) | (6.2) - (6.6), (6.9)\}$. It follows that the optimal objective value of the static counterpart, $z^\star_{SC-s}$, will always be less or equal to the optimal objective value of the multi-period model, $z^\star_{MP}$. This implies that $VMPS_s \geq 0$.

6.2.2. The value of temporal information

Alumur et al. [2012b] and Aras et al. [2015] use a two-step approach to determine the static counterpart. First, the authors first solve a 1-period model in which all parameters are set to their average values over the planning horizon. The model is then solved, and the optimal location decisions are stored. Second, the location decisions are fixed to these values in the multi-period model. The resulting problem is solved to obtain the static counterpart objective. Consequently, the solution found with this static counterpart will also be feasible in the multi-period model as no aggregation is performed on the data. Yet, the location decisions were obtained from a model using averaged values. The approach, therefore, combines elements of a strong and a weak VMPS. We refer to the resulting value as hybrid VMPS and denote it by $VMPS_{w/s}$.

When applying the above to the MP-CFLP, we obtain the location decisions by solving the CFLP and setting D_j, F_i, and Q_i to the average values of d_{jt}, f_{it}, and q_{it} over the planning horizon. We denote the resulting location decisions by $y^\star_{avg}$. Then, we solve a static counterpart for which we add the following constraint to the MP-CFLP

$$y_{it} = y^\star_{avg_i} \qquad \forall i \in I,\ t \in \mathcal{I}(T). \tag{6.10}$$

Again, the solutions of the static counterpart adhere to all additional restrictions im-

plied by the multi-period modeling while missing the added flexibility. Furthermore, in this case, location decisions were made without taking the additional restrictions induced by multiple periods into account. The 1-period, average-based model does not capture the different ratios of total supply and total demand occurring over the course of the planning horizon. Meanwhile, constraint (6.10) contains constraint (6.9). This implies that the set of feasible solutions of the static counterpart to the $VMPS_{w/s}$, $M_{SC-w/s} := \{(x, y)|(6.2) - (6.6), (6.10)\}$, is a subset of M_{SC-s}. It follows that the associated objective value of the optimal solution, $z_{SC-w/s^\star}$, is less or equal $z^\star_{SC-s} \leq z^\star_{MP}$, and therefore, $VMPS_{w/s} \geq VMPS_s \geq 0$.

When is this inequality strict? For $z^\star_{SC-w/s}$ to be strictly greater than $z^\star_{SC-s}$, the set of location decisions $y^\star_{avg}$ must not constitute an optimal solution in the static counterpart for the $VMPS_s$. This implies that making location decisions based on average values without considering the time-induced restrictions leads to location decisions that are suboptimal once these restrictions are taken into account. Therefore, the difference between $z^\star_{SC-w/s}$ and $z^\star_{SC-s}$ can also be interpreted as the price of not knowing the temporal development of the parameters. The absolute difference of the corresponding values of the multi-period solution, $\left(VMPS_{w/s} - VMPS_s\right) \geq 0$, may also be interpreted as the value of temporal information. This implies that the $VMPS_{w/s}$ itself overestimates the added benefit resulting from the additional flexibility in the decision-making as part of it does not come from added flexibility but from the inclusion of additional information in the model.

It is not difficult to find conditions under which the value of temporal information is significantly greater than 0. But are there also situations in which $VMPS_s = VMPS_{w/s}$ and nothing is to be gained from solely including temporal information? In the following, we will show that this is exactly the case when capacities do not restrict the optimal solution, and one, therefore, would not benefit from explicitly modeling their more restrictive bounds. This is, i.e., the case when $q_{it} \geq \sum_j d_{jt}$ for all $i \in I$ and $t \in T$. According to Leung and Magnanti [1989], such a capacitated facility location model equals an uncapacitated facility location model. The capacity constraints, in particular constraint (1.3) in the CFLP and constraint (6.3) in the static counterpart for the $VMPS_s$, can be replaced by

$$x_{ij} \leq y_i \qquad \forall i \in I, \tag{6.11}$$

and

$$x_{ijt} \leq y_{it} \qquad \forall i \in I,\ t \in \mathcal{I}(T), \tag{6.12}$$

respectively. Given this replacement, the two models will have the same optimal objective value if there exists an optimal solution of the static counterpart to the $VMPS_s$, in which the allocation decisions remain the same over the planning horizon. Thus, the following

constraint holds

$$x_{ijt-1} = x_{ijt} \qquad \forall i \in I, \ , \in J, \ t \in \mathcal{I}(T). \tag{6.13}$$

Such a solution would then also be a feasible solution of the CFLP.

For the uncapacitated facility location problem, there always exists an optimal solution with $x_{ij} \in \{0, 1\}$ for all $i \in I, j \in J$ as it will always be an optimal allocation policy to serve each customer fully from the most profitable operating facility. In consequence, if the unit transport costs and profits are time-invariant and there are no capacity restrictions, it is always optimal to allocate each customer to the same, most profitable facility across all periods. In conclusion, we emphasize that in the uncapacitated facility location model, knowledge about multi-period developments without the ability to adapt location decisions accordingly does not yield any benefit. Meanwhile, in two-stage stochastic programming, it is common to assume that only the allocation decisions can be changed over time after uncertainty is disclosed. The above implies that it does not yield any benefit to switch from a two-period to a multi-period two-stage stochastic program unless the capacity restrictions of the underlying problem instance are binding.

6.2.3. The value of moving to a different discretization

In the introductory Example B, we compare the optimal solution of the MP-CFLP with the optimal solution of the CFLP to determine the VMPS. The latter is a 1-period version of the multi-period model in which all parameters are summed over the planning horizon. Evidently, any optimal location-allocation policy will also be optimal in a model in which parameters are averaged and vice versa. But, in contrast to the computation of the $VMPS_{w/s}$, we did not proceed to fix those location decisions in the multi-period model. Instead, we interpreted the CFLP itself as the static counterpart. The solution of the CFLP might not be a feasible solution to the multi-period problem as it may violate the capacity restrictions of individual periods. Therefore, the resulting weak VMPS, further denoted by $VMPS_w$, captures both aforementioned aspects of shifting to a multi-period model: the benefit of added flexibility in the decision-making and the added costs resulting from additional restrictions. Which of these two effects has a stronger impact on the objective value is not clear and will be further examined in Section 5.3. Consequently, neither one of the sets of feasible solutions to the CFLP and the MP-CFLP is a subset of the other, and the $VMPS_w$ may be smaller or greater than 0.

The $VMPS_w$ compares the multi-period model to its aggregated, 1-period static counterpart. However, the effect of switching to a multi-period model and thereby dividing the planning horizon into discrete periods also depends on the chosen discretization T. Each shift from a coarser to a more granular discretization resembles the shift from a static to a multi-period problem in each period individually. Thus, we generalize $VMPS_w$ to evaluate the effect of shifting from one discretization T^1 to a discretization T^2 and define

$$VMPS\Delta_{T^1}^{T^2} := \frac{z^\star_{T^2} - z^\star_{T^1}}{z^\star_{T^2}}, \tag{6.14}$$

whereby $z_T^\star$ denotes the optimal objective value of the multi-period problem corresponding to the discretization T. In the following, if T is a scalar, this implies a division into $|T|$ periods of equal length.

Example B 6.2 (Weak, strong, and hybrid VMPS) Table 6.2 displays the different values for the VMPS for multi-period models corresponding to instances $\mathcal{P}_1^4$ and $\mathcal{P}_2^4$ from Example B. As before, we assume that in the static instances, all parameters are aggregated (summed) over the planning horizon, capacities and costs are time-invariant, and the demand of each customer doubles over the course of the planning horizon, increasing by constant steps each period. We assume all periods are of equal length, implying that the first four periods of the 8-period model correspond to the first two periods of the 4-period model and the first period of the 2-period model, and so on.

For all instances, the hierarchy $VMPS_{ws} \geq VMPS_s \geq VMPS_w$ holds. It is striking that the $VMPS_{ws}$ is extremely high for the multi-period instances of $\mathcal{P}_1^4$, ranging between 60% and 80%. In the corresponding static counterparts, data was averaged over the planning horizon; thus, location decisions were taken without anticipating that in the multi-period setting, the demand in later periods significantly exceeds capacities. This is because several of the facilities only become profitable later on in the planning horizon. While the multi-period problems operate between 12 and 14 facilities in period 1, the static counterpart problem operates 18. These facilities incur high fixed costs despite still serving less demand over the course of the planning horizon, as their capacities do not suffice in the final periods.

We see that neither the $VMPS_s$ nor the $VMPS_w$ increase or decrease monotonically in the number of periods. For $\mathcal{P}_1^4$, we observe that while the 2-period model performs worse than the static model and the 4-period model performs worse than the 2-period model, the performance improves again when comparing the 8-period model with the 4-period model. This non-monotonicity will be further discussed in the upcoming Section 5.3. Furthermore, even though both instances face exactly the same influx in customer demand, there is a notable difference in the value of multi-period modeling. We will discuss which characteristics drive this difference in Section 5.4.

	$\|T\|$	$z_{MP}^\star$	$VMPS_{w/s}$	$VMPS_s$	$VMPS_w$	$VMPS\Delta_{\|T\|/2}^{\|T\|}$
$\mathcal{P}_1^4$	1	2373.1	-	-	-	-
	2	2304.9	62.4	3.8	-3.0	-3.0
	4	2281.8	81.3	10.9	-4.0	-1.0
	8	2296.5	66.3	9.8	-3.3	0.6
$\mathcal{P}_2^4$	1	19803.9	-	-	-	-
	2	19748.8	11.7	5.6	-0.3	-0.3
	4	19709.4	11.4	7.5	-0.4	-0.2
	8	19581.3	11.0	7.2	-1.1	-0.7

Table 6.2.: VMPS derived from different static counterparts for the 2-, 4-, and 8-period model ($\mathcal{P}_1^4$-$\mathcal{P}_2^4$, Ex. B)

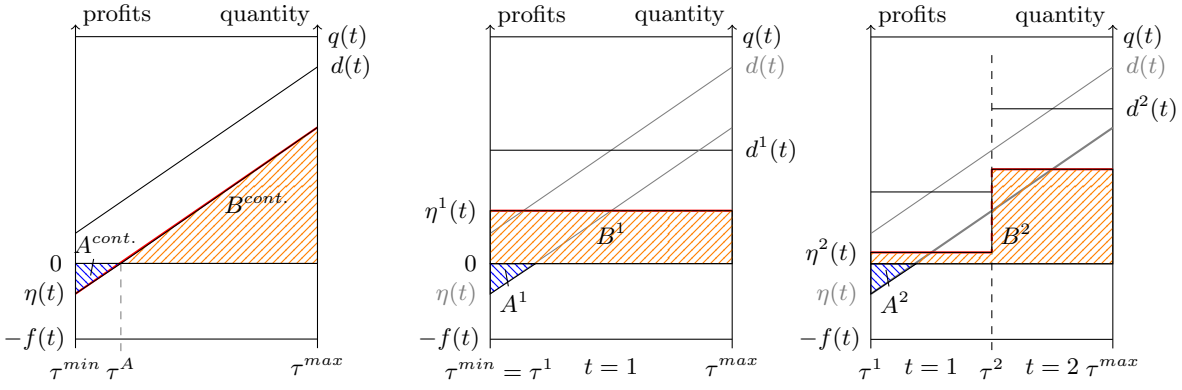
▲

We summarize that the different static counterparts and their corresponding VMPS require different interpretations. They allow for a comprehensive evaluation of the effect of multi-period modeling and separately quantify

- the value of added flexibility in decision-making ($VMPS_s$),
- the value of modeling parameter developments even without the ability to adapt decisions accordingly ($VMPS_{w/s} - VMPS_s$),
- and the value of moving from one discretization of the planning horizon to another ($VMPS\Delta_{T_1}^{T_2}$).

6.3. Effective, restrictive and obsolete breakpoints

Depending on the choice of a static counterpart, the effect of the multi-period model is evaluated differently. Nevertheless, this does not answer whether the mere fact that parameters change throughout the planning horizon suffices to expect a positive effect of a multi-period model. It neither gives insights into determining a suitable number of periods that allows a decision-maker to capture this potential benefit effectively. Under the theoretical assumption of perfect data availability, increasing the number of breakpoints can be seen as an approximation of an underlying continuous-time problem. When the induced number of periods converges to infinity, the optimal objective value converges to that of a continuous-time problem. The following demonstrates that this convergence is not monotone, even in very restricted problem settings. Hence, predicting the effect of adding more breakpoints on the objective value in general problem instances is impossible.



(a) $|T| \to \infty$ **(b)** $|T| = 1$ **(c)** $|T| = 2$

Figure 6.2.: Temporal aggregation in an uncapacitated one-facility-one-customer problem

Consider a one-facility-one-customer problem for a planning horizon spanning the interval $[\tau^{min}, \tau^{max}]$. The fixed cost and the facility's capacity are constant functions of time denoted by $f(t)$ and $q(t)$, respectively. The customer's demand, $d(t)$, is the only parameter that changes over time and increases monotonically in t. For now, assume that the demand $d(t)$ does not exceed the capacity $q(t)$ during the planning horizon, leading to an essentially uncapacitated problem. Unit costs (c) and profits (r) are time-invariant. The net profit $\eta(t)$ can be derived as a function of t with

$$\eta(t) = (r - c)\, d(t) - f(t). \tag{6.15}$$

The situation is depicted in Figure 6.2a. The left axis denotes costs and profits, and the right axis denotes quantities of demand and capacity. It is optimal to open the facility at time

$$\tau^A := \min\left\{\tau \in [\tau^{min}, \tau^{max}] \;\middle|\; \eta(\tau) \geq 0\right\}, \tag{6.16}$$

as soon as variable profits exceed the fixed costs. The optimal objective value of the continuous-time problem is

$$z^\star_{cont.} = \int_{\tau^A}^{\tau^{max}} \eta(t)dt := B^{cont.}. \tag{6.17}$$

Figure 6.2b depicts the situation when the continuous-time setting is represented by a static, single-period problem. The only point in time for decision-making is $\tau^1 = \tau^{min}$, the beginning of the planning horizon. The demand function $d^1(t)$ is a constant function set to the average value of $d(t)$ throughout the planning horizon. As the resulting constant profit, $\eta^1(t)$, is greater than zero, the optimal opening policy is to operate the facility across the planning horizon. However, this implies it also operates throughout the time interval $[0, \tau^A)$. Consequently, it incurs fixed costs during an interval during which the incoming profits do not fully offset them. The amount of these fixed costs accumulates to

$$A^1 = A^{cont.} = \int_{\tau^{min}}^{\tau^A} -\eta(t)dt. \tag{6.18}$$

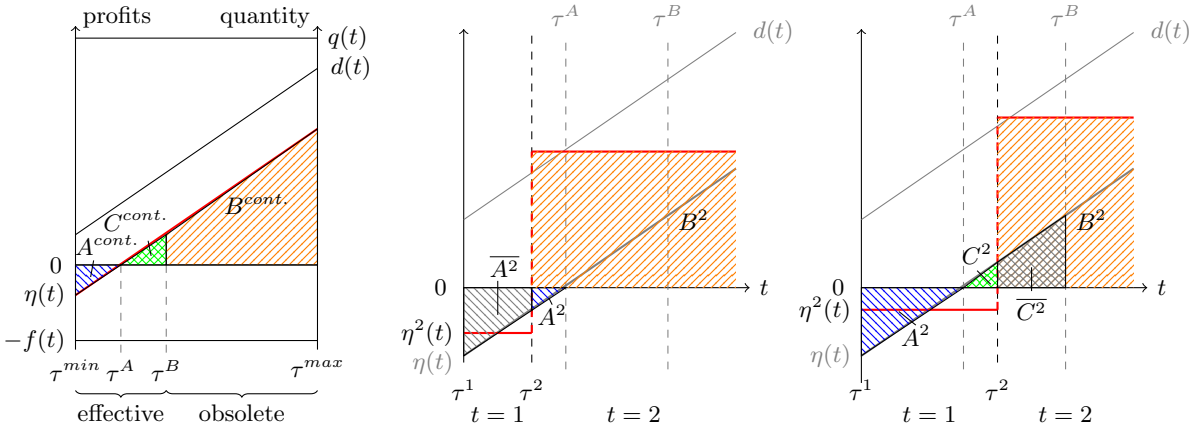
The profit a decision maker obtains from the single-period model is $B^1 = B^{cont.} - A^1$. In Figure 6.2c, the planning horizon is split into two periods of equal length. The demand function $d^2(t)$ is a piece-wise constant function, valuing the average of $d(t)$ for the respective periods. Operating the facility in both periods is optimal as the total attainable profit already exceeds the total incurred fixed costs in the first period. This, however, also implies that the objective value does not improve compared to the static single-period model. $A^2 = A^1$ implies that $B^2 = B^1$. This illustrates that while the temporal aggregation induces a loss in terms of profits, the time at which the planning horizon is divided determines whether that loss can be reduced.

6.3.1. Effective and obsolete breakpoints

We distinguish between effective and obsolete breakpoints.

Definition 6.4. *A breakpoint, τ^t, is **effective** if the optimal objective value of the multi-period problem corresponding to the discretization $T^2 = T^1 \cup \tau^t$ is strictly greater than the optimal objective value of the multi-period problem corresponding to the discretization T^1 with $\tau^t \notin T^1$ and thus $VMPS\Delta^{T^2}_{T^1} > 0$.*

Definition 6.5. *A breakpoint, τ^t, is **obsolete** if the optimal objective value of the multi-period problem corresponding to the discretization $T^2 = T^1 \cup \tau^t$ equals the optimal objective value of the multi-period problem corresponding to the discretization T^1 with $\tau^t \notin T^1$ and thus $VMPS\Delta^{T^2}_{T^1} = 0$.*



(a) $|T| \to \infty$ **(b)** $\tau^2 \in (0, \tau^A)$ **(c)** $\tau^2 \in (\tau^A, \tau^B)$

Figure 6.3.: Effective and obsolete breakpoints in a one-facility-one-customer problem

An obsolete breakpoint increases the size of the problem, but nothing can be gained from it.

In the considered one-facility-one-customer setting, breakpoints are effective if, and only if, they lie in the interval $[0, \tau^B]$ (Figure 6.3a). Thereby, τ^B denotes the point in time after which the incurred fixed costs ($A^{cont.}$) are offset by additional profits ($C^{cont.}$), such that

$$\tau^B := \max \left\{ \tau \in [\tau^{min}, \tau^{max}] \;\middle|\; \int_{\tau^{min}}^{\tau^A} -\eta(t)dt \geq \int_{\tau^A}^{\tau} \eta(t)dt \right\}. \tag{6.19}$$

In particular, the effect of dividing the planning horizon into two periods is based on the position of τ^2, the starting point of period 2, with respect to τ^A and τ^B such that if

$\tau^2 \in (\tau^{min}, \tau^A)$**:** Operating the facility in period 1 is not profitable, and the facility is opened only in period 2. The benefit of the multi-period approach equals the fixed cost overhead in period 1 that can be avoided:

$$\overline{A^2} = \int_{\tau^{min}}^{\tau^2} -\eta(t)dt, \tag{6.20}$$

such that $B^2 = B^1 + \overline{A^2}$, thus $B^2 > B^1$. Consequently, the $VMPS_w$ is positive, and the breakpoint is effective (Figure 6.3b).

$\tau^2 \in (\tau^A, \tau^B)$**:** Operating the facility in period 1 is still not profitable. In the optimal solution, it operates only in period 2. However, contrary to before, the entire fixed cost overhead $A^{cont.}$ can be avoided. Meanwhile, as $\tau^2 > \tau^A$, part of the profit that could be obtained if the facility was opened earlier cannot be attained, particularly the amount

$$C^2 = \int_{\tau^A}^{\tau^2} \eta(t)dt. \tag{6.21}$$

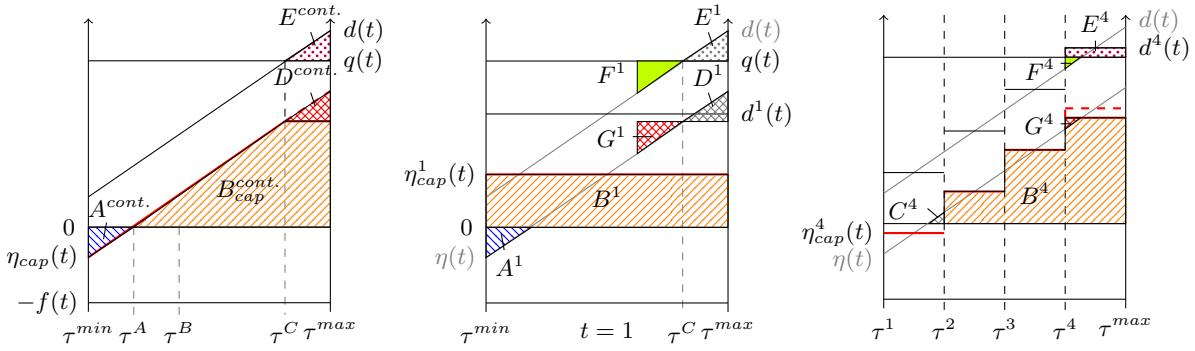
Consequently, the attainable profit is $B^2 = B^{cont.} - C^2$ with $C^2 + \overline{C^2} = C^{cont.} = A^{cont.}$. As $B^1 = B^{cont.} - A^{cont.}$ it holds that $B^2 > B^1$ and $VMPS_w > 0$ (Figure 6.3c).

$\tau^2 = \tau^A$**:** This is the optimal point in time for opening the facility: neither fixed costs can be reduced, nor profits increased and $B^2 = B^{cont.}$.

$\tau^2 \geq \tau^B$**:** Dividing the planning horizon in the interval $[\tau^B, \tau^{max}]$ is obsolete. In the resulting optimal solution, the facility operates throughout the planning horizon as the net profit in period 1 is greater than zero. Consequently, $C^2 = C^{cont.} = A^{cont.}$ and $B^2 = B^1 = B^{cont.} - A^{cont.}$, thus $VMPS_w = 0$.

Dividing the planning horizon into periods is only effective if the time interval during which it is not profitable to operate the facility is separated from a time interval in which it is.

6.3.2. Restrictive breakpoints



(a) $|T| \to \infty$ **(b)** $|T| = 1$ **(c)** $|T| = 4$

Figure 6.4.: Temporal aggregation in a capacitated one-facility-one-customer problem

Now, let there be a moment after which the demand of the customer exceeds the facility's capacity (Figure 6.4a),

$$\tau^C := \min \left\{ \tau \in [\tau^{min}, \tau^{max}] \;\middle|\; d(\tau) \geq q(\tau) \right\}. \tag{6.22}$$

The profit function must be modified as not necessarily the full demand but the minimum of demand and capacity is served

$$\eta^{cap}(t) = (r_j - c_{ij}) \min\{d(t), q(t)\} - f(t). \tag{6.23}$$

After τ^C, not all demand can be served anymore. In particular,

$$E^{cont.} = \int_{\tau^C}^{\tau^{max}} d(t) - q(t) dt, \tag{6.24}$$

cannot be satisfied. In consequence, the associated profit

$$D^{cont.} = (r - c) \int_{\tau^C}^{\tau^{max}} d(t) - q(t) dt = (r_1 - c_{11}) E^{cont.}, \tag{6.25}$$

cannot be attained and the profit in the continuous-time capacitated problem is $B^{cont.}_{cap} = B^{cont.} - D^{cont.}$.

Figure 6.4b illustrates the associated 1-period problem. Ignoring the temporal developments and simply looking at the aggregated parameters implies that the average demand is below the average capacity, and hence, the 1-period problem assumes that all demand can be served, such that

$$B^1_{cap} = B^{cont.} - A^1 + D^1. \tag{6.26}$$

Consequently, D^1 is the benefit that arises from implicit warehousing. In an aggregated time setting, it is implicitly assumed that inventory can be held throughout a period of time without incurring additional costs. As Figure 6.4c depicts, this benefit diminishes when the planning horizon is divided, e.g., into four periods, particularly when breakpoints move towards the point in time τ^C. In period 4, the aggregated demand exceeds the capacity, and the demand E^4,

$$E^4 = \int_{\tau^C}^{\tau} d^4(t) - q(t)dt. \tag{6.27}$$

cannot be served. However, the amount F^4,

$$F^4 = \int_{\tau^4}^{\tau^C} q(t) - d(t)dt = E^{cont.} - E^4, \tag{6.28}$$

can implicitly be pre-produced such that the total profit in the 4-period model is $B^4_{cap} = B^{cont.} - C^4 + G^4$ with $G^4 = (r_j - c_{ij})E^4$ and $G^4 < G^1 = D^1$.

Figure 6.4b illustrates the corresponding single-period problem. Ignoring the temporal developments and simply looking at the aggregated parameters implies that the average demand is below the average capacity throughout the planning horizon. Hence, in the single-period problem, all demand can be served, such that $B^1_{cap} = B^{cont.} - A^1 + D^1$, with $D^1 = D^{cont.}$. Consequently, D^1 is the benefit that arises from implicit warehousing.

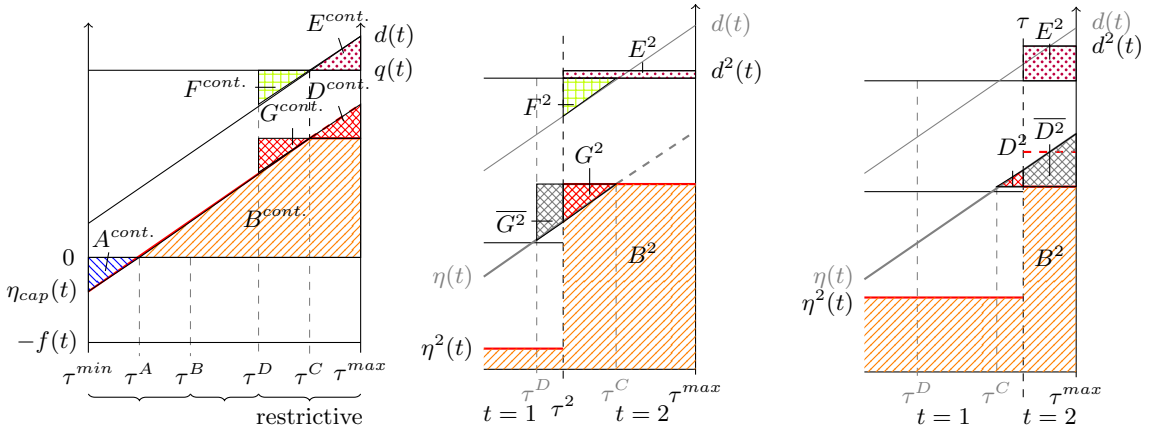
In order to fully attain this benefit, implicit pre-production must start at the latest at the point in time τ^D, where the accumulated unused capacity before τ^C is as large as the missing capacity after τ^C, such that

$$\tau^D := \max\left\{ \tau \in [\tau^{min}, \tau^{max}] \;\middle|\; \int_{\tau^C}^{\tau^{max}} d(t) - q(t)dt \leq \int_{\tau}^{\tau^C} q(t) - d(t)dt \right\}. \tag{6.29}$$

In consequence, any breakpoint within the interval (τ^D, τ^{max}) restricts the ability to benefit of implicit warehousing (Figure 6.5a).

Definition 6.6. *A breakpoint τ^t is **restrictive** if the optimal objective value of the multi-period problem corresponding to the discretization $T^2 = T^1 \cup \tau^t$ is strictly smaller than the optimal objective value of the multi-period problem corresponding to the discretization T^1 with $\tau^t \notin T^1$ and thus $VMPS\Delta^{T^2}_{T^1} < 0$.*

In particular, for a two-period model, the position of τ^2 relative to τ^C and τ^D affects the attainable profit as follows:



(a) $|T| \to \infty$ **(b)** $\tau^2 \in (\tau^D, \tau^C]$ **(c)** $\tau^2 \in (\tau^C, \tau^{max})$

Figure 6.5.: Restrictive breakpoints in a one-facility-one-customer problem

$\tau^2 \in (\tau^D, \tau^C]$: The capacity in period 2 is not sufficient to serve all demand in that period. However, it is possible to partially pre-produce at the beginning of that period. With

$$F^2 = \int_{\tau^2}^{\tau^C} q(t) - d(t) dt, \tag{6.30}$$

and $G^2 = (r-c)F^2$ the additional attainable profit. With $D^{cont.} = G^{cont.} = \overline{G^2} - G^2$, we obtain that $B^2 = B^{cont.}_{cap} + G^2 = B^1 - D^{cont.} + G^2 = B^1 - \overline{G^2}$ and thus $B^2 < B^1$ and $VMPS_w < 0$ (Figure 6.5b).

$\tau^2 \in (\tau^C, \tau^{max})$**:** Throughout period 2, the demand exceeds the total capacity. The excess demand in this period cannot be served. However, as $\tau^2 > \tau^C$ the demand exceeds the capacity already at the end of period 1 when implicit warehousing is possible. Applying the same logic as before yields that $B^2 = B^{cont.}_{cap} + D^2 = B^1 - \overline{D^2}$ (Figure 6.5c), with

$$D^2 = (r-c) \int_{\tau^C}^{\tau^2} d(t) - q(t) dt. \tag{6.31}$$

Consequently, for the one-facility-one-customer problem with constant costs, profits and capacities, and monotonically increasing demands, the planning horizon can be divided into three potentially overlapping intervals of interest: effective, obsolete, and restrictive. Whether or not an additional breakpoint in one of these intervals results in an improvement or deterioration of the objective value depends on the position of other existing breakpoints. In the following, we discuss what this implies for the VMPS.

6.3.3. VMPS

There are two critical moments during the planning horizon: τ^A, the point in time at which the demand suffices such that the facility can operate profitably, and τ^C, the point in time at which the demand exceeds capacity. For any multi-period model, the position of the breakpoints relative to these two moments determines its effectiveness. Let τ^t denote

the starting point of period t. Further, let $t(\tau^A)$ denote the period that comprises the point in time τ^A, such that

$$t(\tau^A) := \max\left\{t \in \mathcal{I}(t) \mid \tau^t \leq \tau^A\right\}. \tag{6.32}$$

Similarly, let $t(\tau^C)$ denote the period comprising τ^C,

$$t(\tau^C) := \max\left\{t \in \mathcal{I}(t) \mid \tau^t \leq \tau^C\right\}. \tag{6.33}$$

With that, we generally define $A(T)$, $C(T)$, $D(T)$, and $G(T)$ as follows:

$$A(T) = \int_{\tau^{t(\tau^A)}}^{\tau^A} -\eta_{cap}(t)dt, \tag{6.34}$$

$$C(T) = \int_{\tau^A}^{\tau^{t(\tau^A)+1}} \eta_{cap}(t)dt, \tag{6.35}$$

$$G(T) = (r-c)\int_{\tau^{t(\tau^C)}}^{\tau^C} q(t) - d(t)dt, \tag{6.36}$$

$$D(T) = (r-c)\int_{\tau^C}^{\tau^{t(\tau^C)+1}} d(t) - q(t)dt. \tag{6.37}$$

With the above, we can set the profit obtained by the T-period model in relation to the profit obtained by the continuous-time model via

$$B(T) = B^{cont.} - \min\{A(T), C(T)\} + \min\{D(T), G(T)\}. \tag{6.38}$$

If $A(T) < C(T)$, then it is profitable to open the facility in period $t(\tau^A)$ as the profit attainable during that period outweighs the fixed costs. Compared to the continuous-time setting, the fixed costs $A(T)$ are incurred additionally and must be subtracted to obtain $B(T)$. If $A(T) > C(T)$, then it is not profitable to operate the facility in period $t(\tau^A)$, and compared to the continuous-time setting, the profit $C(T)$ is not obtained. The second minimum follows a similar logic.

Expressing $B(T)$ relative to the profit attainable in a continuous-time setting allows us to express the $VMPS_s$ and the $VMPS_w$ as follows. The profit attainable by the strong static counterpart of a multi-period problem corresponding to a discretization T is

$$B^1_{SC_s} = B^{cont.} - \min\{A^1, C^1\} + \min\{D(T), G(T)\}, \tag{6.39}$$

as due to the explicit modeling of all capacity constraints the ability to hold implicit warehousing is the same as in the multi-period problem. Therefore, we have that

$$VMPS_s = \frac{B^1_{SC_s} - B(T)}{B(T)} \iff \frac{1}{B(T)} \underbrace{\left(\min\{A^1, C^1\} - \min\{A(T), C(T)\}\right)}_{\geq 0}. \tag{6.40}$$

By definition, in any single-period model, $t(\tau^A) = 1$ and $\tau^1 = \tau^{min}$ which, in turn, is less or equal $t(\tau^A)$ in the $|T|$-period model. Thereby $A(T) \leq A^1$ and their difference is upper

bounded by $A^{cont.}$. Similarly, it holds that $C(T) \leq C^1$ and thus $VMPS_s \geq 0$.

Similarly, we have that $D(T) \leq D^1$ and $G(T) \leq G^1$ and for the $VMPS_w$ we obtain

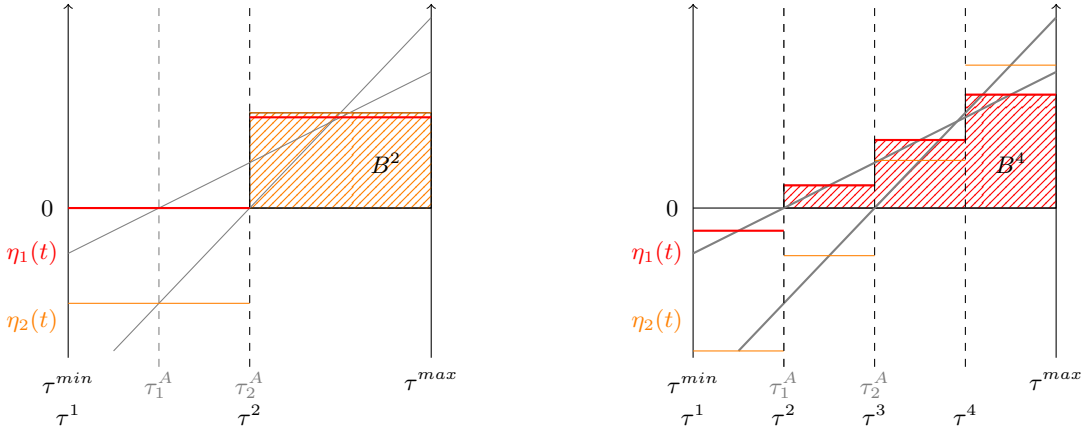
$$VMPS_w = \frac{B^1 - B(T)}{B(T)} \tag{6.41}$$
$$\iff \frac{1}{B(T)}\big(\underbrace{\min\{A^1, C^1\} - \min\{A(T), C(T)\}}_{\geq 0} + \underbrace{\min\{D(T), G(T)\} - \min\{D^1, G^1\}}_{\leq 0}\big).$$

The two addends in the numerator of the $VMPS_w$ have a different sign. General statements on which of the two will be larger in absolute terms cannot be made so that the sign of the $VMPS_w$ can be positive or negative. This reasoning can be transferred to $VMPS\Delta^{T^2}_{T^1}$, whereby we assume that the breakpoints induced by T^2 are a superset of those induced by T^1, such that

$$VMPS\Delta^{T^2}_{T^1} = \frac{B(T^1) - B(T^2)}{B(T^2)} \tag{6.42}$$
$$\iff \frac{1}{B(T^2)}\big(\underbrace{\min\{A(T^1), C(T^1)\} - \min\{A(T^2), C(T^2)\}}_{\geq 0}$$
$$+ \underbrace{\min\{D(T^2), G(T^2)\} - \min\{D(T^1), G(T^1)\}}_{\leq 0}\big).$$

Again, no general statement can be made on the sign of $VMPS\Delta^{T^2}_{T^1}$. When $T^1 \subset T^2$, it holds that $t(\tau^A)$ of T^1 is less or equal $t(\tau^A)$ of T^2 and thus $A(T^2) \leq A(T^1)$. The same holds for $C(T^2)$, $D(T^2)$, and $G(T^2)$. With increasingly granular discretization, both minima converge to zero as we approach τ^A and τ^C, and the profit of the multi-period problem converges to the continuous-time profit. However, no general statement on the convergence speed of the two minima is possible. Therefore, the profit itself needs not to converge monotonically. This implies that even in the one-facility-one-customer problem, the effect of moving to a model with more periods cannot be anticipated.

6.3.4. Multiple facilities and cost-minimizing problems



(a) $|T| = 2$ **(b)** $|T| = 4$

Figure 6.6.: Temporal aggregation in a multi-facility problem

To what degree are the above considerations transferable to, first, a multi-facility, multi-customer setting and, second, a cost-minimizing problem that requires full demand satisfaction? In a multi-facility setting, each facility may have a different profit function. Thus, there are two critical moments τ_i^A and τ_i^C for each facility. Nevertheless, as before, the benefit a decision maker obtains from operating each individual facility in a multi-period setting depends on the position of the breakpoints with respect to τ_i^A and τ_i^C. Figure 6.6 depicts the situation for a problem with two facilities (1 and 2), different profit functions in a 2-period and a 4-period problem, respectively. It illustrates that different opening policies are optimal depending on the relative position of the breakpoint τ to τ_i^A and τ_i^C. Facility 2 incurs higher fixed costs but has a higher unit profit and, consequently, a steeper slope. Operating it becomes profitable starting with the second half of the planning horizon, such that $\tau_2^A = 0.5(\tau^{max} - \tau^{min})$. For facility 1, operations become profitable after a quarter of the planning horizon, thus at $tau\underline{=}^A 0.25(\tau^{max} - \tau^{min})$. When the planning horizon is divided into two periods, the optimal opening policy is to open facility 2 in the second period. However, if one increases the number of periods to 4, opening facility 1 at the beginning of period 1 and leaving facility 2 closed throughout the planning horizon is profitable. Consequently, the discretization of the planning horizon can have a significant effect on the optimality of individual location decisions. Furthermore, the profit functions of individual facilities are not independent of one another. Instead, whenever a facility is opened, the profit functions of the remaining facilities change to the residual profit, the profit a facility generates if operated in addition to the already open ones. As fixed costs and capacity, however, are independent of the opening status of other facilities, this implies that τ_i^A and τ_i^C change and depend on the set of already operating facilities.

The profit-maximizing problem formulation that serves demand only when profitable is an extension of the more conventional formulation that requires total demand satisfaction at minimal costs. In the one-facility-one-customer case, the above considerations are obsolete in the cost-minimizing formulation. Due to the demand satisfaction constraint, the facility must operate as soon as the customer exhibits non-zero demand. However, all considerations can be transferred straightforwardly from the multi-facility setting. Assuming that sufficient capacity has been established to serve all demands, the profit functions of individual facilities can be considered residual profit functions. The resulting effects in the cost-minimizing problem are the same as in the profit-maximizing problem.

To summarize, there are two significant conclusions from the one-facility-one-customer case. Firstly, with an increasing number of periods, the optimal objective value of the multi-period problem converges to that of a continuous-time problem. However, that convergence is not monotone, not even when previous breakpoints are maintained. Due to implicit warehousing in capacitated problems, a more granular discretization has a positive and a negative effect on the objective value. Which one prevails cannot be said even in simple settings. Therefore, given a set of possible breakpoints, an exhaustive search is the only way to determine the optimal set of breakpoints. Secondly, for each facility, there is exactly one point in time at which it is profitable to open this facility, given a particular network configuration. However, this also implies that when, for example, five facilities

operate in the final period of a 32-period planning horizon, at least 27 of these breakpoints are obsolete.

6.4. Drivers of the VMPS

The values of the multi-period solution reported in literature vary tremendously. While some of this variance can be attributed to using different static counterparts (Section 6.2), characteristics of the underlying problem instance have a significant effect. We discuss and validate which characteristics contribute to a large VMPS. In Subsection 6.4.1, we start with a discussion on how isolated parameter changes will affect the VMPS in the one-facility-one-customer setting. We validate that expected effects translate to more complex instances by altering individual parameters of instances $\mathcal{P}_1^4$ and $\mathcal{P}_2^4$ from Example B. It is shown that the relative importance of fixed costs is a main indicator of the potential VMPS. In Subsection 6.4.2, we transfer these findings to instances from Section 1.3. We show that the relative importance of the fixed cost is the most significant indicator of the potential value of a multi-period model. Furthermore, we relate the potential VMPS to the size of the underlying service regions.

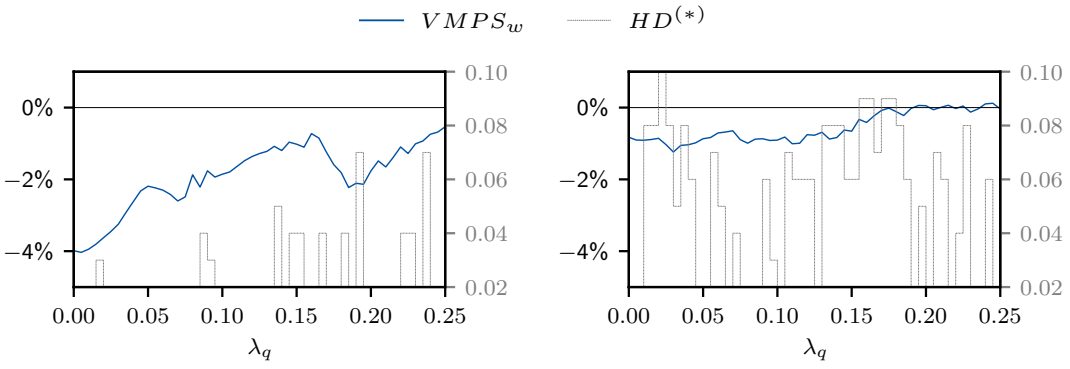
6.4.1. Isolated effects of altering individual parameters

In the following, we examine the isolated effects of parameter changes on the $VMPS_s$, the $VMPS_w$, and the $VMPS_{w/s}$. We discuss the effect on the one-facility-one-customer problem and subsequently examine to which degree these effects translate to more complex instances, in particular, instances derived similarly to $\mathcal{P}_1^4$ and $\mathcal{P}_2^4$ from Example B.

6.4.1.1. Capacity

All other conditions being equal, increasing the facility's capacity in the one-facility-one-customer problem implies that the point in time τ^C occurs later in the planning horizon. This implies that $D^{cont.}$, the profits lost due to the inability to hold implicit warehousing, decreases. As $D^{cont.}$ is an upper bound to the negative addend in the nominator of the $VMPS_w$ (refer to Eq.(6.41)), increasing the capacity suggests that the $VMPS_w$ increases. Meanwhile, the $VMPS_s$ should remain largely unaffected as implicit warehousing is neither possible in the static counterpart nor the multi-period model. For the $VMPS_{w/s}$, the associated static counterpart assumes that implicit warehousing is possible when determining the location decisions. It is likely to open a subset of facilities whose capacities are insufficient to serve the demands in later periods. The less restrictive the capacities are, the lower the potential mismatch between established and required capacities in later periods. With increasing capacities, the $VMPS_{w/s}$ is expected to decrease.

Example B 6.3 (Effect of increasing capacity on the VMPS) We gradually increase the capacity of all facilities of instances $\mathcal{P}_1^4$ and $\mathcal{P}_2^4$ by a constant factor $(1+\lambda_q)$ and evaluate the VMPS for the resulting sequence of instances. Let $\{\lambda_q^l\}$ denote the sequence of increasing factors.



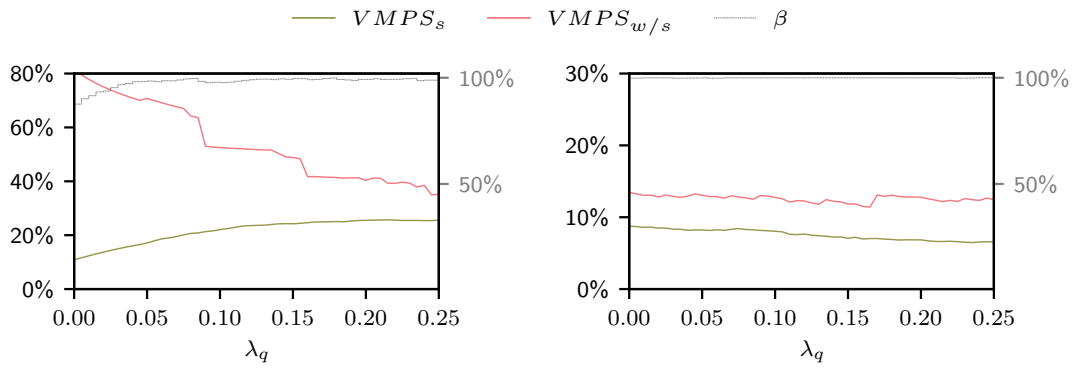
(a) $\mathcal{P}_1^4$ **(b)** $\mathcal{P}_2^4$

Figure 6.7.: Effect of increasing capacity on the $VMPS_w$ ($\mathcal{P}_1^4$-$\mathcal{P}_2^4$, Ex. B); (*)HD: Hamming distance between $Y_{|T|}^{\lambda_q^{l-1}}$ and $Y_{|T|}^{\lambda_q^{l}}$

The results are depicted in Figure 6.7 and Figure 6.8. Figure 6.7 shows the development of the $VMPS_w$ when the capacity of each facility is successively increased by up to 25%. For both sequences of problem instances, the $VMPS_w$ increases with increasing capacities. For $\mathcal{P}_1^4$ (Figure 6.7a), a 25% increase in the facilities' capacities leads to an increase of the $VMPS_w$ from -4.0% to -0.5%. For $\mathcal{P}_2^4$ (Figure 6.7b), which already has relatively large capacities to begin with, a 25% capacity increase results in the $VMPS_w$ rising from to -0.8% to 0.0%.

Furthermore, for $\mathcal{P}_1^4$, we can see that this increase is not linear but drops at several points before increasing again. For example, between $\lambda_q = 0.16$ and $\lambda_q = 0.19$, the $VMPS_w$ drops from -0.7% to -2.1%. Such drops correspond to significant changes in the subset of facilities operating in the final period of the optimal solution to the multi-period problem. When the subset of operating facilities changes, the effects of the one-facility-one-customer problem can no longer be transferred to the multi-facility case straightforwardly. To measure the changes in the set of operating facilities, we compute the Hamming distance between the vector of the optimal location decisions of the final period in the instance with the current capacity increase, $Y_{|T|}^{\lambda_q^{l}}$, and the corresponding vector optimal location decisions of the instance with marginally smaller capacities, $Y_{|T|}^{\lambda_q^{l-1}}$. For $\lambda_q = 0.19$, the Hamming distance between these two decision vectors is 0.07, which implies that the opening status of 7 of the 100 candidate facilities differs.

Figure 6.8 depicts the developments of the $VMPS_s$ and the $VMPS_{w/s}$ together with the service level β, the fraction of the demand that is served over the course of the planning horizon in the multi-period model. Again, effects are more significant for $\mathcal{P}_1^4$, the instance in which original capacities are more restrictive. As anticipated, the $VMPS_{w/s}$ is largest when capacities are most restrictive, and the "false" assumption of implicit warehousing has the most disastrous effect in the multi-period model. For $\mathcal{P}_1^4$ with original capacity levels, the $VMPS_{w/s}$ is at 81.3%. When capacities are increased by 25%, that value decreases to 35.2%. For $\mathcal{P}_2^4$, the same capacity increase only leads to a decrease from 13.5% to 12.5%.



(a) $\mathcal{P}_1^4$ **(b)** $\mathcal{P}_2^4$

Figure 6.8.: Effect of increasing capacity on the $VMPS_s$ and $VMPS_{w/s}$ ($\mathcal{P}_1^4$-$\mathcal{P}_2^4$, Ex. B)

Contrary to initial expectations, for $\mathcal{P}_1^4$ and $\lambda_q \in [0, 0.25]$, the $VMPS_s$ increases from 10.9% to a maximum value of 25.7%. This, however, can be explained when looking at the service level, as with the initial capacities, a larger proportion of the customer demands cannot even be served. We assume that effects in the multi-facility instance correspond to the accumulated effects expected at individual facilities. Hence, when the capacity increases to the point that more facilities become profitable throughout the planning horizon and serve demand that is otherwise not fulfilled, the increase in the pure number of operating facilities leads to a larger accumulated effect and, hence, a larger $VMPS_s$. Similarly, the slight decrease in the $VMPS_s$, for instance, $\mathcal{P}_2^4$ from 8.8% to 6.5% can be explained by the decrease in the number of facilities operating in the final period from 9 to 7. ▲

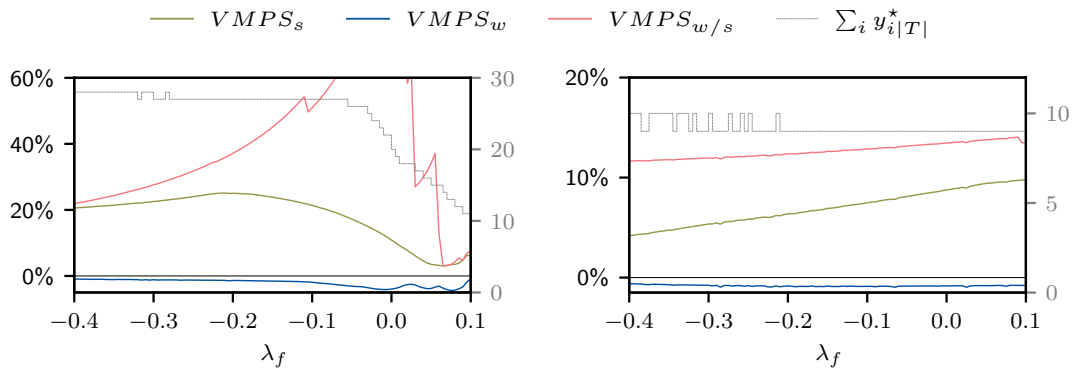
Example B 6.3 confirms that the effect of implicit warehousing decreases with increasing capacities. The $VMPS_w$ increases in the sense that the absolute value of the negative component decreases. The mismatch between the capacity assumed to be available in the static counterpart of the $VMPS_{w/s}$ and the multi-period model decreases, and so does the $VMPS_{w/s}$. While we expected the $VMPS_s$ to remain largely unaffected, we observe that, in particular, when the capacity increase leads to fewer facilities being opened, the $VMPS_s$ also decreases as the accumulated benefit of opening facilities later on in the planning horizon declines.

6.4.1.2. Fixed costs

All other conditions being equal, increasing the fixed costs in the one-facility-one-customer problem implies that the point in time τ^A after which operating the facility becomes profitable occurs later on in the planning horizon and consequently $A^{cont.}$ increases. As $A^{cont.}$ is an upper bound for the positive addend of the nominator for the $VMPS_s$ and the $VMPS_w$, we expect both values to increase with increasing fixed costs. Consequently, the $VMPS_{w/s}$ is also expected to increase.

Example B 6.4 (Effect of increasing fixed costs on the VMPS) Experiments on instances $\mathcal{P}_1^4$ and $\mathcal{P}_2^4$ were conducted in the same way as for the capacity increase in Example B 6.3. This time, only the fixed costs f_{it} are multiplied with the constant factor $(1 + \lambda_f)$. For

$\mathcal{P}_1^4$, small increases in the fixed costs already lead to several facilities being unprofitable. The optimal solution for $\lambda_f > 0.1$ is to open no facility. Therefore, we explored primarily negative values of λ_f. Results for the $VMPS_w$, the $VMPS_s$ and the $VMPS_{w/s}$ are depicted in Figure 6.9.



(a) $\mathcal{P}_1^4$ **(b)** $\mathcal{P}_2^4$

Figure 6.9.: Effect of increasing fixed costs on the $VMPS_w$, $VMPS_s$ and $VMPS_{w/s}$ ($\mathcal{P}_1^4$-$\mathcal{P}_2^4$, Ex. B)

For $\mathcal{P}_1^4$, the $VMPS_s$ increases in the interval of $\lambda_f \in [-0.4, -0.15]$ from 20.5% to 23.9%. For larger values of λ_f, the fixed costs are so large that they significantly alter the set of optimal facilities, eventually leading to only a fraction of the total demand being served. In that interval, the $VMPS_w$ decreases from −1.0% to −1.6%, while the $VMPS_{w/s}$ increases from 21.9% to 45.1%.

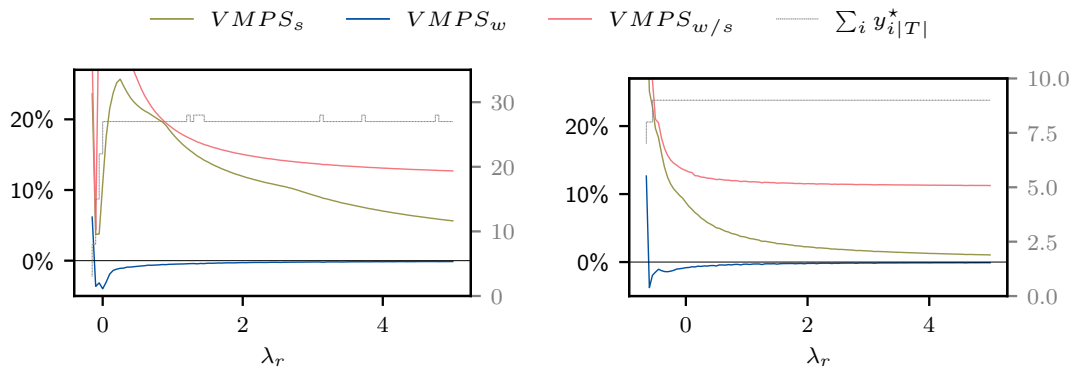
For $\mathcal{P}_2^4$, the $VMPS_s$ and the $VMPS_{w/s}$ increase. Capacities and profits are large enough such that for all values of λ_f all demand is served throughout the planning horizon in the multi-period model. When λ_f increases from −40% to +10%, the $VMPS_s$ increases from 4.2% to 9.8%, and the $VMPS_{w/s}$ increases from 11.7% to 13.5%. Meanwhile, the $VMPS_w$ decreases from −0.6% to −0.8%. ▲

Example B 6.4 confirms the positive effect increasing capacities have on the $VMPS_s$ and the $VMPS_{w/s}$. Meanwhile, the $VMPS_w$ remains largely unaffected and decreases slightly with increasing fixed costs. With increasing fixed costs, capacity becomes relatively more expensive, aggravating the effect of implicit warehousing. The static counterpart establishes successively fewer facilities, assuming capacities to be available later. Thus, the positive and negative components of the $VMPS_w$ increase with increasing fixed costs.

6.4.1.3. Variable profits

Increasing the variable profits implies that τ^A occurs earlier and $A^{cont.}$ decreases. With similar reasoning as before, increasing profits can be expected to have the opposite effect of increasing fixed costs.

Example B 6.5 (Effect of increasing variable profits on the VMPS) In the subsequent experiment, the unit profit of each customer r_j is multiplied by the factor $(1+\lambda_r)$. Figure 6.10 depicts the effects on $\mathcal{P}_1^4$ and $\mathcal{P}_2^4$.



(a) $\mathcal{P}_1^4$ **(b)** $\mathcal{P}_2^4$

Figure 6.10.: Effect of increasing unit profits on the $VMPS_w$, $VMPS_s$ and $VMPS_{w/s}$ ($\mathcal{P}_1^4$-$\mathcal{P}_2^4$, Ex. B)

The $VMPS_w$ converges to zero with increasing profits. Both the positive and negative components decrease with increasing profits and consequently a decreasing relevance of the fixed costs. Similarly, the $VMPS_s$ decreases as the added value of opening facilities later in the planning horizon and saving fixed costs becomes less relevant in the objective value. Interestingly, the $VMPS_{w/s}$ remains relatively constant at approx. 12% for $\mathcal{P}_1^4$, and approx. 11% for $\mathcal{P}_2^4$. We explain this by the fact that the corresponding static counterpart persistently establishes capacity in the "wrong" locations. For $\mathcal{P}_2^4$, the overlap coefficient (refer Def. 2.7) between the index set of facilities operating in the static counterpart and the index set of facilities operating in the final period of the multi-period model is 0.86. ▲

6.4.1.4. Demand development

We assume that customer demand is the only parameter that varies over time. A homogeneous increase of $d(t)$ over the course of the planning horizon will have the opposite effect as an increase in capacity - $D^{cont.}$ decreases - and a similar effect as an increase in unit profit - $A^{cont.}$ increases. But what happens if the total demand over the planning horizon remains the same but the time of its occurrence changes? Keeping the assumption of monotonously increasing demands, this means asking what happens when δ, the ratio between the demand in the final and the first period, increases. Increasing the slope of $d(t)$ implies that τ^A occurs later and τ^D occurs earlier in the planning horizon. Both, $A^{cont.}$ and $D^{cont.}$, increase. As both addends of the sum in the nominator of the $VMPS_w$ increase, the effect can be positive or negative. The $VMPS_s$ and $VMPS_{w/s}$ can be expected to increase.

Example B 6.6 (Effect of demand development on the VMPS) As described in Example B, we assume that the demand increases in constant steps such that the demand in the final period, period 4, equals δ times the demand in the initial period. In the following, we vary δ between 100%, 500%, and 1000%. This means we vary the level of parameter change over the course of the planning horizon. As the sum of all demands stays the same across all considered instances, this means that with increasing δ, the demand in periods 1 and

2 decreases, and the demand in periods 3 and 4 increases. The results are displayed in Table 6.3.

	δ	$VMPS_s$	$VMPS_w$	$VMPS_{w/s}$
$\mathcal{P}_1^4$	100%	10.9%	-4.0%	81.3%
	500%	52.8%	-15.7%	243.1%
	1000%	71.7%	-19.8%	301.9%
$\mathcal{P}_2^4$	100%	7.5%	-0.4%	11.4%
	500%	22.6%	0.2%	31.6%
	1000%	26.6%	0.1%	36.9%

Table 6.3.: Effect of increasing the level of demand increase δ on the $VMPS_w$, $VMPS_s$ and $VMPS_{w/s}$ ($\mathcal{P}_1^4$-$\mathcal{P}_2^4$, Ex. B)

For both instances, with increasing δ the $VMPS_s$ and the $VMPS_{w/s}$ increase. For $\mathcal{P}_1^4$, the $VMPS_w$ decreases, indicating that the negative effect induced by implicit warehousing outweighs the benefit of opening facilities later in the planning horizon. This is different for $\mathcal{P}_2^4$, for which the $VMPS_w$ increases slightly with increasing δ from -0.4% to 0.1% for δ equaling 100% and 1000%, respectively. Nevertheless, the magnitude of the effect of increasing δ on $\mathcal{P}_1^4$ and $\mathcal{P}_2^4$ differs significantly. The increases in the different VMPS are significantly larger for instance $\mathcal{P}_1^4$ with smaller capacities and smaller underlying service regions. We take a closer look at this in the following. ▲

Table 6.4 summarizes the above results. Example B 6.3 to Example B 6.6 indicate that while the degree of change in the time-varying parameters - here, the customer demand - significantly affects the potential benefit of a multi-period model, other characteristics of the underlying problem instance, i.e., the ratio between demand and capacity, as well as the ratio between attainable profits and fixed costs, are significant drivers for what might be gained from moving to a multi-period approach.

			$VMPS_s$	$VMPS_w$	$VMPS_{w/s}$
Capacity	q_{it}	↗	↘	↗	↘
Fixed Costs	f_{it}	↗	↗	↘	↗
Unit profit	r_j	↗	↘	↗	↘
Demand increase	δ	↗	↗	↗ / ↘	↗

Table 6.4.: Expected isolated effect of parameter changes on different VMPS

6.4.2. Experimental validation: indicators for a large VMPS

More interesting than anticipating the effect of isolated parameter changes on the value of the multi-period solution when given a reference point is the degree to which it is possible to anticipate the potential value of a multi-period approach before the problem is solved. This means either from the static counterpart or directly from the problem data. While it is intuitive to suspect that the degree to which parameters change throughout the planning horizon is a key indicator for the potential value of a multi-period approach, the previous experiment indicates that the ratios of time-invariant parameters such as demand versus capacities or fixed costs versus variable profits may outweigh the effect of significant

demand increases. Looking at Table 6.4, one can see that all isolated parameter changes that lead to an increase in the $VMPS_s$ imply an increase in the relevance of the fixed costs. In particular, if - all other conditions being equal - capacities decrease, this implies an increase in the unit fixed costs. So does decreasing unit profits or directly increasing fixed costs. The significant role of the fixed costs relative to the unit profit can also be seen in the one-facility-one-customer problem as in the simplified setting the $VMPS_s$ is upper bounded by $A^\infty/(B^\infty)$.

We quantify this relative importance of the fixed costs by two indicators. The first one is derived from the static, single-period model, the CFLP in which all parameters are aggregated. It sets the fixed costs incurred in the optimal solution in relation to the profits generated by that solution. We denote the resulting ratio by FC_w, such that

$$FC_w := \frac{\sum_i F_i y_i^\star}{\sum_i \sum_j (r_j - c_{ij}) D_j x_{ij}^\star}, \tag{6.43}$$

with $y_i^\star$ and $x_{ij}^\star$ the optimal location and allocation decisions in the CFLP. The second indicator is the profit ratio determined directly from the problem data before solving any mathematical program. It sets the average attainable unit profit in relation to the average unit fixed costs (refer to Eq. (4.2)).

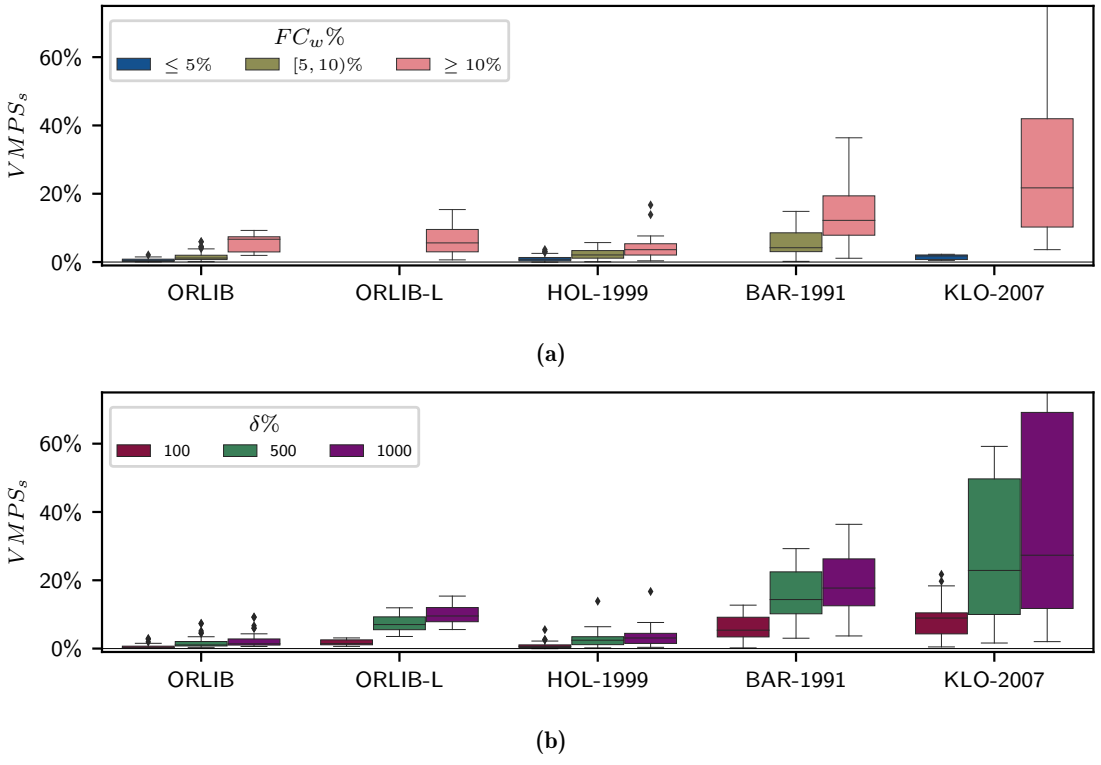


Figure 6.11.: $VMPS_s$ grouped by (a) fixed cost ratio in static problem, (b) relative demand increase

For each instance of the data sets described in Section 1.3, we derive three 4-period instances, postulating a constant demand increase of $\delta = 100\%$, 500%, and 1000%, respectively. Figure 6.11 displays the distribution of the resulting $VMPS_s$ for the different sets, grouped by FC_w (Figure 6.11a) and δ (Figure 6.11b), respectively. Particularly for the

instances from ORLIB, BAR-1991, and KLO-2007, the relative fixed costs FC_w yield more distinct distributions of the $VMPS_s$ than the relative demand increase δ.

Figure 6.12 depicts the $VMPS_s$ for all instances grouped by the profit ratio (Figure 6.12a) and δ (Figure 6.12b), respectively.Furthermore, instances are grouped according to the average dependence density (see Def. 2.13) underlying the aggregated static instance. An average dependence density below 0.1 indicates relatively small service regions in an optimal solution primarily composed of independent facilities. A larger average dependence density indicates larger service regions and optimal solutions primarily composed of interdependent subsets of facilities. Once again, independently of δ, instances with relatively high unit fixed costs (profit ratio≤ 2.5) exhibit a distinctly higher $VMPS_s$ than those with relatively lower fixed costs. Meanwhile, the distribution of the $VMPS_s$ for instances with $\delta = 500\%$ can hardly be distinguished from the distribution obtained when $\delta = 1000\%$. A two-sided t-test cannot reject the null hypothesis that the two means are equal with a statistical significance of 5% (p-value = 0.066).

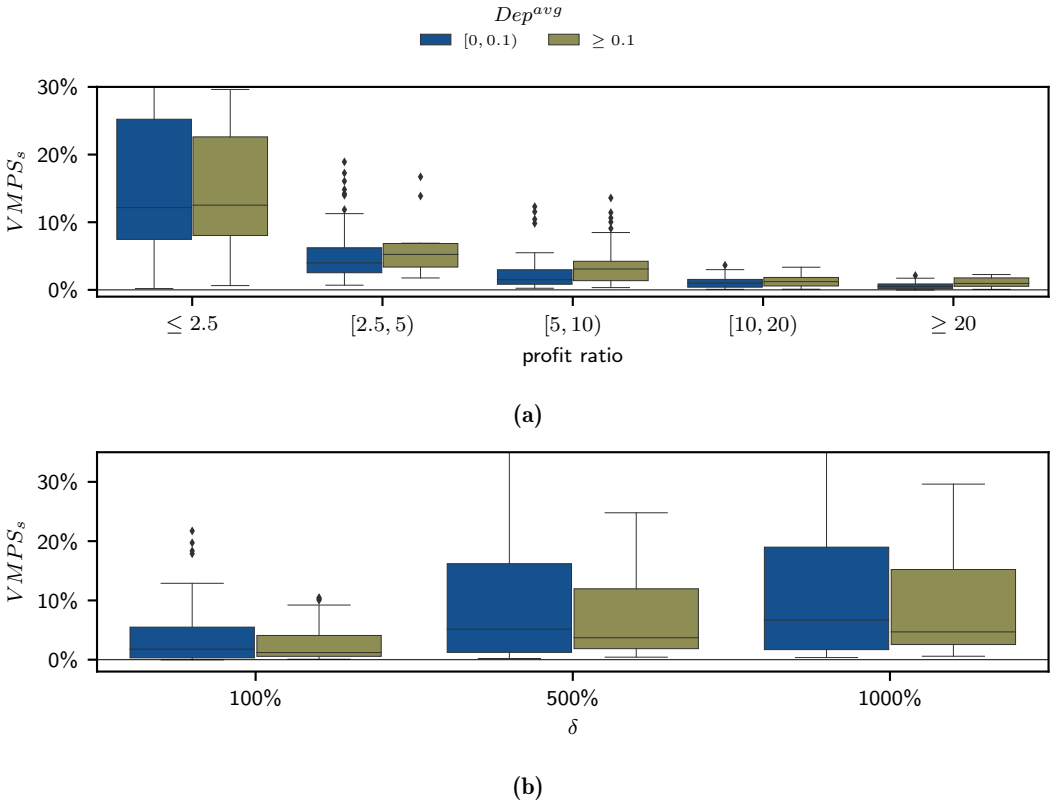


Figure 6.12.: $VMPS_s$ grouped by (a) profit ratio, (b) relative demand increase

Furthermore, we observe that, on average, the $VMPS_s$ between instances with high and low average dependence density are relatively similar. However, in particular, for increasing δ, the variation of the $VMPS_s$ in instances whose optimal solution is primarily composed of independent facilities is significantly larger than for those instances whose optimal solution is primarily composed of interdependent subsets of facilities. This can be explained by the observations from Example B. When facilities are independent, they serve certain customer regions by themselves. In the multi-period context, the customers in individual regions are served only after the facility in that region is opened. Meanwhile, until the demand in

that region suffices to offset the fixed costs in that region, the entire demand is lost. This means that in a static counterpart, either the facility serving a particular region operates throughout the planning horizon and accumulates high fixed costs in the beginning or remains closed, and a lot of the potential profit is lost. If facilities serve larger service regions as interdependent subsets, the multi-period model can open facilities successively, redistributing customers within each service region as demand increases and more facilities operate. At the same time, the static counterpart has more flexibility in finding a subset of facilities that balances the trade-off between lost profits and fixed costs.

In conclusion, the relative importance of the fixed costs is a good indicator of whether or not a model that allows us to open some facilities later on in the planning horizon will benefit the problem described by a particular instance. It is a more reliable indicator of the VMPS than the relative change of customer demand over the planning horizon.

6.5. Conclusion

Despite the VMPS being an established measure to quantify the benefit of a multi-period modeling approach, the reasons for strong variations across different instances and reports in literature are insufficiently explored. Even for a standard multi-period extension of the CFLP as the MP-CFLP with phase-in constraints considered in this chapter, driving factors for a high VMPS were unclear.

We first identified different definitions of the VMPS resulting from different static counterparts as one explanation for the differences in reported values. We distinguish between different aspects of multi-period modeling that can be quantified by different VMPS derived from different static counterparts. In this context, we take a closer look at the fact that in the capacitated case, moving to a multi-period model comes at two costs: increased complexity and added restrictions regarding when to use limited resources. This stands in contrast to the potential gain a decision-maker may derive from the added flexibility in decision-making. Which of these effects dominates depends on the problem at hand. Therefore, choosing the set of periods the planning horizon is divided into with care is essential.

We examine a restricted one-facility-once-customer instance to gain insights into the conditions under which moving to a multi-period model with phase-in constraints can be of value. We show that the potential benefit of a multi-period model depends on the relative position of the breakpoints, the moments in the planning horizon at which decisions can be taken, to the moments at which, firstly, the profits outweigh the fixed costs, and, secondly, the demands exceed the capacity. We use these insights to identify primary drivers in the parameters of a problem instance leading to a high VMPS. Experiments show that not necessarily the degree to which the problem parameters change throughout the planning horizon but rather the degree to which the multi-period model allows reducing the relevant cost components determines the value of a multi-period approach. Thus, only because decisions are taken in a temporal context, and some parameters change significantly over the planning horizon, time is not necessarily an important aspect to consider in the corresponding model.

The main results of the previous chapter can be summarized as follows:

- Moving to a multi-period model not only increases the flexibility in decision-making but also the restrictiveness regarding when to use restricted resources.
- The particular discretization of the planning horizon significantly affects the added value of the multi-period approach and must consequently be chosen with care.
- The main driver of the added value of the multi-period model is not the degree to which problem parameters change but the degree to which the model extension allows reducing relevant cost parameters.

Evaluating firstly which modeling additions enrich decision-making for the data at hand and then, secondly, deriving the level of granularity with which this aspect is included in the model from the problem's parameters are two insufficiently explored steps that may open important new pathways for location modeling. In particular, our insights into the sensitivity of the VMPS towards the position of breakpoints throughout the planning horizon opens interesting avenues for further research. For example, several works in the context of supply chain network design distinguish between tactical and strategic periods [Albareda-Sambola et al., 2012, Badri et al., 2013, Correia and Melo, 2016]. Location decisions can only be taken at strategic periods, a subset of the tactical periods at which allocation decisions can be altered. While the number of strategic periods increases the number of binary decision variables and thereby significantly affects the complexity of the problem, the choice of strategic periods is not further explored at this point. Rather, it is usually assumed to be, e.g., every second or fourth tactical period. Using our newborn insights to determine strategic periods that are effective in the sense that they increase the VMPS may potentially lead to better multi-period decision-making.

7. Service regions and demand uncertainty

The strategic, long-lasting nature of location decisions not only introduces a temporal dimension but also uncertainty into the problem as usually what will happen in several years is, at least to some degree, uncertain. Including uncertainty in mathematical programs is a topic receiving considerable attention in OR literature in general and location problems, in particular. Thereby, different ways to include uncertainty into the mathematical program exist. The two most prominent modeling paradigms are stochastic programming and robust optimization. It is often argued that the availability of information on the uncertain parameter or the risk attitude of the decision-maker specifies the best-suited paradigm. However, with the increasing availability of data and a wide range of options to incorporate different risk attitudes into the objective function, it is ultimately up to the decision-maker to choose the most appropriate paradigm. Meanwhile, guidelines on the impact of this choice on the resulting solutions and, in particular, the resulting decisions are missing.

In the present work, we evaluate the degree to which underlying decision patterns in the form of implicit service regions affect the added benefit of moving to a more comprehensive model. In Chapter 5, we evaluated the effect of including temporal developments into the model. This chapter looks at model formulations that include uncertainty in the demand parameters. This leads to the following research question:

RQ7: When is it worth explicitly considering uncertainty in the CFLP?

To quantify the added value of the multi-period model, we drew on the VMPS as an established measure in literature. Given the various modeling paradigms for including uncertainty, such a measure is not readily available in the uncertainty context. In Section 7.1, we provide a short overview of related work and conclude with an outline of the necessary steps to provide an answer to RQ7.

7.1. Related work on modeling uncertainty in facility location problems

We present a short overview of different modeling paradigms to include uncertainty in a mathematical program in Subsection 7.1.1. Thereby, the overview relies in large parts on Bakker et al. [2020]. In Subsection 7.1.2, we review how these concepts have been applied to facility location problems, particularly the CFLP and its extensions.

7.1.1. Modeling uncertainty

Methods for optimization under uncertainty have been studied intensely over the past decades. Often triggered by a particular application, different paradigms to integrate uncertainty into the decision problem have evolved. Although these models all intend to solve a similar underlying problem, they differ strongly with respect to the uncertainty representation, the prescriptive solution information they provide, and the means of performance

evaluation. A comprehensive review of different uncertainty paradigms with respect to these aspects can be found in Bakker et al. [2020], in which we particularly emphasize the relationship between time and uncertainty. Yet, the main challenges arising with respect to choosing and comparing the solutions between different paradigms hold for static and multi-stage problems alike. The latter assume a successive disclosure of uncertain parameters. Uncertainty in the coefficients and the constraints requires redefining the concepts of optimality and feasibility, making a direct comparison between the solutions obtained with models using different uncertainty paradigms difficult. In the following, we briefly introduce the main uncertainty concepts and the challenges that arise, as pointed out in the review.

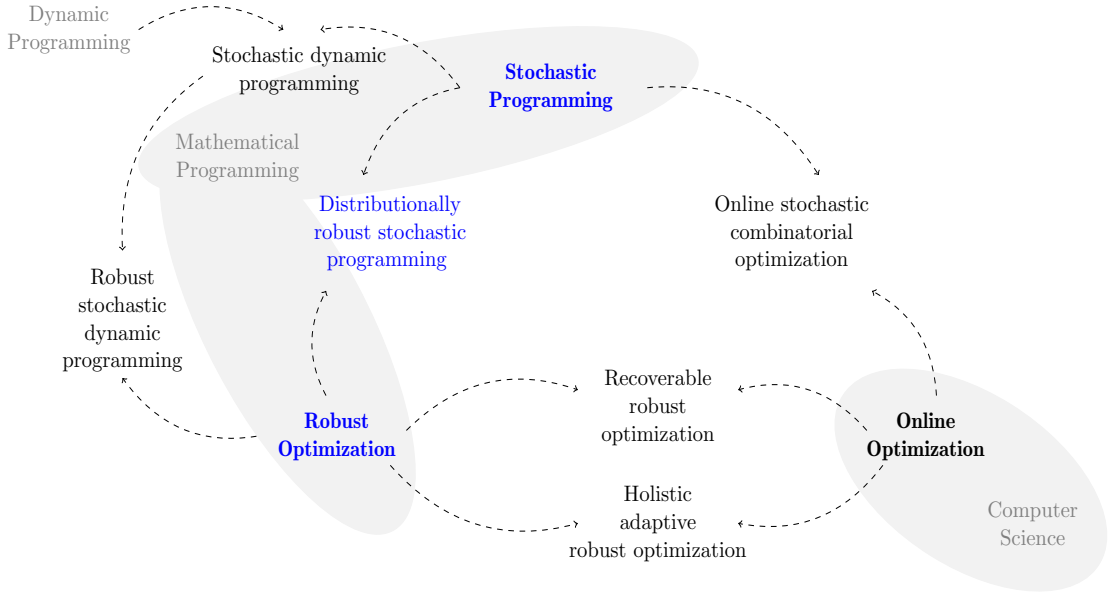


Figure 7.1.: Prominent methods for solving multi-stage stochastic programs have been derived from three basic concepts: stochastic programming, robust optimization, and online optimization, stemming from the fields of mathematical programming and computer science, respectively [Bakker et al., 2020]

Figure 7.1 illustrates the relationship between the most prominent methods to handle uncertainty in decision problems, with those particularly relevant to location modeling highlighted in blue. Particularly when information is disclosed successively, and decisions have to be taken multiple times in response to new inputs, methods from computer science, i.e., online optimization, pose an alternative to modeling paradigms like stochastic programming and robust optimization that stem from mathematical programming. Depending on the application, different methods are dominant, and several approaches combining elements of different methods have been proposed. Given the long-lasting nature of high-investment location decisions, stochastic programming and robust optimization are the most common methods to deal with uncertainty in facility location problems and their extensions. In recent years, a mixture of both, distributionally robust stochastic optimization, gained further attention. We briefly introduce these methods in the following. Thereby, we focus on how uncertainty in the parameters included in the mathematical programming formulation, the definition of optimality and feasibility given this uncertainty,

and how the value of explicitly including uncertainty in the model is quantified. Furthermore, we discuss existing measures quantifying the added value of the uncertainty-aware approach. We review works applying these methods to the CFLP in Subsection 7.1.2.

7.1.1.1. Stochastic programming

Stochastic programming assumes that uncertain parameters are random variables following a known probability distribution. According to the classification of decision environments by Rosenhead et al. [1972], stochastic programming depicts the situation of decision-making under risk. Originating from the seminal paper "Linear programming under uncertainty" by Dantzig [1955], stochastic programming is, to our knowledge, the first approach from within the OR community to deal with uncertainty in mathematical programming-based optimization. To define optimality in the presence of uncertain coefficients in the objective function, the authors propose to optimize the expected value, which implies the decision-maker has a risk-neutral attitude. To model different risk attitudes but still comprise information on the uncertain parameters in a single, deterministic value, other objective criteria have been proposed. The most prominent is the Conditional Value at Risk [Rockafellar and Wets, 1976]. Regarding feasibility, stochastic programming considered temporal relations between decisions and uncertainty observations early on by introducing the concept of recourse — a partial decision to be fixed after uncertainty has been disclosed so that feasibility is ensured, even if possibly at a high cost. This setting is known as two-stage stochastic programming and is formalized by a nested problem formulation in which the first stage decision has to be taken ex-ante, that is before the realization of the uncertain parameter, and the second stage decision is taken ex-post, after uncertainty has been disclosed. When the set of scenarios is discrete, a tractable reformulation of a stochastic program is referred to as a deterministic equivalent.

7.1.1.2. Robust optimization

A practical drawback of stochastic programming is that probability distributions of uncertain parameters are often hard – if not impossible – to identify and hence bound to estimation errors [King and Wallace, 2012]. Robust optimization addresses this problem by assuming that solely information on the set of possible outcomes and not their individual likelihood is available. First proposed by Soyster [1973], it covers the situation of decision-making under ambiguity. Uncertainty is formally described by an uncertainty set that contains all possible realizations of the uncertain problem parameters. Robust feasible solutions must be feasible for all potential realizations of the uncertain parameters. Optimality is usually defined by a min-max cost or min-max regret objective. The uncertainty set allows for a strict separation between the model of uncertainty and the actual optimization problem. Thereby, the uncertain phenomenon is often described by primitive uncertainties residing in box-shaped, ellipsoidal, or polyhedral uncertainty sets. These primitive uncertainties affinely perturb the actual problem parameters from their nominal values. Tractable reformulations for different forms of this uncertainty set have been proposed and are referred to as robust counterparts [Bertsimas and Sim, 2004, Bertsimas and Brown, 2009].

When some decisions are to be taken only after uncertainty has been disclosed, one speaks of adaptive or adjustable robust optimization [Ben-Tal et al., 2004]. As uncertainty sets are often infinite, decision rules that yield second-stage recourse decisions for any realization of the uncertain parameters are optimized as part of the robust counterpart. As general decision rules are not tractable, considerations are usually restricted to a specified functional form, often linear functions, whose parameters are decision variables in the robust counterpart.

7.1.1.3. Distributionally robust optimization

Distributionally robust stochastic optimization combines ideas from robust optimization and stochastic programming. It acknowledges that the probability distribution governing the uncertain data is difficult to obtain and often subject to uncertainty itself [Wiesemann et al., 2014]. Scarf [2005] propose a first formulation in which the optimization problem is reformulated with respect to the worst-case costs over a set of probability distributions assumed to contain the true probability distribution. Different forms of the ambiguity set, the set of potential distributions, have been proposed [Ben-Tal and Nemirovski, 1998, Delage and Ye, 2010], before Wiesemann et al. [2014] present a unifying framework for modeling these sets.

7.1.1.4. Performance evaluation

Besides different concepts of optimality and feasibility and different uncertainty models, stochastic programming and robust optimization primarily differ regarding their performance measures. In stochastic programming, the benefit arising from explicitly considering uncertainty is measured by the Value of the Stochastic Solution (VSS) and the Expected Value of Perfect Information (EVPI). The VSS captures the difference between the objective value of the stochastic program and the expected objective value of the Expected Value Problem (EVP), the expected outcome obtained when the first stage decisions are taken in a purely deterministic setting in which all parameters have been set to their expected values. The VSS hence expresses what is gained by using a stochastic modeling approach. The EVPI, on the other hand, measures what is lost because only probabilistic – and not perfect – information is available. It is the difference between the expected objective value when the realization of the uncertain parameter is known at the first stage and the objective value of the stochastic program [Birge and Louveaux, 2011].

In contrast to stochastic programming, in which the added benefit of the more comprehensive model is quantified, in robust optimization, the price of robustness measures the trade-off between feasibility violation probability and impact on the objective value [Bertsimas and Sim, 2004]. In robust optimization, the objective is a worst-case objective, and the requirement of robust feasibility usually comes at relatively high costs. Thus, this is a loss-oriented view on including uncertainty. It asks how much needs to be paid to protect the solution against all, particularly the worst possible realizations of the uncertain parameters.

Summing up, stochastic programming and robust optimization use different formal descriptions of the uncertain parameters, deploy different concepts of optimality and feasibility under uncertainty, and evaluate the resulting solutions differently. This makes a direct comparison extremely difficult. Nevertheless, both methods have been successfully applied to the CFLP and its extensions, addressing a mutual goal: improving location decisions in the face of parameter uncertainty.

7.1.2. Robust and stochastic variants of the CFLP

Several review papers discuss advances in facility location under uncertainty [Snyder, 2006, Melo et al., 2009]. A recent overview is given in Correia and Saldanha da Gama [2019]. The authors distinguish between three major research directions stemming from different sources of uncertainty: models accounting for "congestions" that address situations in which customer demand for service has a probabilistic behavior, models that handle unexpected disruptions in the network, e.g., the facilities or the transport links, and models considering aspects arising from uncertainty in the problem parameters such as demands or transport costs. In the following, we restrict our review to models of the latter kind. In particular, we restrict ourselves to works considering uncertainty in the customer demand of the CFLP and its extensions.

7.1.2.1. Stochastic facility location models

Many works apply stochastic programming to facility location problems, particularly the CFLP and its extensions in supply chain network design. The strategic location and tactical allocation decisions naturally yield a two-stage decision structure. The location decisions are first-stage, here-and-now decisions. The allocation decisions are second-stage, wait-and-see recourse decisions.

Laporte et al. [1994] present a model in which allocation decisions are also considered in the first stage. The expected net revenue resulting from recourse transportation decisions is considered in the second stage. Albareda-Sambola et al. [2011b] consider location decisions in the first stage and, in the second stage, either allocate the demand to an existing facility or outsource it to external suppliers. Lin [2009] present a stochastic version of the single-source CFLP with service level requirements and uncertain demands. Nickel et al. [2012] present a multi-period supply chain network design problem that includes financial decisions with uncertainty in the demands and the interest rates. The authors not only quantify the relevance of their approach by means of the relative value of the multi-stage stochastic solution but also evaluate the effect the inclusion of uncertainty has on the resulting location decisions. While all of the above works evaluate the added value of the stochastic programming approach by means of the VSS, a comparison of the decisions taken by the stochastic program to those taken by the expected value problem is rarely performed. Nickel et al. [2012] compare the number of facilities operating and the proportion of the customer demand served for a particular problem instance and report significant differences in the decisions taken by the EVP and the stochastic program.

7.1.2.2. Robust facility location models

Several authors address uncertainty in the parameters of facility location problems using robust optimization. For the CFLP, two major contributions have been made by Baron et al. [2010], who present the first static robust counterpart formulation to a location transportation problem, and Ardestani-Jaafari and Delage [2018], who extend this model to an adjustable robust counterpart.

Baron et al. [2010] are the first to present a robust counterpart to a CFLP with capacity choice. The model is static in the sense that all decisions are ex-ante here-and-now decisions. Location decisions have to be fixed at the beginning of the planning horizon. However, customer allocations and capacity choices are made over a $|T|$-period planning horizon. The demand parameters are uncertain and lie in either a box or an ellipsoidal uncertainty set. The authors analyze the robust solutions in the decision space via the number of operating facilities, their respective capacities, and the overall connectivity of the resulting graph. They compare them with solutions from a deterministic model in which all values are set to averages, the nominal values of the uncertainty sets. In their formulation, the robust counterpart must simultaneously respond to higher potential demand – requiring larger capacities – and lower potential revenue resulting from lower potential demand. Consequently, they observe that with increasing uncertainty, the model opens fewer facilities, resulting in lower fixed costs with higher per-facility capacities. Furthermore, they observe that with ellipsoidal uncertainty, the number of customers allocated to more than one facility throughout the planning horizon increases.

Ardestani-Jaafari and Delage [2018] extend the model of Baron et al. [2010] to an adjustable robust model in which only the location decisions are ex-ante here-and-now decisions, and the allocation and capacity decisions are ex-post wait-and-see decisions. These decision rules establish a functional relationship between wait-and-see decisions and uncertain parameters, in this case, the uncertain demand. The authors approximate this relationship by affine functions and evaluate the effect of gradually increasing their complexity. In particular, they successively increase the number of uncertain parameters each decision rule considers, i.e., whether the allocation decisions for a particular customer in a given period consider only the demand of that specific customer in that period, all demands of all customers in that period, or all demand of all customers up to that period. Furthermore, the authors lift the demand uncertainty space to two dimensions, which allows them to react differently to positive and negative perturbations of the nominal demand. Other works presenting robust counterparts for extensions of the CFLP include Gülpınar et al. [2013], who consider a formulation with a service level. Gabrel et al. [2014] and Zeng and Zhao [2013] present cutting-place methods to solve similar problems with budgeted uncertainty.

Several authors present robust counterparts for hub location problems [Boukani et al., 2014, Shahabi and Unnikrishnan, 2014, Zetina et al., 2017]. These works must be mentioned because they commonly evaluate which hub locations are opened with different models. In particular, Zetina et al. [2017] compare the number and indices of the hubs opened with their proposed robust counterpart with budgeted uncertainty to those opened by a two-stage stochastic program. Thereby, demand is assumed to be distributed uniformly over

the support of the uncertainty set. The authors compare the worst-case and average-case performance of both the stochastic and robust solutions and observe that while the robust solution performs only slightly worse on average, it performs significantly better in the worst-case scenario.

7.1.2.3. Distributionally robust stochastic location models

Recently, authors applied distributional robust stochastic optimization to variants of the CFLP. Zhu et al. [2018] present a distributionally robust model for the traffic sensor location problem with uncertainty in the travel time data. Wang et al. [2020] present an adaptive distributionally robust hub location problem with multiple commodities under demand and cost uncertainty. Liu et al. [2022] use distributional robust optimization to model state-dependent customer demands. Depending on, e.g., seasonal or socio-economic conditions, the demand resides in different ambiguity sets, collections of probability distributions for the uncertain demand parameters. The authors conduct a sensitivity analysis to determine the impact of changes in the parameter values, which describe the ambiguity set, on the optimal and worst-case objective values. This analysis is feasible since, when the ex-ante location decisions are fixed, the second stage problem becomes a linear program.

The review of literature demonstrates that while ample research on uncertain model variants of the CFLP has been performed, RQ7, under which conditions it is worth deploying an uncertainty model, remains largely unaddressed. The general notion is that when parameters are uncertain, an uncertainty-aware model is of value, and the appropriate modeling paradigm is a direct consequence of the available data or the risk attitude of the decision-maker. Particularly for the case of the CFLP no comparison of the solutions resulting from stochastic and robust models exists at this point.

In Section 7.2, we present a unified model of the uncertain demand in the CFLP, which can be used in stochastic programming and robust optimization. We subsequently derive a two-stage stochastic program and an adjustable robust counterpart that both address the same problem: finding the optimal location allocation policy given the presented uncertain demand. In Section 7.3, we present concise measures that allow evaluating and comparing the performance of the solutions obtained with different approaches in the objective value space. To answer RQ7, we then turn to the decision space and evaluate to which degree different modeling approaches recommend different decisions and to which degree they recommend the same decisions but evaluate them differently. In this context, we relate our findings to the size of the service regions underlying the problem instance. Up to this point, all experiments are performed on instances from Example C. In Section 7.4, we validate all findings on instances from data sets described in Section 1.3. We summarize our main conclusions in Section 7.5.

7.2. Capacitated facility location with uncertain demand

In the following, we present a two-stage stochastic program and several variants of adjustable robust counterparts to the CFLP with uncertain demand. Variants of both mod-

els can be found in literature. In particular, the presented two-stage stochastic program can be found in Correia and Saldanha da Gama [2019]. The adjustable robust counterpart formulations can be considered a simplification of the location-transportation model presented by Ardestani-Jaafari and Delage [2018]. Nevertheless, to the best of our knowledge, this is the first time these two modeling paradigms are compared on a CFLP with a unified model of the uncertain demand.

7.2.1. A unified model of uncertain demand

Depending on whether a stochastic programming or robust optimization approach is pursued, the uncertain demand is either modeled by a set of discrete scenarios, a probability distribution, or an uncertainty set. In the following, we assume that information on the set of all possible scenarios and their individual likelihood is available. We derive scenario and uncertainty sets that allow comparing the solutions obtained with stochastic programming and robust optimization approaches.

We assume that only the customer demand is uncertain and denote the uncertain demand of customer j by $\tilde{D}_j$. We further assume that the uncertain demand of each customer is linearly dependent on a single primitive uncertainty $\xi \in \mathbb{R}$, such that

$$\tilde{D}_j(\xi) = D_j^0 + \xi \hat{D}_j. \tag{7.1}$$

Thereby, D_j^0 is the nominal demand perturbed by $\xi \hat{D}_j$. The drift parameter $\hat{D}_j$ can be scaled so that the primitive perturbation parameter ξ can be normalized to the interval $[-1, 1]$. The above is a standard way to describe uncertainty in robust optimization, and the interval $[-1, 1]$ is called the uncertainty set $\mathcal{U}$.

To integrate this uncertainty model in a stochastic program, we assume that each primitive uncertainty ξ is a random variable and that the interval $[-1, 1]$ is the support of its probability distribution. Furthermore, we assume that ξ has an expected value of 0. The latter has the effect that the nominal demand D_j^0 equals the expected demand. We assume that ξ follows a uniform distribution, such that $\xi \sim U[-1, 1]$. However, any probability distribution with the above mean and support could equally be applied. The uniform distribution is a continuous probability distribution, which, as such, cannot be included in a stochastic programming formulation. Instead, the probability distribution is approximated by a finite set of scenarios $\Omega := \{\xi^1, \xi^2, \ldots, \xi^{|\Omega|}\}$ that are randomly drawn from $U[-1, 1]$. This is a state-of-the-art procedure, also called sample average approximation. We want to ensure that the robust and the stochastic models include the same information on extreme realizations of the random parameters, which is why for any set Ω, we set $\xi^1 := 1$ and $\xi^2 := -1$. Furthermore, to ensure that Ω correctly captures the expected or nominal value of each parameter, we generate further scenarios in Ω by randomly drawing ξ from $U[0, 1]$ and subsequently adding ξ and $-\xi$ to the set of scenarios.

The above uncertainty model is very versatile and allows capturing a variety of uncertainty settings. We examine three different settings with three different perturbation magnitudes. For the first, we distinguish between global (g), local (l), and regional (r) uncertain effects

by altering the number of primitive uncertainties that the uncertain customer demands depend on. In particular, we distinguish between the following:

Global uncertainty: All demand parameters depend on the same, single primitive uncertainty ξ, as described in Eq. (7.1) and $\mathcal{U}_g := [-1, 1]$. In this case, uncertainty reflects, e.g., the economic situation. Notice that with different ratios between D_j^0 and $\hat{D}_j$, this might still allow one to model an effect of different magnitude for different customers. However, assuming that all demand is affected by the same primitive uncertainty specifies the direction and, hence, allows us to model situations with high, low, and medium demand scenarios as common in stochastic programming.

Local uncertainty: Each parameter is affected by its own primitive uncertainty ξ_j, such that $\tilde{D}_j(\xi_j) = D_j^0 + \xi_j \hat{D}_j$. Thereby, the individual uncertain parameters ξ_j are mutually independent. The uncertainty set takes the form of a hypercube $\mathcal{U}_l := [-1, 1]^{|J|}$. Local uncertainty reflects a situation in which the demand of individual customers is forecasted based on historical data and, consequently, subject to estimation errors.

Regional uncertainty: With regional uncertainty, we explore a mixed form of the above commonly used uncertainty settings. In particular, we assume that there are a few regional impacts that affect the uncertain demand of customers in a certain area of the facility-customer space. The effect in a given region $r \in R$ is modeled by a primitive uncertainty ξ_r. Customers are assigned to a region based on their estimated coordinates obtained with MDS (see Section 1.3). The uncertain demand of customer j belonging to region r is then described by $\tilde{D}_j(\xi_r) = D_j^0 + \xi_r \hat{D}_j$. The primitive uncertainties affected different regions are again independent, and the uncertainty set is a hypercube of reduced dimensionality $\mathcal{U}_r := [-1, 1]^{|R|}$, with $2 \leq |R| < |J|$.

In the following, we refer to the vector of primitive uncertainties as ξ, regardless of its dimensions. We refer to a particular realization of ξ by ξ^ω.

Besides the uncertainty setting, we distinguish between different uncertainty magnitudes, described by the percentage to which the demands maximally deviate from their nominal or expected value. Therefore, $p \in [0, 1]$ denotes the perturbation level, such that $\hat{D}_j = pD_j^0$ for all customers.

7.2.2. Two-stage stochastic capacitated facility location

We consider the classic stochastic programming formulation of the CFLP, which can also be found in Correia and Saldanha da Gama [2019], further referred to as Two-stage Stochastic CFLP (2SS-CFLP). Location decisions are first-stage decisions, and allocation decisions are second-stage recourse decisions. In a profit-maximizing formulation, the second-stage problem $Q(y, \xi)$ has relatively complete recourse. It is always a feasible solution not to serve any customer demand.

2SS-CFLP

$$\max -\sum\nolimits_i F_i y_i + \mathbb{E}_\xi \left[Q\left(y, \xi\right) \right] \tag{7.2}$$

$$\text{s.t.} y_i \in \{0,1\} \qquad i \in I \tag{7.3}$$

with

$$Q\left(y, \xi^\omega\right) := \max \sum\nolimits_i \sum\nolimits_j (r_j - c_{ij}) \tilde{D}_j(\xi^\omega) x_{ij} \tag{7.4}$$

$$\text{s.t.} \sum\nolimits_i x_{ij} \leq 1 \qquad j \in J \tag{7.5}$$

$$\sum\nolimits_j \tilde{D}_j(\xi^\omega) x_{ij} \leq Q_i y_i \qquad i \in I \tag{7.6}$$

$$x_{ij} \geq 0 \qquad i \in I,\ j \in J. \tag{7.7}$$

When ξ resides in a finite set of scenarios Ω, we can formulate a deterministic equivalent to the above problem, which we further denote by Deterministic Equivalent Two-stage Stochastic CFLP (2SS-CFLP-DE). The second-stage recourse decisions $x_{ij\omega}$ denote the fraction of the demand of customer j allocated to facility i in scenario ω. The probability of scenario ω is denoted by π^ω.

2SS-CFLP-DE

$$\max -\sum\nolimits_i F_i y_i + \sum\nolimits_{\omega \in \Omega} \pi_\omega \left[\sum\nolimits_i \sum\nolimits_j \left(r_j - c_{ij}\right) \tilde{D}_j(\xi^\omega) x_{ij\omega} \right] \tag{7.8}$$

$$\text{s.t.} \sum\nolimits_i x_{ij\omega} \leq 1 \qquad j \in J,\ \xi^\omega \in \Omega \tag{7.9}$$

$$\sum\nolimits_j \tilde{D}_j(\xi^\omega) x_{ij\omega} \leq Q_i y_i \qquad i \in I,\ \xi^\omega \in \Omega \tag{7.10}$$

$$y_i \in \{0,1\} \qquad i \in I \tag{7.11}$$

$$x_{ij\omega} \geq 0 \qquad i \in I,\ j \in J,\ \xi^\omega \in \Omega. \tag{7.12}$$

The above formulation can be handed over to off-the-shelf MIP solvers. The optimal objective value to the 2SS-CFLP-DE will subsequently be referred to as $z^\star_{SP}$.

The state-of-the-art measure to quantify the value of a stochastic program compared to a deterministic approach is to determine the VSS. It describes the relative increase in the objective value when comparing $z^\star_{SP}$ to the expected objective value a decision-maker obtains when making first-stage decisions based on the expected value problem. To obtain the latter in the present setting, we solve the CFLP in which all uncertain demand parameters are represented by their expected values which, in this case, implies $\mathbb{E}(\tilde{D}_j(\xi)) = D^0_j + \mathbb{E}(\xi)\hat{D}_j = D^0_j$. The resulting optimal location decisions, $y^\star_{EV}$, are then fixed as the first-stage decisions in the 2SS-CFLP-DE such that

$$y_i = y^\star_{EV_i}. \tag{7.13}$$

The 2SS-CFLP-DE is then solved with these fixed location decisions to obtain the expected objective value, $z^\star_{EV}$. It represents the expected profit if uncertainty is not considered in

the first stage.

Definition 7.1. *The **Value of the Stochastic Solution (VSS)** is the relative arithmetic difference between the optimal objective value of the stochastic program, $z^\star_{SP}$, and the expected objective value when first-stage decisions are based on the expected value problem, $z^\star_{EV}$, such that*

$$VSS = \frac{z^\star_{SP} - z^\star_{EV}}{z^\star_{EV}}. \tag{7.14}$$

As the set of feasible solutions to the deterministic equivalent is a superset of the set of feasible solutions to the deterministic equivalent in which location decisions are fixed to those of the expected value problem, it holds that $VSS \geq 0$. Notice that the concept of the VSS is very similar to the hybrid VMPS, which evaluates the added benefit of a multi-period model by comparing its objective value to that of a model using less, in that case, temporally aggregated information. Based on the same rationale as for the hybrid VMPS, we conclude that the VSS is only greater than 0 if the expected value problem opens a different, non-optimal subset of facilities than the stochastic program.

Example C 7.1 (Stochastic programming vs. expected value solution) Consider instances $\tilde{\mathcal{P}}_1$ and $\tilde{\mathcal{P}}_2$ from Example C and assume the demand is modeled according to the description in Subsection 7.1.1 with three different perturbation levels, $p \in \{5\%, 10\%, 20\%\}$, and global ($g$), local ($l$), and regional ($r$) uncertainty settings. For the latter, we consider two randomly determined regions. For each of the 6 resulting uncertain problems, we determine the optimal solution to the 2SS-CFLP-DE, $s^\star_{SP}$, and the optimal solution to the expected value problem, $s^\star_{EV}$. We derive $|\Omega| = 50$ scenarios to be considered in the deterministic equivalent.

		VSS (%)			$Jacc(I^\star_{SP}, I^\star_{EV})$			$overlap(I^\star_{SP}, I^\star_{EV})$		
	p	g	l	r	g	l	r	g	l	r
$\tilde{\mathcal{P}}_1$	5%	0.0	0.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0
	10%	1.9	0.7	2.5	0.9	0.9	0.9	1.0	1.0	1.0
	20%	22.8	19.4	26.3	0.8	0.9	0.8	1.0	1.0	1.0
$\tilde{\mathcal{P}}_2$	5%	0.9	0.9	0.8	0.6	0.6	0.6	0.7	0.7	0.7
	10%	1.6	1.8	1.4	0.8	0.8	0.8	0.9	0.9	0.9
	20%	3.1	3.7	2.8	0.4	0.4	0.4	0.6	0.6	0.6

Table 7.1.: Difference between stochastic programming and expected value solution in objective value and decision space ($\tilde{\mathcal{P}}_1$-$\tilde{\mathcal{P}}_2$, Ex. C)

Table 7.1 depicts the VSS as a measure of comparison in the objective value space, as well as the Jaccard index (see Eq. (2.4)) and the overlap coefficient (see Def. 2.7) comparing the similarity of the resulting solutions in the decision space – particularly, comparing the set of facilities operating in $s^\star_{SP}$, $I^\star_{SP}$, and the set of facilities operating in $s^\star_{EV}$, $I^\star_{EV}$. As to be expected, the VSS increases with an increasing perturbation level p. The magnitude of this increase, however, significantly differs between $\tilde{\mathcal{P}}_1$ and $\tilde{\mathcal{P}}_2$. While for $\tilde{\mathcal{P}}_2$, the VSS never exceeds 3.7%, for $\tilde{\mathcal{P}}_1$ it increases up to 26.3%. Also, the uncertainty setting affects the

VSS of instances $\tilde{\mathcal{P}}_1$ and $\tilde{\mathcal{P}}_2$ differently. For $\tilde{\mathcal{P}}_1$, throughout all perturbation levels, regional uncertainty (r) yields the greatest VSS. Meanwhile, for $\tilde{\mathcal{P}}_2$, local uncertainty yields the largest VSS throughout.

Regarding the decision space, Table 7.1 compares the subset of facilities operating in the optimal solution to the stochastic program with that of the expected value problem by measures that consider only the location decisions. We will look closer at the differences in the allocation decisions, particularly the implied service regions, in Section 7.3. However, the location decisions indicate a significant difference between the effect that the inclusion of uncertainty has on the optimal facilities in $\tilde{\mathcal{P}}_1$ compared to $\tilde{\mathcal{P}}_2$. Despite a significantly higher VSS, the overlap coefficient for $\tilde{\mathcal{P}}_1$ is 1.0 throughout all 6 considered problems. This means that one of the two sets of facilities is a subset of the other. In particular, the facilities operating in the stochastic program are a subset of those operating in the expected value problem. With an increasing perturbation level, the number of facilities operating in the optimal solution to the stochastic program decreases. While for $p = 5\%$, the stochastic program and the expected value problem operate the same set of facilities, the Jaccard index decreases with increasing p, indicating that fewer facilities operate in the optimal solution to the stochastic program. Even though for $\tilde{\mathcal{P}}_2$, the VSS is lower than for $\tilde{\mathcal{P}}_1$ throughout, the effect uncertainty has on the subset of optimal facilities is more profound as both, the overlap coefficient and the Jaccard index, are significantly lower than 1.0 throughout all considered uncertainty settings. Both decrease with increasing perturbation levels, indicating large differences between the sets of optimal facilities. ▲

7.2.3. Adjustable robust capacitated facility location

We propose adjustable robust counterparts to the CFLP corresponding to the previously presented stochastic program. The different versions deploy different decision rules and different approximations of the uncertainty sets. Neither of them is straightforward, which is why we introduce the concepts one after the other. The presented models can be considered simplifications of the models presented by Ardestani-Jaafari and Delage [2018]. We start by presenting the initial formulation of the robust counterpart, leaning on the formulation presented by Baron et al. [2010]. We then gradually extend the formulation to match state-of-the-art modeling techniques that limit over-conservatism.

Consider the following robust version of the CFLP.

Adjustable Robust CFLP I

$$\max \min_{\xi \in \mathcal{U}} -\sum\nolimits_i F_i y_i + \sum\nolimits_i \sum\nolimits_j (r_j - c_{ij}) \tilde{D}_j(\xi) x_{ij} \tag{7.15}$$

$$\text{s.t.} \sum\nolimits_i x_{ij} \leq 1 \qquad j \in J \tag{7.16}$$

$$\sum\nolimits_j \tilde{D}_j(\xi) x_{ij} \leq Q_i y_i \qquad i \in I,\ \xi \in \mathcal{U} \tag{7.17}$$

$$x_{ij} \geq 0 \qquad i \in I,\ j \in J \tag{7.18}$$

$$y_i \in \{0,1\} \qquad i \in I. \tag{7.19}$$

In contrast to the two-stage stochastic program, where allocation decisions can be taken after uncertainty is disclosed, the above robust formulation requires all decisions to be taken upfront. However, notice that the fractional variable-based formulation of the allocation decisions implies a certain degree of adaptivity. The actual transport volumes in any scenario linearly depend on the realized demand. In particular, this can be considered as a naive decision rule in an adjustable robust counterpart, taking the form $X^I_{ij}(\tilde{D}_j) := x_{ij}(D^0_j + \xi\hat{D}_j)$.

A simple way to transform the max-min profit objective into a simple maximization problem is by means of an auxiliary decision variable v, such that the objective (7.15) is replaced by

$$\begin{aligned}
&\max \quad v && && && (7.20)\\
&\text{s.t.} -\textstyle\sum_i F_i y_i + \sum_i \sum_j (r_j - c_{ij})\tilde{D}_j(\xi)x_{ij} && \geq v && \xi \in \mathcal{U} && (7.21)\\
&\qquad v && \geq 0. && && (7.22)
\end{aligned}$$

Two constraints of the resulting problem are affected by uncertainty: the constraint representing the objective function (7.21), and the capacity constraint (7.17). Evidently, for uncertainty sets that are defined by intervals or hypercubes, these constraints cannot be handed over to state-of-the-art solvers as they require an infinite number of constraints to account for all possible realizations of the uncertain parameters. A straightforward, tractable, but over-conservative robust counterpart formulation is obtained by replacing the uncertain parameters with their extreme values, such that we obtain the following.

Adjustable Robust Counterpart to the CFLP (ARC-CFLP)-I

$$\begin{aligned}
&\max \quad v && && && (7.23)\\
&\text{s.t.} -\textstyle\sum_i F_i y_i\\
&\qquad + \sum_i \sum_j \max\{0, (r_j - c_{ij})\} \min_{\xi\in\mathcal{U}} \tilde{D}_j(\xi)x_{ij} && \geq v && && (7.24)\\
&\quad \textstyle\sum_i x_{ij} && \leq 1 && j \in J && (7.25)\\
&\quad \textstyle\sum_j \max_{\xi\in\mathcal{U}} \tilde{D}_j(\xi)x_{ij} && \leq Q_i y_i && i \in I && (7.26)\\
&\quad v,\ x_{ij} && \geq 0 && i \in I,\ j \in J && (7.27)\\
&\quad y_i && \in \{0,1\} && i \in I. && (7.28)
\end{aligned}$$

Thereby, the minimum and maximum realizations of $\xi \in \mathcal{U}$ can be straightforwardly determined as $(-1,\ldots,-1)$ and $(1,\ldots,1)$, respectively. Consequently, we know that

$$\min_{\xi\in\mathcal{U}} \tilde{D}_j(\xi) = D^0_j - \hat{D}_j, \tag{7.29}$$

and

$$\max_{\xi\in\mathcal{U}} \tilde{D}_j(\xi) = D^0_j + \hat{D}_j. \tag{7.30}$$

The reformulations rely on the fact that the greater equal constraints hold for all realization of the uncertain parameters if they hold for the minimum parameter realization and, similarly, the less or equal constraints hold for all realizations if they hold for the maximum parameter realization. Thereby, the reformulation of constraint (7.21) as constraint (7.24) must further restrict all considerations to allocations generating a positive net unit profit. Consequently, we only consider those allocations that yield a net profit greater than 0 and denote $\eta_{ij} := \max\{0, (r_j - c_{ij})\}$. This way it is ensured that the model never considers it profitable to allocate customers to a facility that requires unit transportation costs that outweigh unit profits. Notice that η_{ij} can be evaluated in the preprocessing of any model as no decision variables are involved.

The above robust counterpart formulation is overly conservative for two reasons. First, since we model the uncertainty set as a hypercube, all demands may take on their worst-case realization simultaneously. While these worst-case scenarios, in which all parameters taken on their minimum or maximum value are also included in the scenario set Ω considered in the stochastic program, the robust counterpart only takes these worst-case realizations into account. This leads to very conservative solutions.

Secondly, as pointed out by Ardestani-Jaafari and Delage [2018], the fractional-variable-based formulation (decision rule X^I_{ij}) makes the model overly conservative. Since the uncertain demand affects both the objective and the capacity constraint, and both are "protected" simultaneously, the model requires building capacities to serve the maximum demand while expecting the minimal demand to generate profit. A consequence of the fractional variable-based formulation is that every small fraction of the demand that is served requires arbitrarily large capacities to be installed. As a result, the model cannot fully exploit established capacities. Neither of the above problems is new, and we address each separately with state-of-the-art countermeasures.

7.2.3.1. Affine decision-rules for allocation decisions

To overcome the conservatism induced by the fractional-variable-based formulation, Ardestani-Jaafari and Delage [2018] suggest adding a constant component to the decision rule representing the allocation decision, such that $X^{II}_{ij}(\tilde{D}_j(\xi)) := w_{ij} + x_{ij}\xi\hat{D}_j$, with $w_{ij} \geq 0$ the constant component of the allocation volume. The functional form of the resulting decision rule is illustrated in Figure 7.2b. As the following robust counterpart formulation shows, adding the constant component to the decision rule implies that uncertainty must also be considered in the demand satisfaction constraint (7.33) and the non-negativity constraint for the total transport volume (7.35). It is easy to see that both constraints are restricted most strongly when the demand is minimal.

ARC-CFLP-II

$$\max\ v \tag{7.31}$$

$$\text{s.t.} -\sum\nolimits_i F_i y_i + \sum\nolimits_i \sum\nolimits_j \eta_{ij} \left(w_{ij} + x_{ij} \xi \hat{D}_j \right) \geq v \qquad \xi \in \mathcal{U} \tag{7.32}$$

$$\sum\nolimits_i w_{ij} + x_{ij} \xi \hat{D}_j \leq D_j^0 + \xi \hat{D}_j \qquad j \in J,\ \xi \in \mathcal{U} \tag{7.33}$$

$$\sum\nolimits_j w_{ij} + x_{ij} \xi \hat{D}_j \leq Q_i y_i \qquad i \in I,\ \xi \in \mathcal{U} \tag{7.34}$$

$$w_{ij} + x_{ij} \xi \hat{D}_j \geq 0 \qquad i \in I,\ j \in J \tag{7.35}$$

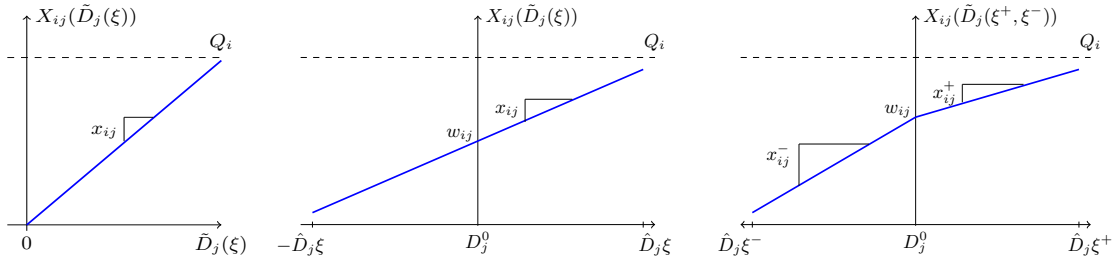
$$v,\ w_{ij},\ x_{ij} \geq 0 \qquad i \in I,\ j \in J,\ \xi \in \mathcal{U} \tag{7.36}$$

$$y_i \in \{0, 1\} \qquad i \in I. \tag{7.37}$$

Furthermore, Ardestani-Jaafari and Delage [2018] suggest lifting the uncertainty space to allow treating positive and negative perturbations differently. As depicted in Figure 7.2c, this means that the allocation decision rule is approximated by a piece-wise linear function. In particular, each primitive uncertainty ξ is split into a positive and a negative component ξ^+ and ξ^-, such that the uncertain demand becomes $\tilde{D}_j(\xi^+, \xi^-) := D_j^0 + \hat{D}_j \xi^+ - \hat{D}_j \xi^-$. Let m denote the number of primitive uncertainties considered in the given uncertainty setting. Then, the lifted uncertainty set is formally described by

$$\mathcal{U}_2 := \left\{ (\xi^+, \xi^-) \in [0,1]^m \times [0,1]^m \middle| \, ||\xi^+ + \xi^-||_\infty \leq 1, \right\}. \tag{7.38}$$

The resulting affine decision rule for the allocation decision becomes $X_{ij}^{III}(\tilde{D}_j(\xi^+, \xi^-)) := w_{ij} + \left(x_{ij}^+ \xi^+ - x_{ij}^- \xi^- \right) \hat{D}_j$. Integrating this decision rule into the adaptive robust CFLP from above is straightforward to derive the ARC-CFLP-III.



(a) X_{ij}^I **(b)** X_{ij}^{II} **(c)** X_{ij}^{III}

Figure 7.2.: Increasingly flexible decision rules Ardestani-Jaafari and Delage [2018]

The optimal min-max objective value of the adjustable robust counterpart is denoted by $z_{RC}^\star$. The results demonstrate the over-conservatism of the fractional-variable-based decision rule (X_{ij}^I). As the robust counterpart maximizes the worst-case profit, one intuitively assumes that $z_{RC}^\star$ is the lowest attainable profit across all potential realizations of the uncertain parameter. However, $z_{WC}^\star$ denotes the "actual" worst-case profit that is obtained when the location decisions from the robust counterpart $y_{RC}^\star$ are fixed in the static CFLP, and the problem is solved for all possible realizations of ξ, in particular for $(-1, \ldots, -1)$.

Definition 7.2. *The **worst-case gap**, $WC-gap$, denotes the relative arithmetic difference between the objective value of the (adjustable) robust counterpart, $z^{\star}_{RC}$, and the worst realization of the objective value in CFLP, $z^{\star}_{RC}(\xi^{\omega})$, in which location decisions have been fixed to those of the (adjustable) robust counterpart and demands are set to the worst-case realization of the uncertain parameter $\xi^{\omega} \in \mathcal{U}$ such that*

$$WC - gap := \max\{0, \frac{z^{\star}_{WC} - z^{\star}_{RC}}{z^{\star}_{WC}}\}, \text{ with } z^{\star}_{WC} = \min_{\xi^{\omega} \in \mathcal{U}} z^{\star}_{RC}(\xi^{\omega}). \tag{7.39}$$

The $WC - gap$ precisely captures the over-conservatism that results from approximating a fully flexible adaptive decision rule by an affine approximation.

Example C 7.2 f (Different affine decision rule approximations in adjustable robust counterparts) Table 7.2 displays the results obtained for robust counterparts with different decision rules for instances $\tilde{\mathcal{P}}_1$ and $\tilde{\mathcal{P}}_2$. The perturbation magnitude p is again varied between 5%, 10%, and 20%. The table displays $z^{\star}_{RC}$, $z^{\star}_{WC}$, and $WC - gap$ as means of comparison in the objective value space and compares the set of solutions operating in the adjustable robust counterpart $I^{\star}_{RC}$ to those operating in the CFLP in which all parameters have been set to their nominal (expected) values, $I^{\star}_{EV}$ using the overlap coefficient and the Jaccard index. The results indicate a significant worst-case gap for instance $\tilde{\mathcal{P}}_1$ and decision rule X^{I}_{ij}. For $p = 5\%$, the robust counterpart underestimates the worst-case profit by 70.4%. For $p = 10\%$, the worst-case profit is underestimated by 95.9%. For $\tilde{\mathcal{P}}_2$, the worst-case gap is significantly smaller and does not exceed 1.6%. The restrictive capacities in $\tilde{\mathcal{P}}_1$ contribute strongly to the worst-case gap as to be profitable, facilities must be highly utilized. However, as the fractional variable-based formulation requires that the capacity constraint must also hold for the highest demand realization, high utilization is impossible. For $\tilde{\mathcal{P}}_2$, lower utilization levels are optimal, reducing the negative effect of the required capacity buffers.

Including a constant allocation volume to the decision rule effectively solves the problem, and the worst-case gap gap reduces to 0 for X^{II}_{ij} and X^{III}_{ij}. In particular, lifting the uncertainty set does not improve the objective value for the considered instances. Ardestani-Jaafari and Delage [2018] already point out that the robust counterpart to a location-transportation problem can ignore positive perturbations as there is always a worst-case optimal solution that adapts to negative demand perturbations but ignores positive perturbations. As a consequence, x^{+}_{ij} is 0 in all optimal solutions obtained with $X_{ij}(\xi)^{III}$.

In addition, Table 7.2 displays the overlap between the subset of facilities operating in the robust counterpart and those facilities operating in expected or nominal value problem, the CFLP in which all parameters are set to their average or nominal values. The results strongly resemble those from stochastic programming. For $\tilde{\mathcal{P}}_1$, the robust counterpart opens a subset of those facilities operating in the expected value problem. Thereby, the number of operating facilities reduces significantly, i.e., in the case of a fractional variable-based decision rule (X^{I}_{ij}). Meanwhile, for $\tilde{\mathcal{P}}_2$, the set of facilities operating in the optimal solution changes substantially throughout all considered settings.

	p	X_{ij}	$z^\star_{RC}$	$z^\star_{WC}$	$WC - gap$	$overlap(I^\star_{RC}, I^\star_{EV})$	$Jacc(I^\star_{RC}, I^\star_{EV})$
$\tilde{\mathcal{P}}_1$	5%	I	506.8	1712.3	70.4	1.0	0.5
		II	2274.9	2274.9	0.0	1.0	0.9
		III	2274.9	2274.9	0.0	1.0	0.9
	10%	I	11.8	288.3	95.9	1.0	0.1
		II	2174.2	2174.2	0.0	1.0	0.9
		III	2174.2	2174.2	0.0	1.0	0.9
	20%	I	0.0	0.0	0.0	0.0	0.0
		II	1952.9	1952.9	0.0	1.0	0.8
		III	1952.9	1952.9	0.0	1.0	0.8
$\tilde{\mathcal{P}}_2$	5%	I	18183.6	18242.4	0.3	0.7	0.6
		II	18756.7	18756.7	0.0	0.5	0.3
		III	18756.7	18756.7	0.0	0.5	0.3
	10%	I	16509.5	16567.1	0.3	0.7	0.6
		II	17682.6	17682.6	0.0	0.7	0.4
		III	17682.6	17682.6	0.0	0.7	0.4
	20%	I	12970.5	13178.1	1.6	0.3	0.2
		II	15511.4	15511.4	0.0	0.6	0.3
		III	15511.4	15511.4	0.0	0.6	0.3

Table 7.2.: Robust counterpart solution for different decision rule approximations ($\tilde{\mathcal{P}}_1$-$\tilde{\mathcal{P}}_2$, Ex. C)

▲

When uncertainty equals a hypercube, distinguishing between different uncertainty settings in the form of a different number of primitive uncertainties is obsolete. All primitive uncertainties take their worst-case values at either 1 or -1, which is all that is considered in the above problem formulations. This changes when we consider polyhedral uncertainty sets in the following.

7.2.3.2. Polyhedral uncertainty

Uncertainty sets in the form of hypercubes imply that the decision maker is willing to protect a solution against the simultaneous deviation of all parameters to their worst-case value. As Example C 7.2 demonstrates, this may come at potentially high costs. Bertsimas and Sim [2004] propose polyhedral uncertainty sets. As pointed out in Section 7.1, polyhedral uncertainty sets reflect the intuitive idea that it is unlikely that all parameters deviate to their worst-case values simultaneously. They allow a decision-maker to choose a desired protection level by specifying a maximum number of uncertain parameters that may deviate to their worst-case realizations simultaneously. Formally, this means that in any constraint with m uncertain parameters, the decision maker may choose a parameter $\Gamma \in [0, m]$ so that no more than Γ uncertain parameters may deviate to their worst-case value at the same time. Depending on whether we consider global, local, or regional uncertainty settings, we consider a different number of primitive uncertainties. This means the number of uncertain parameters that affect a particular constraint differs, hence the upper bound of the protection level Γ. Formally, the polyhedral uncertainty set can be described as follows:

$$\mathcal{U}_2^{pol.} := \left\{(\xi^+, \xi^-) \in [0,1]^m \times [0,1]^m \middle| ||\xi^+ + \xi^-||_\infty \leq 1, ||\xi^+ + \xi^-||_1 \leq \Gamma \right\}. \tag{7.40}$$

Including polyhedral uncertainty sets in the problem formulation is not trivial, and we will describe the implications for each constraint in the following using X_{ij}^{III}. For demonstrative purposes, we assume that each customer is affected by a single primitive uncertainty ξ_j. We discuss the implications of assuming global or regional uncertainty settings after presenting the model.

Uncertain objective

The reformulation idea of Bertsimas and Sim [2004] relies on exploiting the duality of the inner maximization problem. Thus, for the objective constraint, we first reformulate objective (7.15) from a max-min profit to a min-max net cots formulation, such that

$$\min_{\mathbf{x}^+,\mathbf{x}^-,\mathbf{w},\mathbf{y}} \max_{(\xi^+,\xi^-)\in\mathcal{U}_2^{pol.}} \sum\nolimits_i F_i y_i + \sum\nolimits_i \sum\nolimits_j -\eta_{ij}\left(w_{ijt} + (x_{ij}^+\xi^+ - x_{ij}^-\xi^-)\hat{D}_j\right). \tag{7.41}$$

Then, we separate the deterministic from the uncertain part, such that

$$\underbrace{\sum\nolimits_i F_i y_i + \sum\nolimits_i \sum\nolimits_j -\eta_{ij} w_{ij}}_{\text{deterministic}} + \underbrace{\max\nolimits_{(\xi_j^+,\xi_j^-)\in\mathcal{U}_2^{pol.}} \left\{\sum\nolimits_i \sum\nolimits_j -\eta_{ij}\hat{D}_j(x_{ij}^+\xi_j^+ - x_{ij}^-\xi_j^-)\right\}}_{\beta^{obj}(\Gamma^{obj},x_{ij}^{+\star},x_{ij}^{-\star})} \leq v. \tag{7.42}$$

The inner maximization problem $\beta^{obj}(\Gamma^{obj}, x_{ij}^{+\star}, x_{ij}^{-\star})$, is defined as follows

$$\max_{(\xi_j^+,\xi_j^-)\in\mathcal{U}_2^{pol.}} \sum\nolimits_j \hat{D}_j\left(\sum\nolimits_i -\eta x_{ij}^{+\star}\xi_j^+ - \sum\nolimits_i -\eta x_{ij}^{-\star}\xi_j^-\right) \tag{7.43}$$

$$\text{s.t.} \sum\nolimits_j \left(\xi_j^+ + \xi_j^-\right) \leq \Gamma^{obj} \qquad \left(Z^{obj}\right) \tag{7.44}$$

$$0 \leq \xi_j^+ + \xi_j^- \leq 1 \qquad \forall j \in J \quad \left(p_j^{obj}\right), \tag{7.45}$$

the dual to which is

$$\min_{p,Z} \Gamma^{obj} Z^{obj} + \sum\nolimits_j p_j^{obj} \tag{7.46}$$

$$\text{s.t. } Z^{obj} + p_j^{obj} \geq \sum\nolimits_i -\eta_{ij} x_{ij}^{+\star}\hat{D}_j \qquad \forall j \in J \tag{7.47}$$

$$Z^{obj} + p_j^{obj} \geq -\sum\nolimits_i -\eta_{ij} x_{ij}^{-\star}\hat{D}_j \qquad \forall j \in J \tag{7.48}$$

$$Z^{obj}, p_j^{obj} \geq 0 \qquad \forall j \in J. \tag{7.49}$$

Given that $\eta_{ij} \geq 0$ per definition, as well as $x_{ij}^{+\star} \geq$ and $\hat{D}_j \geq$, it is obvious that constraint (7.47) is always fulfilled as it is implicitly ensured by the non-negativity constraints on the dual variables (7.49). It can thus be omitted, and only constraints (7.48) and (7.49) must be included in the robust counterpart formulation. This reformulation thus shows what we concluded from the robust counterpart with box uncertainty: the worst-case objective is only affected by negative perturbations of the demand parameters.

Demand satisfaction constraint

Separating the deterministic and the uncertain part of the demand satisfaction constraint yields

$$\sum\nolimits_i w_{ij} + \underbrace{\max\nolimits_{(\xi_j^+,\xi_j^-)\in\mathcal{U}_2^{pol.}} \hat{D}_j \left\{\sum\nolimits_i x_{ij}^+\xi_j^+ - \sum\nolimits_i x_{ij}^-\xi_j^- - \xi_j^+ + \xi_j^-\right\}}_{\beta_j^{dem}\left(\Gamma_j^{dem},x_{ij}^{+\star},x_{ij}^{-\star}\right)} \leq D_j^0 \quad \forall j \in J. \tag{7.50}$$

As for all considered uncertainty settings, the uncertain customer demand is affected by at most one primitive uncertainty, we obtain the following inner maximization problem

$$\max_{(\xi_j^+,\xi_j^-)\in[0,1]^2} \hat{D}_j\Big(\left(\sum\nolimits_i x_{ij}^{+\star} - 1\right)\xi_j^+ + \left(1 - \sum\nolimits_i x_{ij}^{-\star}\right)\xi_j^-\Big) \tag{7.51}$$

$$\text{s.t.} \xi_j^+ + \xi_j^- \leq 1 \qquad \left(Z^{dem}\right) \tag{7.52}$$

$$\xi_j^+, \xi_j^- \geq 0 \qquad \forall j \in J. \tag{7.53}$$

Thus, the dual is

$$\min_{p,Z} Z_j^{dem} \tag{7.54}$$

$$\text{s.t.} Z_j^{dem} \geq \left(\sum\nolimits_i x_{ij}^{+\star} - 1\right)\hat{D}_j \tag{7.55}$$

$$Z_j^{dem} \geq \left(1 - \sum\nolimits_i x_{ij}^{-\star}\right)\hat{D}_j \tag{7.56}$$

$$Z_j^{dem} \geq 0. \tag{7.57}$$

As $\sum_i x_{ij}^+$ is the fraction of the excess demand served by any facility, we can assume that in the optimal solution, it holds that $\sum_i x_{ij}^+ \leq 0$. Constraint (7.55) can thus be omitted from the formulation, and, again, the constraint is only affected by negative demand perturbations.

Capacity constraint

Separating the deterministic and the uncertain part yields:

$$\sum\nolimits_j w_{ij} + \underbrace{\max\nolimits_{(\xi^+,\xi^-)\in\mathcal{U}_2^{pol.}} \sum\nolimits_j \hat{D}_j\left(x_{ij}^+\xi_j^+ - x_{ij}^-\xi_j^-\right)}_{\beta_i^{cap}\left(\Gamma_i^{cap},x_{ij}^{+\star},x_{ij}^{-\star}\right)} \leq Q_i y_i \qquad \forall i. \tag{7.58}$$

The inner maximization problem is

$$\max_{(\xi_j^+,\xi_j^-)\mathcal{U}_2^{pol.}} \sum\nolimits_j \hat{D}_j\left(x_{ij}^{+\star}\xi_j^+ - x_{ij}^{-\star}\xi_j^-\right) \tag{7.59}$$

$$\text{s.t.} \sum\nolimits_j \left(\xi_j^+ + \xi_j^-\right) \leq \Gamma_i^{cap} \qquad \left(Z_i^{cap}\right) \tag{7.60}$$

$$0 \leq \xi_j^+ + \xi_j^- \leq 1 \qquad \forall j \in J \qquad \left(p_{ij}^{cap}\right). \tag{7.61}$$

which yields the following dual

$$\min_{p,Z} \Gamma_i^{cap} Z_i^{cap} + \sum\nolimits_j p_{ij}^{cap} \tag{7.62}$$

$$\text{s.t.} Z_i^{cap} + p_{ij}^{cap} \geq x_{ij}^{+\star} \hat{D}_j \qquad \forall j \in J \tag{7.63}$$

$$Z_i^{cap} + p_{ij}^{cap} \geq -x_{ij}^{-\star} \hat{D}_j \qquad \forall j \in J \tag{7.64}$$

$$Z_i^{cap}, p_{ij}^{cap} \geq 0 \qquad \forall j \in J. \tag{7.65}$$

Given that $x_{ij}^{-\star} \geq 0$ and $\hat{D}_j \geq 0$, it is evident that Equation 7.64 can be dropped and only positive demand perturbations affect the uncertain capacity constraints.

Non-negative transport volumes

Separating the deterministic and the uncertain parts yields

$$w_{ij} + \min\nolimits_{(\xi^+,\xi^-)\in\mathcal{U}_2^{pol.}} \hat{D}_j \left(x_{ij}^+ - x_{ij}^-\right) \geq 0 \tag{7.66}$$

$$\Leftrightarrow w_{ij} - \underbrace{\max\nolimits_{(\xi^+,\xi^-)\in\mathcal{U}_2^{pol.}} \hat{D}_j \left(-x_{ij}^+ x_{ij}^-\right)}_{\beta_{ij}^{nn}\left(\Gamma_{ij}^{nn}, x_{ij}^{+\star}, x_{ij}^{-\star}\right)} \geq 0. \tag{7.67}$$

Again, there is at most one primitive uncertainty affecting the non-negativity constraint, such that the inner optimization problem is:

$$\max_{(\xi_j^+,\xi_j^-)\mathcal{U}_2^{pol.}} \hat{D}_j(-x_{ij}^{+\star} + x_{ij}^{-\star}) \tag{7.68}$$

$$\text{s.t.} \xi_j^+ + \xi_j^- \leq 1 \qquad (Z_i^{nn}) \tag{7.69}$$

$$\xi_j^+, \xi_j^- \geq 0. \tag{7.70}$$

Thus, the dual is

$$\min_{p,Z} Z_{ij}^{nn} \tag{7.71}$$

$$\text{s.t.} Z_{ij}^{nn} \geq -x_{ij}^{+} \hat{D}_j \tag{7.72}$$

$$Z_{ij}^{nn} \geq x_{ij}^{-} \hat{D}_j \tag{7.73}$$

$$Z_{ij}^{nn} \geq 0. \tag{7.74}$$

Once again, the non-negativity constraint is only affected by negative demand perturbations.

Adaptive robust counterpart with polyhedral uncertainty

Given the above reformulation of the uncertain constraints, we obtain the following adaptive robust counterpart with polyhedral uncertainty, which we use in the following.

ARC-CFLP

$$\min_{\substack{x^+,x^-,v,\\ w,y,p,Z}} \; v \tag{7.75}$$

$$\text{s.t.} \sum\nolimits_i F_i y_i + \sum\nolimits_i \sum\nolimits_j -\eta_{ij} w_{ij} + \sum\nolimits_j p_j^{obj} + Z^{obj}\Gamma^{obj} \leq v \tag{7.76}$$

$$Z^{obj} + p_j^{obj} \geq -\sum\nolimits_i -\eta_{ij}\hat{D}_j x_{ij}^- \quad \forall j \in J \tag{7.77}$$

$$\sum\nolimits_i w_{ij} + Z_j^{dem} \leq D_j^0 \quad \forall j \in J \tag{7.78}$$

$$Z_j^{dem} \geq \left(1 - \sum\nolimits_i x_{ij}^-\right)\hat{D}_j \quad \forall j \in J \tag{7.79}$$

$$\sum\nolimits_j \left(w_{ij} + p_{ij}^{cap}\right) + \Gamma^{cap} Z_i^{cap} \leq Q_i y_i \quad \forall i \in I \tag{7.80}$$

$$Z_i^{cap} + p_{ij}^{cap} \geq \hat{D}_j x_{ij}^+ \quad \forall j \in J, i \in I \tag{7.81}$$

$$w_{ij} - Z_{ij}^{nn} \geq 0 \quad \forall j \in J, i \in I \tag{7.82}$$

$$Z_{ij}^{nn} \geq x_{ij}^- \hat{D}_j \quad \forall j \in J, i \in I \tag{7.83}$$

$$Z^{obj}, p_j^{obj}, Z_j^{dem}, Z_i^{cap}, p_{ij}^{cap}, x_{ij}^+, x_{ij}^-, w_{ij} \geq 0 \quad \forall j \in J, i \in I \tag{7.84}$$

$$y_i \in \{0,1\} \quad \forall i \in I. \tag{7.85}$$

The ARC-CFLP was derived assuming a local uncertainty setting in which each customer is affected by a single primitive uncertainty. When moving, e.g., to a global uncertainty setting, there is only a single primitive uncertainty ξ, and the uncertainty set $\mathcal{U}^{box}$ equals the interval $[-1, 1]$. Restricting this interval by setting a budget of uncertainty $\Gamma \in [0, 1]$ essentially implies shrinking this interval. This means that the reformulation of those constraints that consider more than one primitive uncertainty simplifies. In particular, the reformulation of the objective function simplifies from (7.76) and (7.77) to

$$\sum\nolimits_i F_i y_i + \sum\nolimits_i \sum\nolimits_j -\eta_{ij} w_{ij} + Z^{obj}\Gamma^{obj} \leq v \tag{7.86}$$

$$Z^{obj} \geq -\sum\nolimits_j \sum\nolimits_i -\eta_{ij}\hat{D}_j x_{ij}^-. \tag{7.87}$$

Similarly, the reformulation of the capacity constraints (7.80) and (7.81) simplifies to

$$\sum\nolimits_j \left(w_{ij} + p_{ij}^{cap}\right) + \Gamma^{cap} Z_i^{cap} \leq Q_i y_i \quad \forall \in I \tag{7.88}$$

$$Z_i^{cap} \geq \sum\nolimits_j \hat{D}_j x_{ij}^+ \quad \forall i \in I. \tag{7.89}$$

Meanwhile, the reformulations of those constraints that have always been affected by only a single primitive uncertainty, i.e., the demand satisfaction constraints (7.78)-(7.79) and the non-negativity constraints (7.82)-(7.83), remain unchanged. The robust counterpart protects individual constraints against uncertain perturbations independently of one another. Thus, whether each constraint is affected by the same or different primitive uncertainties does not make any difference. Obtaining reformulation for a regional uncertainty setting is straightforward. In the following, we assume the same protection level Γ for each constraint and denote the optimal objective value to the ARC-CFLP for a given value of Γ

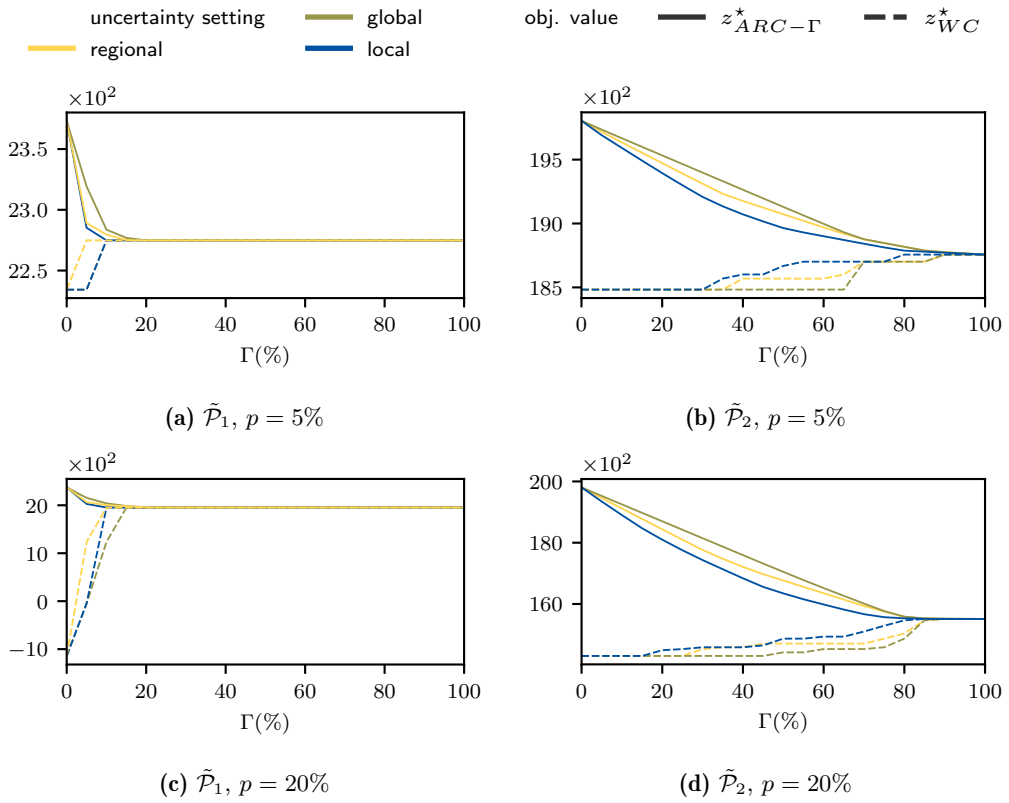
by $z^{\star}_{ARC-\Gamma}$.

Example C 7.3 (Adjustable robust counterparts with polyhedral uncertainty sets) We solve $\tilde{\mathcal{P}}_1$ and $\tilde{\mathcal{P}}_2$ with different uncertainty settings and perturbation levels $p = 5\%$ and $p = 20\%$ assuming a polyhedral uncertainty set with different budgets of uncertainty Γ. Different uncertainty settings are dependent on a different number of primitive uncertainties. Consequently, there are different upper bounds for Γ. Therefore, we express the uncertainty budget as a percentage of its maximum value. If $\Gamma = 0\%$, uncertainty is completely ignored, and, essentially, the nominal (or expected) value problem is solved. If $\Gamma = 100\%$, the model protects the decision-maker against all parameters deviating to their worst case simultaneously, thus reflecting an uncertainty set in the form of a hypercube.

The development of $z^{\star}_{ARC-\Gamma}$ and the $z^{\star}_{WC}$ are displayed Figure 7.3 for different protection levels Γ. Notice that this time, the objective value of the robust counterpart generally exceeds the worst-case realization of the objective value. Decision rule X^{II} is sufficiently flexible to approximate the adjustable decision rule for the allocation decisions. Instead, the lower protection levels Γ result in the ARC-CFLP underestimating the negative effect of potential perturbations. The worst-case realization of the objective value in the CFLP is obtained for a realization of the uncertain parameter that is not considered in the polyhedral uncertainty set with protection level Γ. Throughout all considered problems, we observe that reducing Γ may result in significantly lower worst-case objective values than had one protected against that worst-case realization from the beginning. In other words, for $\Gamma = 0\%$, $z^{\star}_{WC}$ is less than $z^{\star}_{WC}$ for $\Gamma = 100\%$. Furthermore, one can see that the discrepancy between the objective value and the worst-case realization is largest for global uncertainty, which considers only a single primitive uncertainty parameter. The discrepancy is smallest for local uncertainty. This indicates that a budget of uncertainty is particularly effective when it allows the consideration of trade-offs between different primitive uncertainties rather than simply reducing the considered interval of a single primitive uncertainty.

Once again, there is a significant difference between the development of $z^{\star}_{ARC-\Gamma}$ and $z^{\star}_{WC}$, in $\tilde{\mathcal{P}}_1$ and $\tilde{\mathcal{P}}_2$. For $\tilde{\mathcal{P}}_1$, the $z^{\star}_{ARC-\Gamma}$ and $z^{\star}_{WC}$ coincide already for a budget of $\Gamma = 20\%$. This shows that the problem reacts very sensitively to the perturbation of some of the demand parameters, approximately 20%. These customers are critical. Then, there is another 80% of the customers for which the demand may decrease significantly. Yet, it has no effect on the worst-case objective value, nor does it require a change in the location-allocation strategy. For $\tilde{\mathcal{P}}_2$, the situation is contrary. The robust counterpart and the worst-case objective decrease almost linearly with an increasing uncertainty budget Γ up until about 80%. This implies approximately 80% of the customers are critical to attaining the worst-case objective value, while 20% are irrelevant.

Notice that the analysis of Figure 7.3 is usually what is referred to as analyzing the "price of robustness" in robust optimization literature.



(a) $\tilde{\mathcal{P}}_1$, $p = 5\%$ **(b)** $\tilde{\mathcal{P}}_2$, $p = 5\%$

(c) $\tilde{\mathcal{P}}_1$, $p = 20\%$ **(d)** $\tilde{\mathcal{P}}_2$, $p = 20\%$

Figure 7.3.: Robust counterpart objective and worst-case objective value for different uncertainty settings and magnitudes with an increasing uncertainty budget Γ ($\tilde{\mathcal{P}}_1$-$\tilde{\mathcal{P}}_2$, Ex. C)

▲

In our view, the fact that robust counterparts have been developed for row-wise uncertainty models and the customer demand in the CFLP is a parameter that affects several constraints, leads to a discrepancy between what one intuitively wants to achieve with a robust approach and what the model actually does. Baron et al. [2010] claim that their robust counterpart formulation simultaneously protects a decision maker against high demand, and therefore, high capacity requirements, and low demands, and therefore, low potential profit. However, Ardestani-Jaafari and Delage [2018] already pointed out what we illustrated in Example C 7.3: the resulting model is over-conservative and underestimates the actual worst-case attainable profit. However, all reformulations presented by Ardestani-Jaafari and Delage [2018] lead to the fact that the model solely protects a decision-maker against low demands as these are the only ones that affect the objective value. Thereby, the model completely ignores the potential effect of high-demand realizations, which is a direct byproduct of the row-wise uncertainty model. In our view, this is a severe impediment to making robust counterpart reformulation attractive for practical applications, e.g., in supply chain network design.

Up to this point, we evaluated the solutions to the stochastic program and the robust counterparts separately and compared solutions solely to deterministic formulations that are oblivious to the underlying uncertainty. In the following, we compare the solutions obtained with a stochastic program to those obtained with robust counterparts regarding

their performance in the objective value space, as well as their similarities in the decision space. We then relate this performance to the underlying service regions.

7.3. Comparison of robust counterpart and stochastic programming solutions

We compare the solutions obtained with a stochastic program to those obtained with an adjustable robust counterpart with polyhedral uncertainty and different protection levels Γ. We start by looking at the performance in the objective value space (Subsection 7.3.1) and subsequently compare solutions in the decision space (Subsection 7.3.2).

7.3.1. Average- and worst-case performance

Despite not previously performed on instances of the CFLP, the performance of robust and stochastic solutions has been compared in several works. Usually, these works present a novel robust counterpart formulation for a problem for which a stochastic programming formulation already exists and seek to justify the relevance of the robust approach. The standard procedure considers the set of scenarios Ω used in the stochastic programming formulation. For each scenario ξ^ω in Ω, a deterministic CFLP is solved in which first-stage decisions are fixed either to those from the stochastic program or the robust counterpart. This yields an empirical distribution of the objective values obtained with either model formulation and, i.e., allows to compare the empirical worst-case and average-case realization. To the best of our knowledge, most works conclude that while the robust counterpart performs slightly worse on average, it performs significantly better in the worst case. An example can be found in Zetina et al. [2017].

We perform a similar evaluation, fixing the location decisions in the respective deterministic problems to those obtained from stochastic programming or adjustable robust optimization.

Definition 7.3. *The **AVUS** in a set of scenarios Ω with known probability π_ω is the average probability-weighted difference in the optimal objective values between the CFLPs in which demands are set to their realizations in the respective scenarios, $D_j = \hat{D}_j(\xi^\omega)$, and location decisions are fixed to either the optimal location decisions in the uncertainty-aware solution $\tilde{s}$, $z^\omega_{\tilde{s}}$, or the optimal location decisions of the expected or nominal value problem, z^ω_{EV}, such that*

$$AVUS := \sum\nolimits_{\omega\in\Omega} \pi_\omega \frac{z^\omega_{\tilde{s}} - z^\omega_{EV}}{z^\omega_{EV}}. \tag{7.90}$$

The AVUS generalizes the idea of the VSS. If $\tilde{s}$ is the optimal solution to a stochastic program, then AVUS equals the VSS. Similarly, we compare the added benefit of an uncertainty-aware approach in the worst case.

Definition 7.4. *The **WCVUS** in a set of scenarios Ω is the relative difference between the worst-case objective value across all realizations of the uncertain parameter of the CFLP*

in which location decisions are fixed to the optimal location decisions in the uncertainty-aware solution $\tilde{s}^\star$, $z^\omega_{\tilde{s}}$, and the optimal location decisions of the expected or nominal value problem, z^ω_{EV}, such that

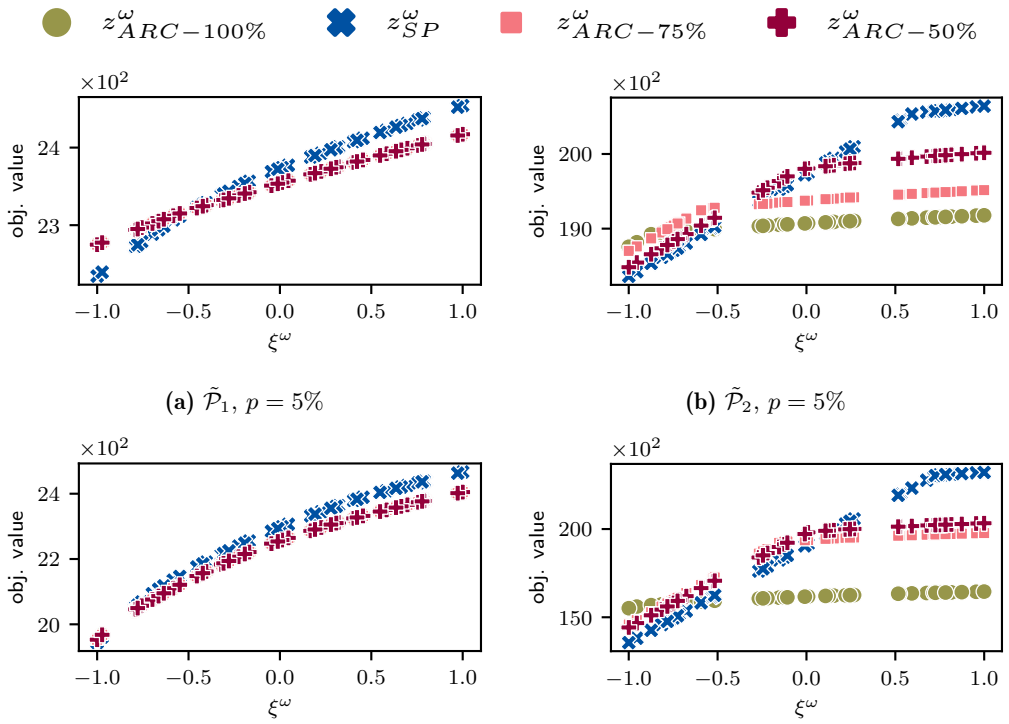
$$WCVUS := \frac{\min_{\omega\in\Omega} z^\omega_{\tilde{s}} - \min_{\omega\in\Omega} z^\omega_{EV}}{\min_{\omega\in\Omega} z^\omega_{EV}}. \tag{7.91}$$

Notice that the WCVUS compares the minimum objective value obtained with the location decisions from $\tilde{s}$ to the minimum objective value obtained with the location decisions from the expected value problem. The scenarios for which these minima are attained may differ. This, WCVUS measures the relative improvement of the objective value in the worst case and not the worst-case improvement of the objective value.

Example C 7.4 (AVUS and WCVUS) Figure 7.4 displays the empirical distribution of objective values obtained with the location decision from different models for $\tilde{\mathcal{P}}_1$ and $\tilde{\mathcal{P}}_2$, in particular, the 2SS-CFLP-DE (z^ω_{SP}), and the ARC-CFLP ($z^\omega_{RC-\Gamma}$) with protection levels $\Gamma \in \{50\%, 75\%, 100\%\}$, respectively. The figures are derived for a global uncertainty setting, allowing the realization of the uncertain parameter ξ to be displayed on a single axis. The perturbation magnitude p is 5% and 20%, respectively. Throughout all settings and solutions, we see that the robust counterpart solutions perform better in low-demand scenarios ($\xi^\omega \to -1$) and, with increasing demands ($\xi^\omega \to 1$), are at some point outperformed by the stochastic programming solution. For example, for $\tilde{\mathcal{P}}_1$ and $p = 5\%$, one can see that for low-demand realizations, the objective values attained with robust counterpart solutions exceed those obtained with the locations from stochastic programming. When ξ exceeds -0.45, the stochastic programming solution outperforms the robust counterpart. For $p = 20\%$, the objective value attained with the stochastic programming solution outperforms that of the robust counterpart already for $\xi = -0.80$. Notice that this implies that the robust counterpart is only better in about 5% of all possible scenarios. Recall that Example C 7.3 has shown that for $\tilde{\mathcal{P}}_1$, only 20% of the customers have a critical effect on the worst-case objective value. Consequently, the empirical objective values obtained for different uncertainty budgets Γ all take the same value, and the positive effect of reducing the uncertainty budget is not visible.

For instance $\tilde{\mathcal{P}}_2$, perturbations of up to 80% of the customers affected the worst-case realization. Figure 7.4b and Figure 7.4d demonstrate that decreasing the uncertainty budget Γ provides a trade-off between worst-case protection and average-case performance. The solutions obtained with $\Gamma = 75\%$ and $\Gamma = 50\%$ perform better than the stochastic programming solution and worse than the adjustable robust counterpart solution for $\Gamma = 100\%$ in the worst case. Meanwhile, with decreasing Γ, they continue outperforming the stochastic programming solution for increasingly high values of ξ^ω.

Table 7.3 displays the AVUS and WCVUS. The results confirm observations from Figure 7.4. However, we observe that the impact of explicitly modeling uncertainty is significantly higher in $\tilde{\mathcal{P}}_1$ than in $\tilde{\mathcal{P}}_2$. For $\tilde{\mathcal{P}}_1$ and $p = 20\%$, the average-case performance of models, including uncertainty, exceeds that of the expected value problem by a min-



(a) $\tilde{\mathcal{P}}_1$, $p = 5\%$ **(b)** $\tilde{\mathcal{P}}_2$, $p = 5\%$

(c) $\tilde{\mathcal{P}}_1$, $p = 20\%$ **(d)** $\tilde{\mathcal{P}}_2$, $p = 20\%$

Figure 7.4.: Empirical distribution of objective values of different realizations $\xi^{\omega} \in \Omega$ for location decisions from stochastic programming or adjustable robust counterparts ($\tilde{\mathcal{P}}_1$-$\tilde{\mathcal{P}}_2$, Ex. C)

			AVUS (%)					WCVUS (%)				
			EV	SP	ARC-Γ			EV	SP	ARC-Γ		
	p	ξ			100%	75%	50%			100%	75%	50%
$\tilde{\mathcal{P}}_1$	5%	g	2363.9	0.0	-0.5	-0.5	-0.5	2234.3	0.0	1.8	1.8	1.8
		l	2365.3	0.0	-0.6	-0.6	-0.6	2234.3	0.0	1.8	1.8	1.8
	20%	g	1848.4	22.8	20.7	20.7	20.7	-1145.0	269.8	270.6	270.6	270.6
		l	1911.9	19.4	16.8	16.8	16.8	-1145.0	194.1	270.6	270.6	270.6
$\tilde{\mathcal{P}}_2$	5%	g	19495.6	0.9	-2.3	-1.1	0.0	18482.8	-0.7	1.5	1.2	0.0
		l	19542.9	0.9	-2.5	-0.9	-0.7	18482.8	-0.7	1.5	1.2	1.0
	20%	g	18384.4	3.1	-12.4	-0.8	0.2	14303.8	-5.2	8.4	2.0	0.8
		l	18583.8	3.7	-13.2	-9.5	-2.2	14303.8	-2.8	8.4	6.9	3.9

Table 7.3.: AVUS and WCVUS for stochastic programming and adjustable robust counterpart solutions ($\tilde{\mathcal{P}}_1$-$\tilde{\mathcal{P}}_2$, Ex. C)

imum of 16.8%, e.g., for the adjustable robust counterparts. Meanwhile, the worst-case performance improves by up to 270.6%. At the same time, however, for perturbation magnitudes less than 20%, the stochastic programming solution equals that of the expected value solution, and AVUS and WCVUS are 0. In contrast, for $\tilde{\mathcal{P}}_2$, the AVUS does not exceed 3.7% for any of the uncertainty-aware models. The WCVUS is at most 8.4% for the stochastic program and a perturbation level of 20%. At the same time, it is noticeable that for $\tilde{\mathcal{P}}_1$, the adjustable robust counterparts outperform the expected value solutions even in the average case. This can be attributed to the extremely bad performance of that solution in low-demand scenarios. ▲

Example C 7.4 demonstrates that the robust counterpart and the stochastic programming solutions perform differently. As expected, the robust counterpart solution performs slightly better in the worst case and slightly worse in the average case than the stochastic programming solution. This is in line with previous findings and suggests that it is indeed up to the risk preferences of the decision-maker which modeling approach is to be preferred. Once again, the high discrepancy between the effect of uncertainty and the two different problem instances is significant, and we take a closer look at what this implies in terms of decisions taken in the following.

7.3.2. Service regions in stochastic and robust solutions

Example C 7.4 suggests that there are significant differences between the effect different degrees and different models of uncertainty in the demand parameters have on the optimal location decisions. In some instances, a subset of location decisions is remarkably robust towards both model choice and perturbation levels, while in other instances, minor perturbations result in significant changes in the optimal facilities.

In Chapter 2, we identified service regions of varying size as an underlying persistent decision pattern in the optimal solutions to CFLP instances. We observed that when these service regions are small in the sense that their customers are best served by a single facility, these facilities appear to be relatively persistent throughout well-performing solutions. However, when these service regions are large and optimally served by a combination of interdependent facilities, these subsets of facilities are relatively sensitive to minor changes in the data. Meanwhile, the implied division into service regions is relatively persistent and denotes the characterizing decision pattern underlying optimal and near-optimal solutions to that instance.

We evaluate the degree to which the decisions of robust counterparts and stochastic programs facing the same level of demand uncertainty differ. While the differences in AVUS and WCVUS of robust and stochastic solutions indicate that different decisions are taken, it is unclear to which degree the robust counterpart solutions simply open fewer facilities than the stochastic programming solutions. Furthermore, we evaluate the degree to which the set of optimal deterministic solutions to the individual scenarios in Ω yields insights into the solutions obtained by a stochastic program or an adjustable robust counterpart. We denote the set of solutions $\mathcal{S}^{det} := \{s^1, s^2, \ldots, s^\omega\}$. In particular, we are interested

in evaluating whether a high kernel persistence value $KER(\mathcal{S}^{det})$ is an indicator that the same facilities or a subset thereof will also operate in the stochastic programming or robust counterpart solution.

Secondly, we evaluate the degree to which the implied customer service regions derived from $\mathcal{S}^{det}$ represent the allocations in the stochastic programming or robust counterpart solution. For this purpose, we derive a set of service regions $\mathcal{R}$ from the $\mathcal{S}^{det}$ with the help of *RegClus* (see Algorithm 2 in Chapter 3). We then derive $\mathcal{S}^{SP}$ and $\mathcal{S}^{ARC-\Gamma}$ as the sets of optimal solutions provided by the stochastic program and the adjustable robust counterpart for each scenario in Ω. We fix the location decisions to those obtained with the uncertainty-aware models and solve the CFLP for the different demand scenarios. We then evaluate the degree to which the service regions obtained from the deterministic solutions represent the allocations in $\mathcal{S}^{SP}$ and $\mathcal{S}^{ARC-\Gamma}$ by determining the external loss ($\ell_{\alpha}^{external}(\mathcal{R}, \mathcal{S})$, see Def. 3.8).

Example C 7.5 (Service regions in uncertainty-aware solutions) In the following, we restrict all considerations to a local uncertainty setting, for which reducing the uncertainty budget in the adjustable robust counterpart has the most profound effect. Table 7.4 depicts $KER(\mathcal{S}^{det})$ as well as the overlap coefficient and the Jaccard index for the pairwise comparison between the set of optimal facilities operating in the expected value problem, the stochastic program, and adjustable robust counterparts. For $\tilde{\mathcal{P}}_1$, the kernel persistence is relatively high. It exceeds 85.1% for all perturbation levels. Meanwhile, there is a significant overlap between the facilities operating in the optimal solutions to the deterministic problem and the optimal solutions to the stochastic program and robust counterparts. In particular, the latter operate subsets of the expected value problem throughout. For $\tilde{\mathcal{P}}_2$, the kernel persistence in $\mathcal{S}^{det}$ is significantly lower and ranges between 54.2% and 34.2% for perturbation levels of 5% and 20%, respectively. Meanwhile, the set of facilities operating in the expected value problem differs significantly from those operating in the robust counterpart and the stochastic programming solutions. This suggests that a high kernel persistence indicates that the facilities that persist throughout the deterministic optimal solutions are also an essential component of the optimal solutions to models explicitly considering uncertainty.

	p	$KER(\mathcal{S}^{det})$ (%)	$overlap(I^{\star}_{\cdot}, I^{\star}_{EV})$			$Jacc(I^{\star}_{\cdot}, I^{\star}_{EV})$		
			SP	ARC−100%	ARC−50%	SP	ARC−100%	ARC−50%
$\tilde{\mathcal{P}}_1$	5 %	97.9	1.0	1.0	1.0	1.0	0.9	0.9
	10 %	97.1	1.0	1.0	1.0	0.9	0.9	0.9
	20 %	85.1	1.0	1.0	0.9	0.8	0.8	0.8
$\tilde{\mathcal{P}}_2$	5 %	54.2	0.7	0.5	0.7	0.6	0.3	0.4
	10 %	46.4	0.9	0.7	0.7	0.8	0.4	0.4
	20 %	34.2	0.6	0.6	0.8	0.4	0.3	0.6

Table 7.4.: Persistence of location decisions throughout $\mathcal{S}^{det}$ as well as overlap-coefficient and Jaccard index of facilities operating in EV and SP, ARC−100%, and ARC−50% ($\tilde{\mathcal{P}}_1$-$\tilde{\mathcal{P}}_2$, Ex. C)

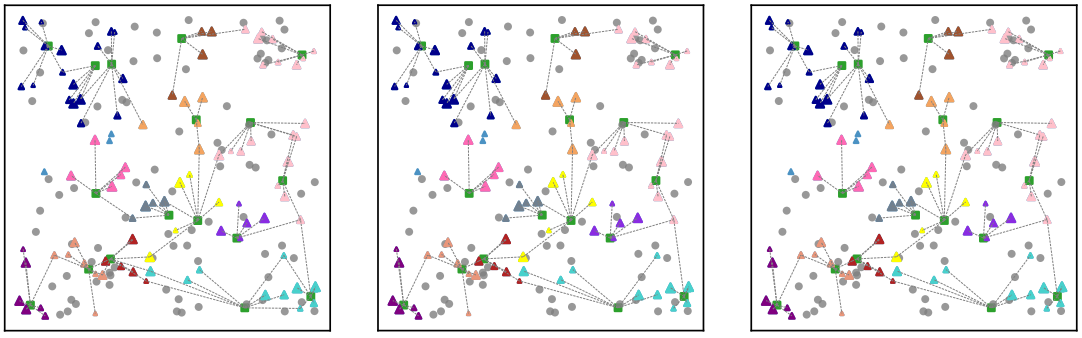
Furthermore, we determine service regions based on the set of deterministic solutions

	p (%)	$\lvert\mathcal{R}'\rvert$	$\ell_\alpha^{external}(\mathcal{R}', .)$ $\mathcal{S}^{SP}$	$\mathcal{S}^{ARC-100\%}$	$\mathcal{S}^{ARC-50\%}$	average
$\tilde{\mathcal{P}}_1$	5	11	0.99	0.97	0.97	0.98
	10	10	0.98	0.91	0.91	0.93
	20	6	0.97	0.93	0.93	0.94
$\tilde{\mathcal{P}}_2$	5	4	0.87	0.98	0.96	0.94
	10	5	0.80	0.82	0.92	0.85
	20	3	0.96	0.78	0.97	0.90

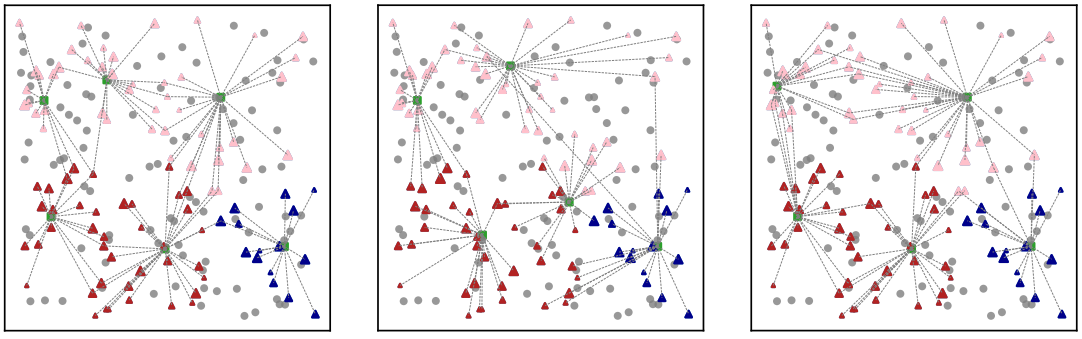
$\mathcal{R}' = \mathcal{R}(\mathcal{S}^{det}, \alpha^{profit}, 0.05)$

Table 7.5.: Persistence of implied service regions from deterministic solutions in solutions to the stochastic program and adjustable robust counterparts ($\tilde{\mathcal{P}}_1$-$\tilde{\mathcal{P}}_2$, Ex. C)

$\mathcal{S}^{det}$. Table 7.5 displays the number of regions the facility-customer space is separated into for different perturbation levels, the aggregation function *profit* and a target level of coherence of $\theta = 5$ %. We see that for both instances and all perturbation levels, several service regions can be distinguished. It furthermore shows that the optimal solutions to the stochastic program and robust counterparts adhere to these regions as the average $\ell_\alpha^{external}$ across all uncertainty-aware solutions significantly exceed 0.9 in 5 out of 6 cases. Consequently, the implied division into the customer service region is a robust decision pattern that characterizes not only optimal and near-optimal deterministic solutions but also the optimal solutions to the models explicitly considering uncertainty.



(a) $\tilde{\mathcal{P}}_1$, $\mathcal{R}$, $\mathcal{S}^{SP}$ **(b)** $\tilde{\mathcal{P}}_1$, $\mathcal{R}$, $\mathcal{S}^{ARC-100\%}$ **(c)** $\tilde{\mathcal{P}}_1$, $\mathcal{R}$, $\mathcal{S}^{ARC-50\%}$



(d) $\tilde{\mathcal{P}}_2$, $\mathcal{R}$, $\mathcal{S}^{SP}$ **(e)** $\tilde{\mathcal{P}}_2$, $\mathcal{R}$, $\mathcal{S}^{ARC-100\%}$ **(f)** $\tilde{\mathcal{P}}_2$, $\mathcal{R}$, $\mathcal{S}^{ARC-50\%}$

Figure 7.5.: Visualization of customer service regions determined from the set of deterministic solutions and $\mathcal{S}^{SP}$, $\mathcal{S}^{ARC-100\%}$, and $\mathcal{S}^{ARC-50\%}$, $p = 20\%$ ($\tilde{\mathcal{P}}_1$-$\tilde{\mathcal{P}}_2$, Ex. C)

The latter is confirmed when looking at the visualizations in Figure 7.5. It shows the

regions and the solutions to the stochastic program and robust counterparts for a perturbation level of 20%. In particular, for $\tilde{\mathcal{P}}_2$, Figure 7.5d and Figure 7.5f illustrate that when moving from a stochastic program to an adjustable robust counterpart with an uncertainty budget of 50%, the same subsets of customers are served jointly, but with fewer operating facilities. Customers are redistributed within the customer service regions that were already identified by solely considering the deterministic solutions. ▲

7.4. Experimental validation: service regions and the value of modeling uncertainty

We validate the main observations from Example C 7.4 and Example C 7.5 on the instances from the sets BAR-1991, HOL-1999, and ORLIB from Section 1.3. For each instance, we assume a local uncertainty setting and a perturbation level of $p = 20\%$. We randomly generate 50 scenarios for every instance.

Table 7.6 depicts the degree to which service regions obtained from $\mathcal{S}^{det}$ reflect the implied service regions in the solutions to individual scenarios obtained with the stochastic program, and the adjustable robust counterparts for uncertainty budgets of $\Gamma = 100\%$ and $\Gamma = 50\%$. The number of service regions identified across all instances ranges between 1 and 12. Thereby, a small number of regions implies larger customer regions and an increasing likelihood that a subset of interdependent facilities serves these regions. In contrast, a high number of regions implies that these regions are relatively small in terms of the number of customers and likely served by just a single independent facility. The high values of $\ell_\alpha^{external}(\mathcal{R}, \mathcal{S})$ indicate that the service regions derived from $\mathcal{S}^{det}$ match the service regions implied by the allocation decisions to the stochastic programming and adjustable robust counterpart solutions.

	$\lvert\mathcal{R}\rvert$										
	1	2	3	4	5	6	7	8	9	10	12
# of instances	31	54	40	16	9	7	2	6	2	3	1
$\ell_\alpha^{external}(\mathcal{R}, \mathcal{S}^{SP}))$	1.00	0.97	0.98	0.96	1.00	0.99	0.97	0.99	1.00	1.00	0.99
$\ell_\alpha^{external}(\mathcal{R}, \mathcal{S}^{ARC-100\%}))$	0.99	0.93	0.92	0.91	0.91	0.95	0.90	0.79	0.91	0.95	0.74
$\ell_\alpha^{external}(\mathcal{R}, \mathcal{S}^{ARC-50\%}))$	1.00	0.95	0.94	0.96	0.93	0.94	0.89	0.84	1.00	0.89	0.85

Table 7.6.: Average persistence of service regions in adjustable robust counterpart and stochastic programming solutions (α=*profit*, $\theta = 5\%$)

Figure 7.6 depicts the performance in terms of AVUS and WCVUS grouped by the number of service regions the facility-customer space can be partitioned into based on $\mathcal{S}^{det}$. Firstly, the results confirm previous findings from literature in the sense that while the adjustable robust solutions perform worse on average, they perform significantly better in the worst case. Yet, what is remarkable is that the magnitude with which the performance of both the stochastic programming and the adjustable robust counterpart solutions deviate from the performance of the deterministic expected value solution decreases with an increasing number of implied customer service regions.

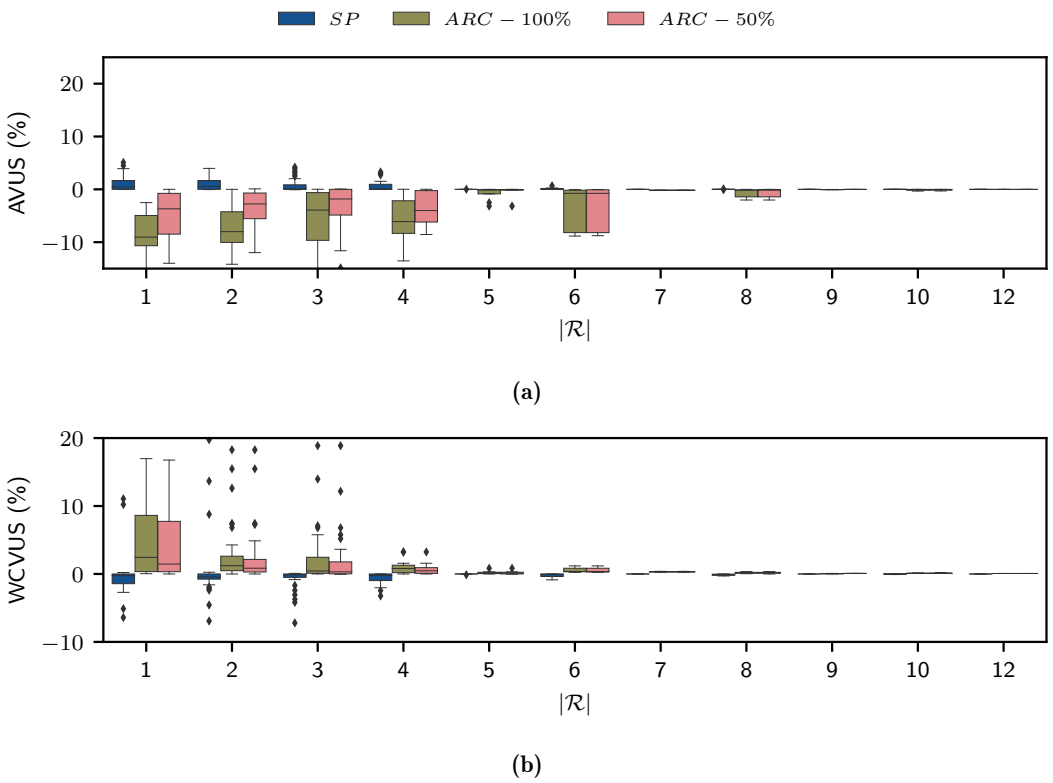


Figure 7.6.: Performance of adjustable robust counterpart and stochastic programming solutions for instances with different numbers of identifiable service regions (α=*profit*, $\theta = 5\%$) and a perturbation level p of 20%

Consider the implications of this observation. If good solutions to a particular problem instance, and in particular also to corresponding instances with perturbed demand, can be characterized by several small service regions that are served by only a few facilities, using a model that explicitly accounts for uncertainty has little advantage toward simply using the solution obtained from the expected value problem. The reason for this is simple. If, in $\mathcal{S}^{det}$, we observe that customers are not re-distributed, but for all scenarios, the implicit division into service regions remains the same, then this implies that for each region, the model has two choices: operating the facility that optimally serves this region, or not. This means that independently of the choice of the uncertainty model, the mathematical program makes the same choice for each region in this persistent decision pattern, and there is little to be gained (or lost) by moving to a more comprehensive modeling approach.

We assume, and this would be supported by the observations on $\tilde{\mathcal{P}}_1$ from Example C 7.5, that when the perturbation magnitude increases further, at some point, the robust counterpart and stochastic programming solutions suggest not to open certain facilities and consequently not to serve certain regions at all. In this case, we can expect both the AVUS and WCVUS to increase sharply, as observed for perturbations with a magnitude of 20% for $\tilde{\mathcal{P}}_1$ from Example C.

Notice that this also explains why, for $\tilde{\mathcal{P}}_1$, varying the uncertainty budget was of limited value as already a perturbation of at most 20% of the customers sufficed for the model to attain the worst-case objective value (see Figure 7.3). If the overall perturbation magnitude, the unrestricted uncertainty set, contains realizations for which it is not profitable

to operate a facility in a particular, relatively small service region, then perturbing only the customers in that region suffices to realize the aforementioned worst-case. While on a global scale, this seems like there are only a few customers with perturbed demand, in instances implicitly divided into several small service regions, the worst-case perturbation implies that all these negative perturbations concentrate on just a single service region.

7.5. Conclusion

We compare the performance of a two-stage stochastic programming formulation and an adjustable robust counterpart formulation of a CFLP with uncertainty in the customer demands. We present the average and worst-case value of the uncertainty-aware solution as a measure to quantify and compare the added value of a stochastic program or an adjustable robust counterpart. Even though we are not the first to follow this idea, we are the first to provide a formal description of this approach. Furthermore, we evaluate the degree to which uncertainty-aware models alter the optimal decisions compared to a deterministic, expected value problem. We derive characteristic decision patterns from the set of deterministic optimal solutions to all scenarios and observe that these patterns persist throughout the uncertainty-aware solutions. We subsequently analyze the degree to which the size of the underlying service regions affects the added value of uncertainty-aware models.

The main findings of the previous chapter can be summarized as follows:

- The solutions obtained with the adjustable robust counterpart perform slightly worse on average but significantly better in the worst case than those obtained with a two-stage stochastic program.
- Presented with a set of scenarios, the set of optimal deterministic solutions to these scenarios can be used to identify service regions that persist throughout the uncertainty-aware solutions.
- The size of these service regions significantly affects the potential added value of moving to an uncertainty-aware model. In particular, the added value decreases with increasing separability into smaller service regions as the options of the mathematical program to redistribute customers to hedge against uncertainty decrease.

We also observe that in instances primarily composed of small service regions served by independent facilities, the worst-case realization is reached even for a low uncertainty budget. This suggests that the worst-case demand realization is primarily characterized by perturbation in a particular (small) service region. An intriguing pathway for future research is to identify this high-risk service region by backtracking the worst-case parameter realization of the uncertain parameter using duality theory and complementary slackness. This would allow for the identification of “high risk” demand profiles for a particular problem instance or types of instances classified according to the size of the underlying service regions.

8. Conclusion

The capacitated facility location problem is a more than well-studied problem in location literature. Nevertheless, a comprehensive understanding of what properties of the problem instance cause the extremely varying behavior regarding the persistence of individual facilities in well-performing solutions, the performance of different algorithms, and the effect of different modeling extensions has been missing. These relationships cannot be derived by looking purely at the input data. Meaningful insights can only be taken in combination with the mathematical program, i.e., the objective and the constraints. One way to implicitly consider this information is by using the information from one or more well-performing solutions to the particular instance. We summarize our main insights regarding the three components of our superordinate research question as follows.

Characteristic decision patterns and link to input data Rather than by a subset of favorable core facilities, well-performing solutions to a particular CFLP instance are characterized by an implicit division of the facility-customer space into service regions. Throughout these solutions, customers may be served from different facilities but are persistently served from facilities within their service region. When a service region is small in the sense that its customers can be served by a single facility optimally, this facility persists throughout well-performing solutions to that instance, appearing to be part of a set of favorable core facilities. Meanwhile, when the service regions are large in that their customers are served by more than one facility, these facilities interdepend on one another and are often optimal only in combination.

We show that service regions found in well-performing solutions to a static, deterministic instance also characterize optimal solutions to model extensions that explicitly include temporal developments or uncertainty in the demand parameters. They abstractly describe a sense of regional coherence induced by the spatial patterns of candidates and customers, the demand-capacity ratio, and the fixed-costs-profit ratio that results in allocation strategies adhering to these regions performing better than others. However, we show that the combination of overlaying effects in practically infinite variations of problem instances makes it difficult to derive service regions directly from the data. Rather, it is necessary to combine this with information from the objective and constraints. In particular, we show that service regions that characterize the optimal solution can be derived from the allocation matrices of sets of well-performing and even integer-infeasible solutions to that instance. We propose spectral biclustering as a pattern recognition technique to derive service regions on arbitrary sets of solutions.

Link to solution algorithms The size of the service regions determines the level of interdependence between facilities operating in the optimal solution. We show that the interdependence of facilities serving the same service region significantly affects the performance of both exact and heuristic solution algorithms. This can partially be attributed to the fact that, at this point, the implied service regions and, thus, the

implied relative position of individual facilities in the facility-customer space are largely ignored. We show that it is possible to determine service regions early on in the search process from infeasible solutions by means of pattern recognition. We demonstrate the potential use of specifically acknowledging these service regions by including them in the branch variable selection routine of CPLEX's branch-and-cut procedure.

Link to model extensions We examine the relationship between service regions and the added value of modeling extensions, in particular, extensions that consider temporal developments and uncertainty in the demand parameters. We draw on the value of the multi-period solution to quantify the benefit of the multi-period approach and present measures that allow quantifying and comparing the added value of a two-stage stochastic program and an adjustable robust counterpart presented with the same uncertainty setting. We conclude that to be of value, the extension must allow the reduction of relevant cost components in the given problem instance. Furthermore, the instance must bear sufficient flexibility in the well-performing decision to react to the added information. For multi-period models, we conclude that simply because decisions are taken in a temporal context, and relevant parameters vary over time, time is not necessarily an essential aspect to consider in the model. For uncertainty-aware problems, we observe that uncertainty-aware solutions yield different insights depending on the size of the underlying service regions. For instances with large service regions, uncertainty-aware models often choose different locations than their deterministic counterparts, which perform better in the face of uncertainty. In contrast, for instances with small service regions, the implied separation of the facility-customer space is very rigid. It restricts the degree to which the model may hedge against uncertainty, often leading to the same location decisions as the deterministic counterpart.

This work can be seen as a first, partially experimental step towards addressing the overarching goal of improving modeling and solution techniques through an improved understanding of the interplay between data, decisions, models, and solution algorithms. However, the focus was restricted to the profit-maximizing version of the capacitated facility location problem as a representative of discrete location problems. While it is reasonable to assume that insights into the implied regional patterns translate to other problem formulations, further experiments are necessary. On the one hand, it must be tested to which degree the presented insights hold for cost-minimizing formulations of the problem and other core location problems like the p-median problem or coverage location problems. On the other hand, it must be evaluated whether the present findings hold when the presented profit-maximizing formulation is extended to include, e.g., capacity expansion or inventory decisions.

The presented results yield ample opportunity for further research, and we provided detailed descriptions in the conclusions of the individual chapters. In particular, we see the following main avenues.

Improve solution procedures The presented algorithm to determine service regions in a given problem instance already from non-optimal or infeasible solutions allows access to information on the spatial relationships of facilities in the implied facility-customer space during the search process. This gives insights into which facilities are close to each other and compete to serve the same or overlapping subsets of customers. First attempts to exploit this information during the search process are promising, yet several further opportunities present themselves. For heuristics, we particularly see significant potential in including information on service regions during the creation of the buckets during kernel search. For exact branching procedures, we propose to combine the information on service regions with novel approaches to multi-variable branching that, up to this point, require external information on the regional structure.

While modern MIP solvers perform well for the CFLP, they stumble when tackling more complex problems like stochastic programs or adjustable robust counterparts. Notably, the decision patterns seen in static instances persist in these challenging scenarios. We see potential in exploring heuristic pattern-based separation procedures to solve such problems.

Data-driven modeling Just because problem parameters change over time does not make time a valuable aspect to be considered in the model. Similarly, a more granular model of time does not necessarily lead to better decisions, nor do periods of equal length ensure that relevant moments in time for decision-making are included. The distinction between effective, obsolete, and restrictive breakpoints provides valuable insights into the potential effect of a particular discretization of the planning horizon. Yet, a directive on identifying the most effective discretization to a particular instance purely from the data is still missing. Identifying characteristic developments of the demand parameters that indicate effective breakpoints is a promising avenue for further research and can contribute to deriving models from the data.

Classify discrete location instances The size of the service regions underlying different problem instances significantly impacts the performance of solution algorithms, the added value of modeling extensions, and the sensitivity of an optimal solution to minor perturbations. This work solely considers the average number of dependence relationships between facilities operating in the optimal solution as an indicator to group instances into distinct categories. However, further opportunities to classify problem instances regarding the number, size, and separability of their service regions present themselves. They potentially yield insights into suitable choices for model extensions or allow for the identification of high-risk perturbation profiles.

Discrete location problems are not the only class of problems in which certain subsets of binary decision variables are likely to interdepend more strongly than others. Identifying these problem classes and developing a standardized framework to detect regions of stronger and weaker coherence can be an interesting pathway to explore for mixed-integer programming.

Bibliography

R. K. Ahuja, J. B. Orlin, S. Pallottino, M. P. Scaparra, and M. G. Scutellà. A multi-exchange heuristic for the single-source capacitated facility location problem. *Management Science*, 50(6):749–760, 2004. doi: 10.1287/mnsc.1030.0193.

M. Albareda-Sambola, E. Fernández, Y. Hinojosa, and J. Puerto. The multi-period incremental service facility location problem. *Computers & Operations Research*, 36(5): 1356–1375, 2009. doi: 10.1016/j.cor.2008.02.010.

M. Albareda-Sambola, E. Fernández, and G. Laporte. A computational comparison of several models for the exact solution of the capacity and distance constrained plant location problem. *Computers & Operations Research*, 38(8):1109–1116, 2011a. doi: 10.1016/j.cor.2010.10.030.

M. Albareda-Sambola, E. Fernández, and F. Saldanha da Gama. The facility location problem with bernoulli demands. *Omega*, 39(3):335–345, 2011b. doi: 10.1016/j.omega.2010.08.002.

M. Albareda-Sambola, E. Fernández, and S. Nickel. Multiperiod location-routing with decoupled time scales. *European Journal of Operational Research*, 217(2):248–258, 2012. doi: 10.1016/j.ejor.2011.09.022.

S. A. Alumur, S. Nickel, and F. Saldanha da Gama. Hub location under uncertainty. *Transportation Research Part B: Methodological*, 46(4):529–543, 2012a. doi: 10.1016/j.trb.2011.11.006.

S. A. Alumur, S. Nickel, F. Saldanha da Gama, and V. Verter. Multi-period reverse logistics network design. *European Journal of Operational Research*, 220(1):67–78, 2012b. doi: 10.1016/j.ejor.2011.12.045.

E. Angelelli, R. Mansini, and M. G. Speranza. Kernel search: A general heuristic for the multi-dimensional knapsack problem. *Computers & Operations Research*, 37(11): 2017–2026, 2010. doi: 10.1016/j.cor.2010.02.002.

N. Aras, A. Korugan, G. Büyüközkan, F. S. Şerifoğlu, İ. Erol, and M. N. Velioğlu. Locating recycling facilities for it-based electronic waste in turkey. *Journal of Cleaner Production*, 105:324–336, 2015. doi: 10.1016/j.jclepro.2015.02.046.

A. Ardestani-Jaafari and E. Delage. The value of flexibility in robust location-transportation problems. *Transportation Science*, 52(1):189–209, 2018. doi: 10.1287/trsc.2016.0728.

P. Avella, M. Boccia, A. Sforza, and I. Vasil'ev. An effective heuristic for large-scale capacitated facility location problems. *Journal of Heuristics*, 15(6):597–615, 2009. doi: 10.1007/s10732-008-9078-y.

H. Badri, M. Bashiri, and T. H. Hejazi. Integrated strategic and tactical planning in a supply chain network design with a heuristic solution method. *Computers & Operations Research*, 40(4):1143–1154, 2013. doi: 10.1016/j.cor.2012.11.005.

H. Bakker and S. Nickel. The value of the multi-period solution revisited: When to model time in capacitated location problems. *Computers & Operations Research*, 161:106428, 2024. doi: 10.1016/j.cor.2023.106428.

H. Bakker, F. Dunke, and S. Nickel. A structuring review on multi-stage optimization under uncertainty: Aligning concepts from theory and practice. *Omega*, 96:102080, 2020. doi: 10.1016/j.omega.2019.06.006.

B. Balcik and B. M. Beamon. Facility location in humanitarian relief. *International Journal of Logistics Research and Applications*, 11(2):101–121, 2008. doi: 10.1080/13675560701561789.

J. Barcelo, E. Fernández, and K. O. Jörnsten. Computational results from a new lagrangean relaxation algorithm for the capacitated plant location problem. *European Journal of Operational Research*, 53(1):38–45, 1991. doi: 10.1016/0377-2217(91)90091-9.

O. Baron, J. Milner, and H. Naseraldin. Facility location: A robust optimization approach. *Production and Operations Management*, 20(5):772–785, 2010. doi: 10.1111/j.1937-5956.2010.01194.x.

J. E. Beasley. An algorithm for solving large capacitated warehouse location problems. *European Journal of Operational Research*, 33(3):314–325, 1988. doi: 10.1016/0377-2217(88)90175-0.

J. E. Beasley. Or-library: Distributing test problems by electronic mail. *Journal of the Operational Research Society*, 41(11):1069–1072, 1990. doi: 10.1057/jors.1990.166.

A. Ben-Tal and A. Nemirovski. Robust convex optimization. *Mathematics of Operations Research*, 23(4):769–805, 1998. doi: 10.1287/moor.23.4.769.

A. Ben-Tal, A. Goryashko, E. Guslitzer, and A. Nemirovski. Adjustable robust solutions of uncertain linear programs. *Mathematical Programming*, 99(2):351–376, 2004. doi: 10.1007/s10107-003-0454-y.

B. J. L. Berry and H. G. Barnum. Aggregate relations and elemental components of central place systems†. *Journal of Regional Science*, 4(1):35–68, 1962. doi: 10.1111/j.1467-9787.1962.tb00896.x.

D. Bertsimas and D. B. Brown. Constructing uncertainty sets for robust linear optimization. *Operations Research*, 57(6):1483–1495, 2009. doi: 10.1287/opre.1080.0646.

D. Bertsimas and M. Sim. The price of robustness. *Operations Research*, 52(1):35–53, 2004. doi: 10.1287/opre.1030.0065.

J. R. Birge and F. Louveaux. *Introduction to stochastic programming*. Springer series in operations research and financial engineering. Springer, New York, NY, 2. ed. edition, 2011. ISBN 978-1-4614-0236-7. doi: 10.1007/978-1-4614-0237-4.

C. M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer, 1 edition, 2007. ISBN 0387310738.

F. H. Boukani, B. F. Moghaddam, and M. S. Pishvaee. Robust optimization approach to capacitated single and multiple allocation hub location problems. *Computational and Applied Mathematics*, 35(1):45–60, Sept. 2014. doi: 10.1007/s40314-014-0179-y.

S. Busygin, O. Prokopyev, and P. M. Pardalos. Biclustering in data mining. *Computers & Operations Research*, 35(9):2964–2987, 2008. doi: 10.1016/j.cor.2007.01.005. Part Special Issue: Bio-inspired Methods in Combinatorial Optimization.

M. Caserta and S. Voß. A general corridor method-based approach for capacitated facility location. *International Journal of Production Research*, 58(13):3855–3880, 2020. doi: 10.1080/00207543.2019.1636320.

C.-J. T. Chia-Ho Chen. Combining lagrangian heuristic and ant colony system to solve the single source capacitated facility location problem. *Transportation Research Part E: Logistics and Transportation Review*, 44(6):1099–1122, 2008. doi: 10.1016/j.tre.2007.09.001.

P. J. Clark and F. C. Evans. Distance to nearest neighbor as a measure of spatial relationships in populations. *Ecology*, 35(4):445–453, 1954. doi: 10.2307/1931034.

I. A. Contreras and J. A. Díaz. Scatter search for the single source capacitated facility location problem. *Annals of Operations Research*, 157(1):73–89, 2007. doi: 10.1007/s10479-007-0193-1.

Á. Corberán, M. Landete, J. Peiró, and F. Saldanha da Gama. The facility location problem with capacity transfers. *Transportation Research Part E: Logistics and Transportation Review*, 138:101943, 2020. doi: 10.1016/j.tre.2020.101943.

G. Cornuejols, R. Sridharan, and J. M. Thizy. A comparison of heuristics and relaxations for the capacitated plant location problem. *European Journal of Operational Research*, 50(3):280–297, 1991. doi: 10.1016/0377-2217(91)90261-S.

I. Correia and T. Melo. Multi-period capacitated facility location under delayed demand satisfaction. *European Journal of Operational Research*, 255(3):729–746, 2016. doi: 10.1016/j.ejor.2016.06.039.

I. Correia and T. Melo. Integrated facility location and capacity planning under uncertainty. *Computational and Applied Mathematics*, 40(5), 2021. doi: 10.1007/s40314-021-01560-0.

I. Correia and F. Saldanha da Gama. Facility location under uncertainty. In *Location Science*, pages 185–213. Springer International Publishing, 2019. doi: 10.1007/978-3-030-32177-2_8.

I. Correia, S. Nickel, and F. Saldanha da Gama. A stochastic multi-period capacitated multiple allocation hub location problem: Formulation and inequalities. *Omega*, 74: 122–134, 2018. doi: 10.1016/j.omega.2017.01.011.

M. J. Cortinhal, M. J. Lopes, and M. T. Melo. Dynamic design and re-design of multi-echelon, multi-product logistics networks with outsourcing opportunities: A computational study. *Computers & Industrial Engineering*, 90:118–131, 2015. doi: 10.1016/j.cie.2015.08.019.

G. B. Dantzig. Linear programming under uncertainty. *Management Science*, 1(3/4): 197–206, 1955. doi: 10.2307/2627159.

M. W. Dawande and J. N. Hooker. Inference-based sensitivity analysis for mixed integer/linear programming. *Operations Research*, 48(4):623–634, 2000. doi: 10.1287/opre.48.4.623.12420.

N. M. M. de Abreu. Old and new results on algebraic connectivity of graphs. *Linear Algebra and its Applications*, 423(1):53–73, 2007. doi: 10.1016/j.laa.2006.08.017. Special Issue devoted to papers presented at the Aveiro Workshop on Graph Spectra.

E. Delage and Y. Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3):595–612, 2010. doi: 10.1287/opre.1090.0741.

H. Delmaire, J. A. Díaz, E. Fernández, and M. Ortega. Reactive grasp and tabu search based heuristics for the single source capacitated plant location problem. *INFOR: Information Systems and Operational Research*, 37(3):194–225, 1999. doi: 10.1080/03155986.1999.11732381.

I. S. Dhillon. Co-clustering documents and words using bipartite spectral graph partitioning. In *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '01, page 269–274, New York, NY, USA, 2001. Association for Computing Machinery. ISBN 158113391X. doi: 10.1145/502512.502550.

M. Di Francesco, M. Gaudioso, E. Gorgone, and I. Murthy. A new extended formulation with valid inequalities for the capacitated concentrator location problem. *European Journal of Operational Research*, 289(3):975–986, 2021. doi: 10.1016/j.ejor.2019.07.008.

J. A. Díaz and E. Fernández. A branch-and-price algorithm for the single source capacitated plant location problem. *Journal of the Operational Research Society*, 53(7): 728–740, 2002. doi: 10.1057/palgrave.jors.2601353.

Z. Drezner. Sensitivity analysis of the optimal location of a facility. *Naval Research Logistics Quarterly*, 32(2):209—224, 1985. doi: doi:10.1002/nav.3800320203.

Z. Drezner and H. W. Hamacher, editors. *Facility location: Applications and theory*. Springer, Berlin, 1. ed., 2. printing edition, 2004. ISBN 9783540213451.

F. Dunke, I. Heckmann, S. Nickel, and F. Saldanha da Gama. Time traps in supply chains: Is optimal still good enough? *European Journal of Operational Research*, 264 (3):813–829, 2018. doi: 10.1016/j.ejor.2016.07.016.

M. Eskandarpour, P. Dejax, and O. Péton. A large neighborhood search heuristic for supply chain network design. *Computers & Operations Research*, 80:23–37, 2017. doi: 10.1016/j.cor.2016.11.012.

E. Fadda, D. Manerba, G. Cabodi, P. E. Camurati, and R. Tadei. Comparative analysis of models and performance indicators for optimal service facility location. *Transportation Research Part E: Logistics and Transportation Review*, 145:102174, 2021. doi: 10.1016/j.tre.2020.102174.

X. Fang, J. Zhou, H. Zhao, and Y. Chen. A biclustering-based heterogeneous customer requirement determination method from customer participation in product development. *Annals of Operations Research*, 309:817–835, 2022. doi: 10.1007/s10479-020-03607-7.

E. Fernández and M. Landete. Fixed-charge facility location problems. In *Location Science*, pages 67–98. Springer International Publishing, 2019. doi: 10.1007/978-3-030-32177-2_4.

L. E. Ferro-Diez, N. M. Villegas, and J. Diaz-Cely. Location data analytics in the business value chain: A systematic literature review. *IEEE Access*, 8:204639–204659, 2020. doi: 10.1109/ACCESS.2020.3036835.

C. Filippi, G. Guastaroba, and M. G. Speranza. On single-source capacitated facility location with cost and fairness objectives. *European Journal of Operational Research*, 289(3):959–974, 2021. ISSN 03772217. doi: 10.1016/j.ejor.2019.07.045.

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. doi: 10.1016/j.ejor.2016.03.002.

M. Fischetti, I. Ljubić, and M. Sinnl. Redesigning benders decomposition for large-scale facility location. *Management Science*, 63(7):2146–2162, 2017. doi: 10.1287/mnsc.2016.2461.

N. Fraiman and Z. Li. Biclustering with alternating k-means. *ArXiv*, abs/2009.04550, 2020.

V. Gabrel, M. Lacroix, C. Murat, and N. Remli. Robust location transportation problems under uncertain demands. *Discrete Applied Mathematics*, 164:100–111, 2014. doi: 10.1016/j.dam.2011.09.015.

S. L. Gadegaard, A. Klose, and L. R. Nielsen. A bi-objective approach to discrete cost-bottleneck location problems. *Annals of Operations Research*, 267(1-2):179–201, 2018. doi: 10.1007/s10479-016-2360-8.

X. Geng and Y. Wang. A model for reverse logistics with collection sites based on heuristic algorithm. In W. E. Wong and T. Zhu, editors, *Computer Engineering and Networking*, pages 395–402, Cham, 2014. Springer International Publishing. ISBN 978-3-319-01766-2.

A. M. Geoffrion and R. Nauss. Exceptional paper—parametric and postoptimality analysis in integer linear programming. *Management Science*, 23(5):453–466, 1977. doi: 10.1287/mnsc.23.5.453.

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. doi: 10.1287/ijoc.1110.0468.

F. Grenouilleau, A. Legrain, N. Lahrichi, and L.-M. Rousseau. A set partitioning heuristic for the home health care routing and scheduling problem. *European Journal of Operational Research*, 275(1):295–303, 2019. doi: 10.1016/j.ejor.2018.11.025.

G. Guastaroba and M. G. Speranza. Kernel search for the capacitated facility location problem. *Journal of Heuristics*, 18(6):877–917, 2012. doi: 10.1007/s10732-012-9212-8.

G. Guastaroba and M. G. Speranza. A heuristic for bilp problems: The single source capacitated facility location problem. *European Journal of Operational Research*, 238 (2):438–450, 2014. doi: 10.1016/j.ejor.2014.04.007.

C. S. Guazzelli and C. B. Cunha. Exploring k- best solutions to enrich network design decision-making. *Omega*, 78:139–164, 2018. doi: 10.1016/j.omega.2017.06.009.

N. Gülpınar, D. Pachamanova, and E. Çanakoğlu. Robust strategies for facility location under uncertainty. *European Journal of Operational Research*, 225(1):21–35, 2013. doi: 10.1016/j.ejor.2012.08.004.

R. Hamming. *Coding and information theory*. Englewood Cliffs N.J: Prentice-Hall, 1980.

R. Henriques, C. Antunes, and S. C. Madeira. A structured view on pattern mining-based biclustering. *Pattern Recognition*, 48(12):3941–3958, 2015. ISSN 0031-3203. doi: 10.1016/j.patcog.2015.06.018.

P. Hertz. Über den gegenseitigen durchschnittlichen abstand von punkten, die mit bekannter mittlerer dichte im raume angeordnet sind. *Mathematische Annalen*, 67(3):1432–1807, 1909. doi: 10.1007/BF01450410.

F. S. Hillier, T. Gal, and H. J. Greenberg. *Advances in Sensitivity Analysis and Parametric Programming*, volume 6. Springer US, Boston, MA, 1997. doi: 10.1007/978-1-4615-6103-3.

S. Hochreiter, U. Bodenhofer, M. Heusel, A. Mayr, A. Mitterecker, A. Kasim, T. Khamiakova, S. V. Sanden, D. Lin, W. Talloen, L. Bijnens, H. W. H. Göhlmann, Z. Shkedy, and D.-A. Clevert. FABIA: factor analysis for bicluster acquisition. *Bioinformatics*, 26 (12):1520–1527, 2010. doi: 10.1093/bioinformatics/btq227.

K. Holmberg, M. Rönnqvist, and D. Yuan. An exact algorithm for the capacitated facility location problems with single sourcing. *European Journal of Operational Research*, 113 (3):544–559, 1999. doi: 10.1016/S0377-2217(98)00008-3.

S. Hu, C. Han, Z. S. Dong, and L. Meng. A multi-stage stochastic programming model for relief distribution considering the state of road network. *Transportation Research Part B: Methodological*, 123:64–87, 2019. doi: 10.1016/j.trb.2019.03.014.

IBM Corporation. Mip variable selection strategy, 2021a. URL `https://www.ibm.com/docs/de/icos/12.10.0?topic=parameters-mip-variable-selection-strategy`.

IBM Corporation. *IBM ILOG CPLEX Optimization Studio V12.10.0 documentation*. IBM, 2021b.

S. K. Jacobsen. Heuristics for the capacitated plant location model. *European Journal of Operational Research*, 12(3):253–261, 1983. doi: 10.1016/0377-2217(83)90195-9.

V. Jayaraman, R. A. Patterson, and E. Rolland. The design of reverse distribution networks: Models and solution procedures. *European Journal of Operational Research*, 150 (1):128–149, 2003. doi: 10.1016/S0377-2217(02)00497-6.

S. Jena, J.-F. Cordeau, and B. Gendron. Dynamic facility location with generalized modular capacities. *Transportation Science*, 49(3):484–499, 2015. doi: 10.1287/trsc.2014.0575.

H. Jia, F. Ordóñez, and M. Dessouky. A modeling framework for facility location of medical services for large-scale emergencies. *IIE Transactions*, 39(1):41–55, 2007. doi: 10.1080/07408170500539113.

D. Jones, S. Firouzy, A. Labib, and A. V. Argyriou. Multiple criteria model for allocating new medical robotic devices to treatment centres. *European Journal of Operational Research*, 297(2):652–664, 2022. doi: 10.1016/j.ejor.2021.06.003.

A. José-García, J. Jacques, V. Sobanski, and C. Dhaenens. *Biclustering Algorithms Based on Metaheuristics: A Review*, page 39–71. Springer Nature Singapore, 2022. ISBN 9789811938887. doi: 10.1007/978-981-19-3888-7_2.

M. Kchaou Boujelben, C. Gicquel, and M. Minoux. A milp model and heuristic approach for facility location under multiple operational constraints. *Computers & Industrial Engineering*, 98:446–461, 2016. doi: 10.1016/j.cie.2016.06.022.

F. Kılcı, B. Y. Kara, and B. Bozkaya. Locating temporary shelter areas after an earthquake: A case for turkey. *European Journal of Operational Research*, 243(1):323–332, 2015. ISSN 03772217. doi: 10.1016/j.ejor.2014.11.035.

A. J. King and S. W. Wallace. *Modeling with Stochastic Programming*. Springer New York, New York, NY, 2012. doi: 10.1007/978-0-387-87817-1.

A. Klose and A. Drexl. Facility location models for distribution system design. *European Journal of Operational Research*, 162(1):4–29, 2005. doi: 10.1016/j.ejor.2003.10.031.

A. Klose and S. Görtz. A branch-and-price algorithm for the capacitated facility location problem. *European Journal of Operational Research*, 179(3):1109–1125, 2007. doi: 10.1016/j.ejor.2005.03.078.

A. A. Kuehn and M. J. Hamburger. A heuristic program for locating warehouses. *Management Science*, 9(4):643–666, 1963.

M. Labbé, J.-F. Thisse, and R. E. Wendell. Sensitivity analysis in minisum facility location problems. *Operations Research*, 39(6):961–969, 1991.

G. Laporte, F. V. Louveaux, and L. van Hamme. Exact solution to a location problem with stochastic demands. *Transportation Science*, 28(2):95–103, 1994. doi: 10.1287/trsc.28.2.95.

J. M. Y. Leung and T. L. Magnanti. Valid inequalities and facets of the capacitated plant location problem. *Mathematical Programming*, 44(1-3):271–291, 1989. doi: 10.1007/BF01587093.

C. Lin. Stochastic single-source capacitated facility location model with service level requirements. *International Journal of Production Economics*, 117(2):439–451, 2009. doi: 10.1016/j.ijpe.2008.11.009.

T. Liu, F. Saldanha da Gama, S. Wang, and Y. Mao. Robust stochastic facility location: Sensitivity analysis and exact solution. *INFORMS Journal on Computing*, 34(5):2776–2803, 2022. doi: 10.1287/ijoc.2022.1206.

Y. Liu, Z. Li, H. Xiong, X. Gao, and J. Wu. Understanding of internal clustering validation measures. In *2010 IEEE International Conference on Data Mining*, pages 911–916, 2010. doi: 10.1109/ICDM.2010.35.

A. Lodi, L. Mossina, and E. Rachelson. Learning to handle parameter perturbations in combinatorial optimization: An application to facility location. *EURO Journal on Transportation and Logistics*, 9(4):100023, 2020. doi: 10.1016/j.ejtl.2020.100023.

N. Loree and F. Aros-Vera. Points of distribution location and inventory management model for post-disaster humanitarian logistics. *Transportation Research Part E: Logistics and Transportation Review*, 116:1–24, 2018. doi: 10.1016/j.tre.2018.05.003.

S. Madeira and A. Oliveira. Biclustering algorithms for biological data analysis: a survey. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 1(1):24–45, 2004. doi: 10.1109/TCBB.2004.2.

F. Mai, M. J. Fry, and J. W. Ohlmann. Model-based capacitated clustering with posterior regularization. *European Journal of Operational Research*, 271(2):594–605, 2018. doi: 10.1016/j.ejor.2018.04.048.

V. Marianov, M. Mizumori, and C. ReVelle. The heuristic concentration-integer and its application to a class of location problems. *Computers & Operations Research*, 36(5): 1406–1422, 2009. doi: 10.1016/j.cor.2008.02.011.

A. Marín, L. I. Martínez-Merino, A. M. Rodríguez-Chía, and F. Saldanha da Gama. Multi-period stochastic covering location problems: Modeling framework and solution approach. *European Journal of Operational Research*, 268(2):432–449, 2018. doi: 10.1016/j.ejor.2018.01.040.

N. Marković, I. O. Ryzhov, and P. Schonfeld. Evasive flow capture: A multi-period stochastic facility location problem with independent demand. *European Journal of Operational Research*, 257(2):687–703, 2017. doi: 10.1016/j.ejor.2016.08.020.

T. Matos, Ó. Oliveira, and D. Gamboa. A simple dual-ramp algorithm for the capacitated facility location problem. *Matsatsinis N., Marinakis Y., Pardalos P. (eds) Learning and Intelligent Optimization. LION 2019. Lecture Notes in Computer Science, vol 11968. Springer, Cham.*, 11968:240–252, 2020.

M. T. Melo, S. Nickel, and F. Saldanha da Gama. Dynamic multi-commodity capacitated facility location: a mathematical modeling framework for strategic supply chain planning. *Computers & Operations Research*, 33(1):181–208, 2006. doi: 10.1016/j.cor.2004.07.005.

M. T. Melo, S. Nickel, and F. Saldanha da Gama. Facility location and supply chain management – a review. *European Journal of Operational Research*, 196(2):401–412, 2009. doi: 10.1016/j.ejor.2008.05.007.

J. E. Mendoza, L.-M. Rousseau, and J. G. Villegas. A hybrid metaheuristic for the vehicle routing problem with stochastic demand and duration constraints. *Journal of Heuristics*, 22(4):539–566, 2016. doi: 10.1007/s10732-015-9281-6.

D. R. Morrison, S. H. Jacobson, J. J. Sauppe, and E. C. Sewell. Branch-and-bound algorithms: A survey of recent advances in searching, branching, and pruning. *Discrete Optimization*, 19:79–102, 2016. doi: 10.1016/j.disopt.2016.01.005.

W. Mu and D. Tong. A spatial–knowledge–enhanced heuristic for solving the p –median problem. *Transactions in GIS*, 22(2):477–493, 2018. doi: 10.1111/tgis.12322.

G. F. Mulligan. The effects of multiplier shifts in a hierarchical city-size model. *Regional Science and Urban Economics*, 10(1):77–90, 1980. doi: 10.1016/0166-0462(80)90049-6.

S. Nickel and F. Saldanha da Gama. Multi-period facility location. In G. Laporte, S. Nickel, and F. Saldanha da Gama, editors, *Location Science*, pages 303–326. Springer International Publishing, Cham, 2019. ISBN 978-3-030-32177-2. doi: 10.1007/978-3-030-32177-2_11.

S. Nickel, F. Saldanha da Gama, and H.-P. Ziegler. A multi-stage stochastic supply network design problem with financial decisions and risk management. *Omega*, 40(5):511–524, 2012. doi: 10.1016/j.omega.2011.09.006.

V. A. Padilha and R. J. G. B. Campello. A systematic comparative evaluation of biclustering techniques. *BMC Bioinformatics*, 18(1):55, 2017. doi: 10.1186/s12859-017-1487-1.

A. Pagès-Bernaus, H. Ramalhinho, A. A. Juan, and L. Calvet. Designing e-commerce supply chains: a stochastic facility-location approach. *International Transactions in Operational Research*, 26(2):507–528, 2019. doi: 10.1111/itor.12433.

S. B. Park. Performance of successively complex rules for locational decision-making. *Annals of Operations Research*, 18(1):323–343, 1989. doi: 10.1007/BF02097811.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

M. Peker, B. Y. Kara, J. F. Campbell, and S. A. Alumur. Spatial analysis of single allocation hub location problems. *Networks and Spatial Economics*, 16(4):1075–1101, 2016. doi: 10.1007/s11067-015-9311-9.

B. Pelegrín, P. Fernández, M. Dolores García Pérez, and S. Cano Hernández. On the location of new facilities for chain expansion under delivered pricing. *Omega*, 40(2): 149–158, 2012. doi: 10.1016/j.omega.2011.04.005.

B. Pontes, R. Girldez, and J. S. Aguilar-Ruiz. Quality measures for gene expression biclusters. *PLOS ONE*, 10(3):1–24, 2015a. doi: 10.1371/journal.pone.0115497.

B. Pontes, R. Giráldez, and J. S. Aguilar-Ruiz. Biclustering on expression data: A review. *Journal of Biomedical Informatics*, 57:163–180, 2015b. doi: 10.1016/j.jbi.2015.06.028.

C. S. ReVelle, H. A. Eiselt, and M. S. Daskin. A bibliography for some fundamental problem categories in discrete location science. *European Journal of Operational Research*, 184(3):817–848, 2008. doi: 10.1016/j.ejor.2006.12.044.

R. T. Rockafellar and R. J.-B. Wets. Stochastic convex programming: Relatively complete recourse and induced feasibility. *SIAM Journal on Control and Optimization*, 14(3):574–589, 1976. doi: 10.1137/0314038.

L. Rokach and O. Maimon. Clustering methods. In *Data Mining and Knowledge Discovery Handbook*, pages 321–352. Springer-Verlag, 2005. doi: 10.1007/0-387-25465-x_15.

F. Role, S. Morbieu, and M. Nadif. Coclust: A python package for co-clustering. *Journal of Statistical Software*, 88(7):1–29, 2019. doi: 10.18637/jss.v088.i07.

G. M. Roodman and L. B. Schwarz. Extensions of the multi-period facility phase-out model: New procedures and application to a phase-in/phase-out problem. *A I I E Transactions*, 9(1):103–107, 1977. doi: 10.1080/05695557708975128.

J. Rosenhead, M. Elton, and S. K. Gupta. Robustness and optimality as criteria for strategic decisions. *Operational Research Quarterly (1970-1977)*, 23(4):413, 1972. doi: 10.2307/3007957.

K. Rosing and C. ReVelle. Heuristic concentration: Two stage solution construction. *European Journal of Operational Research*, 97(1):75–86, 1997. doi: 10.1016/S0377-2217(96)00100-2.

K. Rosing, C. ReVelle, and D. Schilling. A gamma heuristic for the p-median problem. *European Journal of Operational Research*, 117(3):522–532, 1999. doi: 10.1016/S0377-2217(98)00268-9.

A. Saif and E. Delage. Data-driven distributionally robust capacitated facility location problem. *European Journal of Operational Research*, 291(3):995–1007, 2021. doi: 10.1016/j.ejor.2020.09.026.

E. Sanci and M. S. Daskin. Integrating location and network restoration decisions in relief networks under uncertainty. *European Journal of Operational Research*, 279(2):335–350, 2019. doi: 10.1016/j.ejor.2019.06.012.

H. Scarf. *A Min-Max Solution of an Inventory Problem*, page 19–27. Palgrave Macmillan UK, 2005. ISBN 9781137024381. doi: 10.1057/9781137024381_3.

M. Shahabi and A. Unnikrishnan. Robust hub network design problem. *Transportation Research Part E: Logistics and Transportation Review*, 70:356–373, 2014. doi: 10.1016/j.tre.2014.08.003.

L. V. Snyder. Facility location under uncertainty: A review. *IIE Transactions*, 38(7):547–564, 2006. doi: 10.1080/07408170500216480.

A. L. Soyster. Constraints and applications to inexact linear programming. *Operations Research*, 21(5):1154–1157, 1973. doi: 10.1287/opre.21.5.1154.

T. Stidsen, K. A. Andersen, and B. Dammann. A branch and bound algorithm for a class of biobjective mixed integer programs. *Management Science*, 60(4):1009–1032, 2014. doi: 10.1287/mnsc.2013.1802.

I. Tari and S. A. Alumur. Collection center location with equity considerations in reverse logistics networks. *INFOR: Information Systems and Operational Research*, 52(4):157–173, 2014. doi: 10.3138/infor.52.4.157.

V. Verter and M. C. Dincer. Facility location and capacity acquisition: An integrated approach. *Naval Research Logistics*, 42(8):1141–1160, 1995. doi: 10.1002/1520-6750(199512)42:8{\textless}1141::AID-NAV3220420803{\textgreater}3.0.CO;2-B.

M. Vijaymeena and K. Kavitha. A survey on similarity measures in text mining. *Machine Learning and Applications: An International Journal*, 3(1):19–28, 2016. doi: 10.5121/mlaij.2016.3103.

U. von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17:395–416, 2007. doi: 10.1007/s11222-007-9033-z.

S. Wang, Z. Chen, and T. Liu. Distributionally robust hub location. *Transportation Science*, 54(5):1189–1210, 2020. doi: 10.1287/trsc.2019.0948.

W. Wiesemann, D. Kuhn, and M. Sim. Distributionally robust convex optimization. *Operations Research*, 62(6):1358–1376, 2014. doi: 10.1287/opre.2014.1314.

Z. Yang, F. Chu, and H. Chen. A cut-and-solve based algorithm for the single-source capacitated facility location problem. *European Journal of Operational Research*, 221 (3):521–532, 2012. doi: 10.1016/j.ejor.2012.03.047.

B. Yildiz, N. Boland, and M. Savelsbergh. Decomposition branching for mixed integer programming. *Operations Research*, 70(3):1854–1872, 2022. doi: https://doi.org/10.1287/opre.2021.2210.

S. Yu. Multiclass spectral clustering. In *Proceedings Ninth IEEE International Conference on Computer Vision*, pages 313–319 vol.1, 2003. doi: 10.1109/ICCV.2003.1238361.

B. Zeng and L. Zhao. Solving two-stage robust optimization problems using a column-and-constraint generation method. *Operations Research Letters*, 41(5):457–461, 2013. doi: 10.1016/j.orl.2013.05.003.

C. A. Zetina, I. Contreras, J.-F. Cordeau, and E. Nikbakhsh. Robust uncapacitated hub location. *Transportation Research Part B: Methodological*, 106:393–410, 2017. doi: 10.1016/j.trb.2017.06.008.

Y. Zhang, F. Chu, A. Che, Y. Yu, and X. Feng. Novel model and kernel search heuristic for multi-period closed-loop food supply chain planning with returnable transport items. *International Journal of Production Research*, 57(23):7439–7456, 2019. doi: 10.1080/00207543.2019.1615650.

H. Zhao, A. W.-C. Liew, D. Z. Wang, and H. Yan. Biclustering analysis for pattern discovery: Current techniques, comparative studies and applications. *Current Bioinformatics*, 7(1):43–55, 2012. doi: 10.2174/157489312799304413.

N. Zhu, C. Fu, and S. Ma. Data-driven distributionally robust optimization approach for reliable travel-time-information-gain-oriented traffic sensor location model. *Transportation Research Part B: Methodological*, 113:91–120, 2018. doi: 10.1016/j.trb.2018.05.009.

Appendices

A. Expected kernel persistence

In a Laplace experiment the probability of an event equals the number of favorable events divided by the total number of events. In the following we assume that $\tilde{\mathcal{S}}$ is the result of $|\tilde{\mathcal{S}}| = \sigma$ random drawings of $\overline{n}$ out of $|I|$ facilities, whereby $\overline{n}$ is the ceiled average number of facilities operating in the solutions in the set $\mathcal{S}$ such that

$$\overline{n} := \left\lceil \frac{\sum_{s\in\mathcal{S}} \sum_{i\in I} y_i^s}{|\mathcal{S}|} \right\rceil .$$

In Equation 2.6 we saw that the kernel persistence can be rewritten as the average number of facilities operating across all solutions divided by the set of relevant candidates. Consequently, if we approximate that average by $\overline{n}$ and consider the set of relevant candidates, $|I^R(\tilde{\mathcal{S}})|$ a discrete random variable, we can determine the expected value of $KER(\tilde{\mathcal{S}})$ via

$$\mathbb{E}\left(KER(\tilde{\mathcal{S}})\right) = \overline{n} \cdot \left(\mathbb{E}\left(|I^R(\tilde{\mathcal{S}})|\right)\right)^{-1} = \overline{n} \cdot \left(\sum\nolimits_{r=0}^{|I|} r \cdot P\left(|I^R(\tilde{\mathcal{S}})| = r\right)\right)^{-1}. \tag{A.1}$$

Thereby, $P\left(|I^R(\tilde{\mathcal{S}})| = r\right)$ denotes the probability that the cardinality of $I^R(\tilde{\mathcal{S}})$ equals r.

Thereby, $I^R(\tilde{\mathcal{S}}) = \left\{i \in I \middle| \sum_{s\in\tilde{\mathcal{S}}} y_i^s > 0\right\}$ is the result of two nested Laplace experiments. In the first experiment we draw $\overline{n}$ out of $|I|$ candidates in order to produce a single solution s. The set of possible outcomes has the size $\binom{|I|}{\overline{n}}$. Then, in a second drawing, we draw σ solutions out of these outcomes at random. Depending on whether a single solution may occur more than once or not the number of possible outcomes of this second experiment is either

$$\binom{\binom{|I|}{\overline{n}} + \sigma - 1}{\sigma} \tag{A.2}$$

combinations with repetitions, or

$$\binom{\binom{|I|}{\overline{n}}}{\sigma} \tag{A.3}$$

combinations without repetitions.

How do we derive the number of favorable events, that is, the number of events in which we obtain a set $\tilde{\mathcal{S}}$ in which exactly r distinct facilities operate across all solutions? For a defined subset of candidates L with cardinality $|L| = r$ we can derive the subset of events in which only facilities of L occur in $\tilde{\mathcal{S}}$ by

$$\binom{\binom{r}{\overline{n}} + \sigma - 1}{\sigma}. \tag{A.4}$$

However, if $\overline{n} < r$ then (A.4) still contains those events in which $\tilde{\mathcal{S}}$ contains some but not all of the candidates in L and consequently $|I^R(\tilde{\mathcal{S}})| < r$. Therefore, (A.4) counts the number of events in which $|I^R(\tilde{\mathcal{S}})| \leq r$.

Staying with the fixed subset L, to determine the number of events in which $I^R(\tilde{\mathcal{S}}) = L$

and not $I^R(\tilde{\mathcal{S}}) \subset L$ we must subtract all events in which $|I^R(\tilde{\mathcal{S}})| \leq r-l$ for $l \in 1, \ldots, r-\overline{n}$. Therefore, it is sufficient to subtract all events in which $|I^R(\tilde{\mathcal{S}})| \leq r-1$. So, for example, if $r = 5$, then we must subtract from (A.4) all events in which only 4 different candidates operate across all facilities. For a fixed subset of $r-1$ out of these r facilities in L this is easy and can be determined according to (A.4) straightforwardly. However, we need to subtract these events for all possible subsets of $r-1$ out of r facilities. The events that only facilities in different subsets of L of cardinality $r-1$ operate in any solution are not mutually exclusive. In particular, as we subtract the probability that "at most" $r-1$ out of $r-2$ facilities are operating, different subsets of cardinality $r-1$ have a significant overlap of subsets if cardinality $r-2$, $r-3$, etc. Therefore, when considering the events in which $I^R(\tilde{\mathcal{S}}) \subset L$ with $|I^R| = |L|-1$ we are looking at $\binom{|L|}{|L|-1}$ non-mutually exclusive events. The probability of the union of non-mutually exclusive events E_k, $k \in K$ is known to be

$$P\left(\bigcup_{k\in K} E_k\right) = \sum_{m=1}^{|K|} \left\{ (-1)^{m-1} \sum_{\substack{M\subseteq K,\\ |M|=m}} P(E_M) \right\}, \tag{A.5}$$

with

$$P(E_M) = P\left(\bigcap\nolimits_{k\in M} E_k\right). \tag{A.6}$$

Consequently, when $\tilde{\mathcal{S}}$ is a multiset and individual solutions may occur more than once, then we have

$$P(|I^R(\tilde{\mathcal{S}})| = r) = \frac{\binom{|I|}{r}\left\{\sum_{l=0}^{r}(-1)^l \binom{\binom{r-l}{\overline{n}}+\sigma-1}{\sigma}\binom{r}{r-l}\right\}}{\binom{\binom{|I|}{\overline{n}}+\sigma-1}{\sigma}}. \tag{A.7}$$

Otherwise, we have

$$P(|I^R(\tilde{\mathcal{S}})| = r) = \frac{\binom{|I|}{r}\left\{\sum_{l=0}^{r}(-1)^l \binom{\binom{r-l}{\overline{n}}}{\sigma}\binom{r}{r-l}\right\}}{\binom{\binom{|I|}{\overline{n}}}{\sigma}}. \tag{A.8}$$

Together with (A.1) the expectation of $\tilde{\mathcal{S}}$ can be obtained straightforwardly.

Unfortunately, the closed from expressions of $P(|I^R(\tilde{\mathcal{S}})| = r)$ in (A.7) and (A.8) are computationally inefficient. In particular, with increasing $|I|$, the nested coefficients $\binom{\binom{r-l}{\overline{n}}+\sigma-1}{\sigma}\binom{r}{r-l}$ and $\binom{\binom{r-l}{\overline{n}}}{\sigma}$ quickly exceed limits for infinity for most data types. For example, already for $|I| = 100$, $\overline{n} = 45$ and $\sigma = 4$, when we want to determine $P(|I^R(\tilde{\mathcal{S}})| = 90)$, $\binom{\binom{90-0}{45}}{4}$ for $l = 0$ significantly exceeds 10^{100}.

Therefore, in these cases, we approximate $\mathbb{E}\left(KER\left(\tilde{\mathcal{S}}\right)\right)$ by simple Monte Carlo simulation. We generate K sets of solutions $\tilde{\mathcal{S}}^K$ of cardinality σ composed of solutions which represent random samplings of $\overline{n}$ out of $|I|$ candidates. We subsequently count the number of unique candidates $|I^R(\tilde{\mathcal{S}})|^k$ in each set (or multiset) $\tilde{\mathcal{S}}^k$ and approximate $\mathbb{E}\left(KER\left(\tilde{\mathcal{S}}\right)\right)$ by the average of the observed values of $|I^R(\tilde{\mathcal{S}})|^k$. As Table A.1 displays, for $K = 1000$

we are able to approximate $\mathbb{E}\left(KER\left(\tilde{\mathcal{S}}\right)\right)$ with a precision of 2 digits for problems of significant size.

			K					
$\lvert I\rvert$	σ	$\overline{n}$	10	100	1000	10000	100000	exact
10	5	5	0.5050	0.5198	0.5151	0.5153	0.5155	0.5168
10	10	5	0.5000	0.5010	0.5003	0.5005	0.5004	0.5006
50	10	10	0.2283	0.2254	0.2240	0.2240	0.2240	0.2241
100	10	20	0.2250	0.2240	0.2239	0.2240	0.2241	0.2241

Table A.1.: Monte Carlo estimation for expected kernel persistence for different K