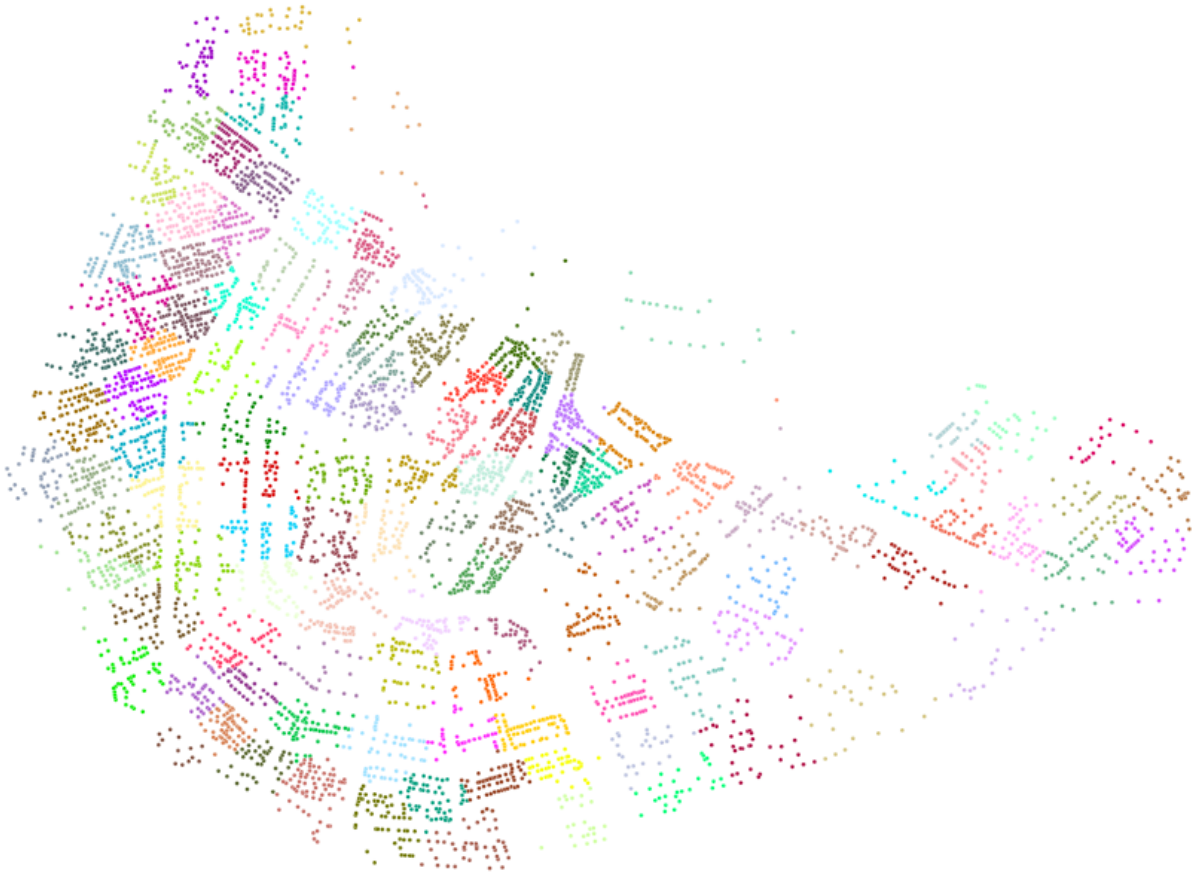
Robust Clustering Methods for 5GDHC Networks in Densely Populated Urban Areas

A Comparative Assessment for Different Participation Scenarios

Jana Teresa Riederer



TU Delft

Robust Clustering Methods for 5GDHC Networks in Densely Populated Urban Areas

A Comparative Assessment for Different Participation Scenarios

Thesis report

by

Jana Teresa Riederer

to obtain the degree of Master of Science
at the Delft University of Technology
to be defended publicly on May 12th, 2025

Thesis committee:

Chair:	Dr. ir. P.W. Heijnen
Supervisors:	Dr. ir. P.W. Heijnen
	Prof. dr. M.E. Warnier
External Supervisor:	Msc. P. Voskuilen (AMS Institute)
Place:	Faculty of Technology, Policy and Management, Delft
Project Duration:	October 2024 - April 2025
Student number:	5832888

An electronic version of this thesis is available at `http://repository.tudelft.nl/`.

Faculty of Technology, Policy and Management · Delft University of Technology

Preface

With this project, I aim to contribute to the implementation of fifth generation district heating and cooling networks in densely populated urban areas. Additionally, I hope to have delivered some insights about the behaviour of clustering methods on different spatial data sets and their comparison for similar applications. I am very grateful for the opportunity to explore this topic in a realistic scenario in the city centre of Amsterdam. Throughout my Master studies, district heating and cooling networks and their potential in the Energy Transition have fascinated me, which is why I really enjoyed this project and the learning opportunities that came with it. I hope to apply and expand the knowledge I acquired in my professional career, while contributing to the Energy Transition.

This project would have not been possible without the continuous support I received from my supervisors. First, I would like to thank Paul Voskuilen from AMS for introducing this topic, providing the relevant data and the inspiring feedback that helped me discover new angles of the topic. I would like to thank Martijn Warnier for his valuable input and encouraging feedback throughout this project. Additionally, I would like to especially thank Petra Heijnen for her crucial support and guidance through many meetings, new ideas and in every step of this project. Finally, I would like to thank my family and friends for their support throughout my academic career. My parents, who have always been there for me and supported me even when I decided to move to the Netherlands. My brother Thomas, who semi-voluntarily proofread earlier versions of this work and is always there for me. And lastly, to Yann, who always supports me.

Jana Teresa Riederer
Delft, April 2025

Executive Summary

The Dutch heating sector largely depends on natural gas. To meet their climate goals, the Netherlands has committed to phasing out natural gas by 2050. In light of recent energy security concerns regarding gas imports, this ambition has become more urgent. Phasing out natural gas requires new heating strategies, especially for the residential sector. Fifth generation district heating and cooling (5GDHC) networks are a promising alternative to natural gas, since they enable the utilization of low temperature heat sources and the balancing of heating and cooling demands between buildings. One important prerequisite for those novel heating and cooling network is a high heat demand density. Therefore, these networks are an attractive solution for densely populated urban areas. However, sufficiently insulated buildings and new infrastructure such as pipes and heat pumps are necessary for the implementation of 5GDHC networks. Therefore, these networks are predominantly implemented in newly build areas. Key challenges for the application of 5GDHC networks in existing densely populated urban areas are the spacial limitations and the proper insulation of old buildings. These challenges occur especially for old Dutch city centres, such as the city centre of Amsterdam.

However, the old building stock needs to be retrofitted to use the network. To implement a 5GDHC network, a modular approach is suitable, where the city is divided into smaller networks that are implemented step by step and later on connected to optimize the overall operation of the networks. This approach ensures a faster implementation and lower initial investment. Additionally, this approach also enables a partially simultaneous implementation of the networks and the retrofitting measures. While networks can be implemented where buildings are already retrofitted, other parts of the cities can be retrofitted at the same time. Therefore, the city centre of Amsterdam needs to be divided into clusters.

Clustering methods are commonly used to group various types of data based on similarity measures. However, there is a knowledge gap regarding how to select the most suitable approach for specific datasets. Furthermore, uncertainty can be an implementation barrier for a district heating and cooling network in densely populated urban areas. Since people might prefer alternatives and there is no obligation to participate in such a network, it is highly uncertain how many building owners will participate. A partition of the city centre that maintains a good performance under participation uncertainty could help overcome this implementation barrier. Therefore, this work aims at identifying the most robust method to cluster buildings for a fifth generation district heating and cooling network facing participation uncertainty.

To be successfully implemented, a fifth generation district heating and cooling network needs to meet the requirements of the involved stakeholders. These requirements imply conditions for a suitable partition and clustering method. A good cluster should be compact to enable short pipes to keep the required space small and investment costs low. Additionally, the cluster should align with the street structure to ensure feasible implementation of infrastructure. The cluster should also be small for a quick implementation of the network. To ensure suitable heating and cooling supply, a good cluster should also have an adequate energy balance. These requirements for a well performing cluster are translated into optimization metrics. Here, the average distance between two buildings in a cluster is used to quantify the compactness of the cluster. Additional optimization metrics are the cluster size, the Block Completeness Coefficient, which captures the alignment with the street structure, and the Demand Fulfilment Coefficient, which indicates the Energy Balance within a cluster. A suitable partition should have a close to optimal average value of those metrics and a low variance across clusters. Besides a good performance measured by the optimization metrics, a suitable clustering method should also be easy to use and easy to explain.

In previous work, Single Linkage Clustering (SLC) and K-Means have been identified as suitable algorithms to cluster buildings for fifth generation district heating and cooling networks in densely populated

urban areas. K-Medians could be a promising alternative to K-Means due to its robustness against outliers. Additionally, Density-Based Spatial Clustering of Applications with noise (DBSCAN) is considered as an alternative, since it may be able to follow the street structure. For all considered algorithms, two post-clustering modification steps are defined to improve the performance measured by the optimization metrics, namely the size and energy optimization step.

Within this work, a methodology to compare the robustness of clustering methods for low temperature district heating and cooling networks is developed. The methodology is a scenario-based approach. Different participation levels, 100%, 95%, 80%, 70% and 50% and a scenario where every building except utility buildings participate are considered. First, the minimax regret method is applied using the Block Completeness Coefficient, the Demand Fulfilment Coefficient and the average distance between two buildings in a cluster as performance indicators. Then the viable clusters for each method across the different scenarios are analysed. Finally, it is investigated for each method if the buildings had been clustered differently if the participation rate was known before by assessing the partition overlap with the Adjusted Rand Index.

With the developed methodology, the K-Means algorithm is identified as the most robust and best performing among the four considered methods for clustering buildings in densely populated areas for 5GDHC networks. The identified viable K-Means clusters are also the most robust in different scenarios and are promising starting points for the implementation of a fifth generation district heating and cooling network in the centre of Amsterdam. However, the K-Medians and SLC also show a robust performance and might be more suitable for other datasets or conditions such as increased importance of the alignment with the street network due to obstacles. The assessment also shows the impact of the participation rate on the viability of cluster networks.

Contents

Nomenclature

List of Abbreviations

5GDHC	Fifth Generation District Heating and Cooling
ARI	Adjusted Rand Index
ATES	Aquifer Thermal Energy Storage
BCC	Block Completeness Coefficient
CoSEM	Complex Systems Engineering and Management
CS	Cluster Size
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DFC	Demand Fulfilment Coefficient
DOC	Demand Overlap Coefficient
NHD	Net Heat Demand
PVT	Photovoltaic thermal collector
RI	Rand Index
SLC	Single Linkage Clustering
WKO	Heat and Cold Storage

List of Symbols

ε	DBSCAN parameter Epsilon
d_{avg}	Average Distance between two buildings in a cluster
$minPts$	DBSCAN parameter minimum Points
t	Time-step
$Q_{c,b}$	Cooling demand of a building b
$Q_{h,b}$	Heating demand of a building b
$Q_{pvt,b}$	Thermal generation of the PVT system of a building b

List of Figures

List of Tables

1 Introduction

To counteract climate change, the European Union aims to achieve net-zero emissions as the first continent by 2050 [1]. These ambitious plans require new solutions across all sectors. In 2022 the heating sector was responsible for 41% of the total energy demand in the Netherlands [2]. Dutch residential and commercial buildings are mostly heated with natural gas [3]. Based on the European goals, the Dutch government has developed the Dutch Climate Agreement, which includes phasing out natural gas for heating buildings [3]. Next to climate change, the public support for phasing out natural gas is also based on the earthquake risks when extracting natural gas, which caused significant damage in the north of the Netherlands [3]. To tackle the challenge of the heating transition, the Netherlands has implemented a policy framework. A key part of this framework is the so-called Transitievisie Warmte (Heat Transition vision), which is a local strategy for achieving the Dutch sustainability goals in the heating sector each municipality is required to develop. Reducing the dependency on natural gas in residential heating is a key part of the Transitievisie Warmte [4]. Since the Netherlands import 65.8% of their natural gas and natural gas reserves are concentrated in certain parts of the world, relying on natural gas also creates dependencies on countries with significant gas resources [5]. As seen in the global energy crisis following Russia's invasion of Ukraine, these dependencies pose energy security risks [5]. This adds to the urgency of the ambition to phase out natural gas. Complementary to the Transitievisie Warmte, regional heat plans called 'Warmteplannen' are made to facilitate cooperation across municipal borders and to optimize the usage of different sources and technologies.

There are several alternative solutions for residential heating. One alternative is direct electric heating, which requires sufficient renewable electricity to reduce greenhouse gas emissions. However, this is would cause residential heating to become very expensive and requires a high net capacity [3]. Finally, the resulting electricity demand from residential heating would compete with the electrification of other sectors [6]. Thus, using electricity for residential heating is not a favourable alternative. Other alternative heating solutions are biogas or hydrogen, which could potentially reuse current gas infrastructure. But they both have a limited supply and a strong dependence on local production and distribution systems, which makes them unattractive as alternatives for natural gas [3, 7]. Additionally, biogas faces significant technical and institutional barriers for large scale implementation in the heating sector [8]. For biogas and hydrogen, there is also a high industrial demand, which competes with other uses such as residential heating.

A promising alternative are fifth-generation district heating and cooling (5GDHC) networks [9], which can provide heating and cooling for all connected buildings and do not require high temperature heat sources. 5GDHC networks enable the exchange of thermal energy for heating and cooling between buildings [10]. Additionally, low temperature heat sources such as data centres, greenhouses and low temperature geothermal sources can be utilized in such a network [10]. 5GDHC systems have the potential to reduce emissions at low costs compared to alternatives and thus become a key technology in the energy transition [9]. Moreover, densely populated urban areas are ideal application areas due to the high heat demand density [3]. Nevertheless, densely populated urban areas exhibit significant challenges for the heat transition, such as spacial limitations. Advancing the heat transition in historic city centres such as the city centre of Amsterdam is especially challenging, due to severe spacial limitations and the poorly

insulated historic building stock.

Currently, natural gas is the main source for residential heating in the city of Amsterdam. As part of their climate ambitions, the city aims to be natural gas free in 2040 [11]. This goal has recently been postponed to 2050, which illustrates the difficulty of this transition. The old buildings in the city centre of Amsterdam have a high heat demand due to lack of insulation and a high cooling demand in summer due to the heat being trapped in the city centre. Since a certain standard of insulation is required, buildings in the centre of Amsterdam would need to be retrofitted for the implementation of a fifth generation district heating and cooling network. If small networks would be implemented in the city centre step-by-step rather than transforming the whole area at once, this retrofitting process could be partially executed in parallel to the implementation of 5GDCHC networks. Additionally, such a modular approach reduces the initial investment costs and enables faster implementation. Therefore, the area needs to be divided into smaller networks. Later on, the overall operation of the networks can be optimized by connecting them. As a first step in the planning process, buildings should be clustered based on optimization criteria, to partition of the area of interests into smaller networks.

1.1. Link to CoSEM

Fifth generation district heating and cooling networks are complex socio-technical systems. Next to the technical design of the network, the objectives of the different actors involved, such as the customers, the energy suppliers, the housing cooperations, network operators and operators of interconnected systems need to be considered for the planning of these systems. The partitioning of the city into cluster networks is a crucial step in the planning process and has a significant impact on the operation of the network. Therefore, the clustering method should create a partition that meets actor objectives and technical requirements. Additionally, the institutional setting is important for the implementation of a 5GDHC network. For example, the climate goals affect the solution space. Therefore, the topic can be clearly linked to the Complex Systems Engineering and Management Master Program (CoSEM). Moreover, especially the Optimization and Network theory skills acquired through the CoSEM program are applied in this thesis project.

1.2. Structure of the Report

The structure of the report is as follows. Firstly, a literature review on district heating and cooling, clustering algorithms and the planning of low temperature district heating and cooling networks is conducted in chapter 2. Based on this review, the research question and sub questions are derived, and the research approach is presented. Secondly, chapter 3 focuses on the problem analysis, which includes state-of-the-art approaches to partitioning cities into networks, an overview over the potential solution and a stakeholder analysis for the case study in Amsterdam. Based on this stakeholder analysis, key requirements for the clustering methods are defined. The methodology is presented in chapter 4 and includes the introduction of the optimization metrics, general design choices related to the case study, the selection of clustering algorithms and the method for the robustness assessment. Then, the data of the case study is analysed in chapter 5. Chapter 6 focuses on the initial tests of the clustering methods on datasets from different densely populated European city centres. Afterwards, the case study results under the assumption that every building participates are presented in chapter 7. The robustness assessment entailing the results of the minimax regret method, the cluster viability analysis and the partition overlap, is the focus of chapter 8. The previously presented research and analysis results are then discussed in chapter 9. General conclusions and policy advice for the case study are given in the penultimate chapter 10. Finally, possibilities for future research are discussed in chapter 11.

2

Literature Review

2.1. District Heating and Cooling Networks

2.1.1. History of District Heating

Traditionally, district heating networks distribute heat through pipes from a centralized source to connected buildings. The first generation dates back to the 1880s. With the development of new technologies, the heat carrier of district heating networks changed over time from steam to superheated water (second generation) and hot water with temperatures below 100°C (third generation). These advancements not only decreased heat losses but also improved the safety of district heating networks, as steam explosions were a significant risk in the first generation of district heating networks. Moreover, the third generation of district heating networks was the first to support the integration of distributed waste heat sources and renewable heat sources due to the lower supply temperature. Fourth generation district heating systems are defined by their low supply temperature, which enables the use of low temperature sources, and the integration into smart energy systems. [12]

2.1.2. 5th Generation District Heating and Cooling Networks

A 5th generation district heating and cooling (5GDHC) network is characterized by the bidirectional exchange of thermal energy between buildings, optimizing the share of distributed low temperature heat sources [13]. Furthermore, 5GDHC networks utilize the synergy of heating and cooling in areas of residential and other types of buildings [14]. Additionally, according to Buffa et al. [15] fifth generation district heating and cooling (5GDHC) networks have operating temperatures close to the ground, enabling the utilization of various waste and renewable heat sources and bidirectional operation. Thus, customers can become prosumers by for example providing waste heat from space cooling or their own Photovoltaic thermal collectors on the roof in a 5GDHC system. The typical network supply temperature ranges from 5° to 35°C, which is too low to be directly used for space heating. Therefore, water source heat pumps (WHSP) are used to elevate the temperature to accomodate space heating, which makes them key elements in 5GDHC networks [15]. Since those heatpumps require electricity, implementing a 5GDHC network also has implications for the electricity grid in the respective area. Furthermore, a 5GDHC network typically exhibits a ring topology as seen in figure 2.1.

2.1.3. Planning 5GDHC Networks

An overview of the reviewed literature on recent developments and planning of fifth generation district heating and cooling (5GDHC) networks can be seen in table 2.1. The concept and characteristics of fifth generation district heating and cooling networks are frequently discussed in current research. However, implementations of 5GDHC networks are limited, which also leads to a lack of planning and operation guidelines [20]. Furthermore, Gjoka, Rismanchi, and Crawford [17] and Revesz et al. [18] explicitly discuss the lack of general guidelines for designing and planning 5GDHC systems. General guidelines can be a driver for the large scale deployment of 5GDHC systems. Roosien et al. (2020) developed a technical design framework for low temperature heating and cooling networks to fill this gap [13]. The framework recommends a bottom-up approach for planning 5GDHC systems, which entails first matching customers on a small local scale before connecting smaller networks later on for optimized operation [13]. Hence, for the implementation of smaller networks that can be connected later on, a large, densely populated area has to be divided into building clusters that form the individual networks. However, how

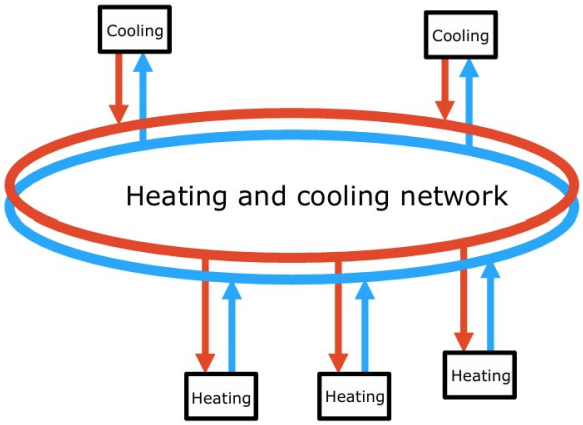


Figure 2.1: Schematic illustration of ring topology adapted from [13]

Table 2.1: Overview Initial Research on 5th Generation District Heating and Cooling Networks

Article	Requirements & Characteristics	Planning	Existing Networks	Implementation Barriers
Boesten et al. (2019) [16]	✓		✓	
Buffa et al. (2019) [15]	✓		✓	
Gjoka et al. (2023) [17]	✓			
Gjoka et al. (2024) [9]		(✓)		
Lund et al. (2021) [14]	✓		(✓)	✓
Revesz et al. (2020) [18]		(✓)	✓	
Rossien et al. (2020) [13]	✓	✓		
Verhoeven et al.(2014) [10]	(✓)		✓	
Volkova et al.(2022) [19]	✓			✓
Wirtz et al. (2022) [20]	✓		✓	✓
Wirtz et al. (2023) [21]	✓			

to divide a city into smaller networks is not discussed by the author. Moreover, the framework of Roosien et al. is the only concrete planning guideline found in literature (see table 2.1). Gjoka et al. (2024) performed a Life-cycle analysis of 5GDHC networks and evaluates the environmental impact, which is aimed at supporting early planning stages but does not give further insights into the planning process [9]. Moreover, the authors do not present any planning guidelines. Revesz et al. (2020) propose integrated smart energy systems connecting thermal and power networks with smart controls for big cities such as London [18]. Furthermore, they propose the GreenSCIES (Green Smart Community Integrated Energy Systems) design methodology for those systems [18]. The steps in this design methodology are rather high level with the initial steps being (1) identify physical boundaries, (2) find initial energy network clusters and (3) map sources and demands [18]. The following steps focus on control, financing and modelling the network [18]. While the presented idea of an integrated smart energy system is interesting, it creates additional uncertainty (such as impacts on the electricity grid) for the heating transition in cities. Therefore, first developing a 5th generation district heating and cooling network and then connecting it to the electricity grid for a smart and optimized integrated energy system might be preferable. Furthermore, for the development of a 5GDHC network Revesz et al. (2020) does not provide detailed guidelines.

The majority of implemented 5GDHC networks have a length under 10km, with many being around 1km long [20][15]. This ensures lower heat losses and lower costs [12]. Therefore, a larger densely populated

urban area needs to be divided into smaller building clusters. Each cluster is an own small scale network, where all connected buildings can exchange heat and cold [10]. This also has the advantage that rather than implementing a district heating system all at once in a city, a step-by-step approach is possible.

Moreover, the majority of existing networks are implemented in newly developed areas with new buildings [20]. Building a 5GDHC network in an already existing densely populated urban exhibits increased complexities such as the lack of space, the state of thermal insulation of the buildings and compatibility of existing equipment such as radiators. Additionally, the question arises if building owners are willing to connect to the network. In the survey on currently implemented 5GDHC networks by Wirtz et al. (2022) the authors state that for many of the built networks, participation was mandatory for some or all buildings [20]. These political decisions were made to ensure low costs and the energy balance of the network. For an area with only existing buildings, it will most likely not be an option to make the connection to the network mandatory. Building owners' unwillingness to participate can impact the overall costs, energy balance and reliability of the network and is thus an implementation barrier. One reason why people might hesitate to participate is that once connected, the supplier cannot be switched easily, which leads to a monopoly position of the network operator [20, 15]. Furthermore, building owners might prefer other heating solutions, such as their own ground source heat pump. Kontu et al. [22] found that for an area in Helsinki depending on the sources and design of the district heating network, people preferred having their own ground source heat pump or pellet boiler (using biomass). Next to the availability of alternatives also social factors, such as whether neighbouring buildings participate, play a role in the decision of each building owner. However, accounting for these social factors is out of the scope of the analysis within this project.

2.2. Clustering Algorithms

2.2.1. Classification

Clustering Algorithms divide data into groups based on a similarity measure [23, 24]. Since many different fields such as computational biology, image processing and spatial database applications utilize them, numerous methods have been proposed in research. However, there is no agreed upon approach for choosing the best suitable clustering method for a specific dataset [25]. Berkhin [23] presents a classification of clustering methods into Hierarchical methods, Partitioning relocation methods, Density-based partitioning methods and others. Furthermore, Hierarchical clustering algorithms combine (agglomerative approach) or divide (divisive approach) clusters into new clusters and thereby create a dendrogram, which is a tree structure illustrating the hierarchy of clusters [23, 24]. Benefits of hierarchical clustering methods are their independency of initial conditions and their applicability to a wide range of data even if the number of clusters is unknown [24, 23]. Single Linkage Clustering is an hierarchical clustering method that is suggested in previous work for clustering buildings for 5GDHC networks [26]. Partition relocation clustering methods on the other hand, divide the respective data into subsets, which are then iteratively improved with optimization techniques [23]. This gradual improvement of clusters by moving entities from one cluster to another is an advantage compared to hierarchical methods [23]. One example for partition relocation clustering is the K-Means algorithm, which was previously applied to 5GDHC networks with Aquifer Thermal Energy Storage (ATES) [27]. Density-Based Partitioning relies on the idea that a cluster is a connected dense component, which expands in any direction that density leads [23]. Therefore, density-based methods can detect arbitrary shapes and are resistant against outliers, which may be advantageous depending on the data [23]. Since a metric space is a necessity for density-based algorithms, they are commonly applied on spatial data [23].

2.2.2. Related Applications

Schiefelbein et al. [28] propose a clustering method based on street structure for clustering buildings for energy system planning. Furthermore, they use the street network in the form of a graph and buildings as nodes and compare their method to the K-Means algorithm. According to their research the compared methods all led to similar partitions and no clear best approach was identified. The authors also point out that it is a disadvantage of clustering methods to not take into account street patterns, since energy infrastructure is often installed along street networks [28]. Pilehforooshha and Karimi [29] compare a modification of DBSCAN with the basic DBSCAN algorithm to cluster polygonal buildings into urban blocks.

Within this work they acknowldge limitations of DBSCAN when clustering buildings in areas where the density varies. Loustau et al. [30] utilize K-Medians and a density-based approached called the GaussianMixture to group districts of a city for urban energy systems. Here, districts are clustered based on the similarity in characteristics instead of location. The authors state that K-Means is often and effectively used in the context of energy modelling [30]. However, for their study they chose the K-Medians over the K-Means due to its robustness against outliers in the data.

2.3. Knowledge Gaps

From the initial literature review, it can be concluded that one of the first steps while planning a 5GDHC network in a densely populated urban area is to divide the buildings into clusters. However, there is a knowledge gap regarding how to best partition an urban area into clusters for planning 5th generation district heating and cooling networks. Additionally, to mitigate the uncertainty regarding the cost and energy balance of the system, the cluster networks should be viable, even if not all building choose to connect to the network. How to deal with participation uncertainty is also not discussed in literature. Furthermore, the literature review on clustering methods and their application revealed a knowledge gap regarding the selection of the most suitable clustering method for the application on specific data. Therefore, the following main research question is formulated as follows based on the identified knowledge gaps:

Research Question

What is the best clustering method in terms of robustness for a fifth generation district heating and cooling (5GDHC) network in densely populated areas?

2.4. Research Approach

This research builds on previous work by Sarah van Burk [26] and Thom van den Akerboom [27], who developed two specific clustering methodologies to cluster buildings in densely populated urban areas. Moreover, Sarah van Burk [26] used Single Linkage clustering and graph theory, while Thom van den Akerboom [27] used the K-means algorithm for clustering and improved the overall heat balance by adding ATES. In contrast to the previous research, this work focuses on comparing different clustering methods and selecting the most promising method. Moreover, the methods developed in the previous work will be compared to other approaches identified in this Master Thesis project.

The aim of this research is to compare different clustering methods in terms of robustness. For this project robustness of clustering method in the context of 5GDHC networks is defined as follows:

Robustness is the ability of a clustering method to produce partitions that maintain the desired performance facing uncertainties in the planning process such as the participation rate of building owners.

This definition was derived based on the definition by Chalupnik (2013) for robustness in the context of building performance [31]. To identify the most robust clustering method, a methodology for comparing clustering approaches based on their performance and robustness for 5GDHC networks will be developed within this thesis. Moreover, within this work a case study on the city centre of Amsterdam is used to enable an in-depth analysis within a real setting. By applying the developed method to the city centre of Amsterdam and evaluating the results, this research will also provide insights for the next steps for the 'High-hanging fruit' project of the AMS institute. The 'High-hanging fruit' project aims at advancing the heating transition for the centre of Amsterdam while preserving the historical buildings [7]. Moreover, the three main objectives of the project are to develop an approach for retrofitting historical buildings, to investigate low temperature heat sources and to develop a strategy for the step-by-step energy transition of the inner city [32].

2.4.1. Sub-Questions

To answer the main research questions different steps are necessary. First, important factors that play a role in how the area should be divided into clusters must be identified. Furthermore, to find good partitions for the investigated area, the identified requirements should be translated into optimization criteria. Next, suitable clustering methods that can optimize the partition of the city for the specified metrics need to be identified. When applying the clustering algorithms to the area of interests the unwillingness of building owners to connect to the network might affect the quality of the clustering solution later on. Finally, the different clustering results for the participation samples and clustering methods need to be compared. Significant changes in the partition created by a clustering method can be a barrier in the planning process. This leads to the following sub-questions:

SQ1 - What are the requirements for dividing densely populated areas such as the city centre of Amsterdam into clusters for 5GDHC networks?

SQ2 - Which metrics can be utilized as optimization criteria for the clustering algorithms based on the identified requirements?

SQ3 - Which algorithms are promising for clustering areas with respect to district heating and cooling networks?

SQ4 - How does uncertainty regarding participation affect the performance of the different clustering algorithms?

SQ5 - Which clustering algorithm is the most robust and what policy conclusion can be drawn for the particular case study?

3

Problem Analysis

3.1. Clustering Buildings for 5GDHC

The most commonly used tool for planning District Heating and Cooling networks is VESTA Mais (conversation with P. Voskuilen, January 2025). VESTA Mais is an Energy system model for the Built Environment developed by PBL Netherlands Environmental Assessment Agency [33]. For low temperature district heating networks VESTA Mais divides buildings into clusters based on profitability [33]. Moreover, first the region where it is theoretically possible to implement a Heat and Cold Storage (WKO) is determined. Then it is assessed per building whether it is economically beneficial to connect the building to the network based on the building demand. Buildings that are deemed not profitable to connect will not be considered for the rest of the analysis. These buildings are then assumed to implement stand-alone systems even though these stand-alone systems might not be feasible, available or affordable for the respective buildings. The connections between buildings are then determined by iterating over all buildings and clusters and assessing the economic benefit of each connection. Although Vesta Mais has the advantage that decisions are based on economic implications, which directly enforces the affordability principle of the dutch heating transition, it also exhibits drawbacks. One major disadvantage of VESTA Mais is that it only considers inidividual buildings or small parts of the potential network rather than interconnected components or the system as a whole [34]. Moreover, buildings that are deemed unprofitable in the first assessment are overlooked when determining the best solution for the respective neighbourhood. Leaving out certain buildings from the neighbourhood solution without ensuring the availability, feasibility and affordability of alternatives for those buildings is problematic, since feasible heating solutions for all buildings in the neighbourhood must be found. Additionally, Vesta Mais considers demand and supply in seperate steps rather than simultaneosly to optimize the overall system.

Another approach to dividing a city into clusters for a district heating and cooling networks is the neighbourhood based approach as recommended by Nijpels [35]. In this approach district heating and cooling networks are planned for each neighbourhood separately which leaves out a considerable amount of possibilities to connect buildings. This limitation of the solution space potentially excludes connections between buildings that improve the energy balance of the overall solution or have other advantages such as reducing the overall pipe length.

3.2. Case Study Amsterdam City Centre

To enable an in-depth and realistic assessment of the robustness of clustering methods for 5GDHC networks in densely populated urban areas a case study is conducted within this work. The focus of this case study is the city centre of Amsterdam, which is particularly challenging due to poorly insulated building stock, many historic monuments and limited space. This limits the implementation of new infrastructure [36], which is required for a low temperature district heating and cooling network.

3.2.1. 5GDHCN Potential

The heat demand density is the heat demand per area per year and an important indicator of the potential for the application of 5GDHCN of an area. A higher heat demand density indicates a higher potential to utilize 5GDHC networks. The lower boundary of the heat demand density for a viable network is 165 $\frac{\text{MWh}}{\text{ha}\cdot\text{yr}}$

(personal correspondence with Meave Dang, January 2025). This corresponds with 5.94 $\frac{\text{TJ}}{\text{km}^2 \cdot \text{yr}}$. The heat demand density in the city centre of Amsterdam is visualized in 3.1., extracted from the Pan-European Thermal Atlas [37]. It can be seen that the heat demand density in the city centre of Amsterdam exceeds the lower limit significantly with a large share of areas exceeding a heat demand density of 300 TJ/km^2. This indicates a high potential for the application of a 5GDHC network in the city centre of Amsterdam. One implementation barrier for the application of 5GDHC systems in the centre of Amsterdam is that the majority of the building stock is poorly insulated. Thus, to be able to utilize a low temperature network buildings need to be retrofitted. Nevertheless, low temperature heating and cooling networks are flexible and enable a step-by-step connection of buildings [32]. This means that the processes of retrofitting the buildings and implementing the 5GDHC system can be executed in parallel to some degree.

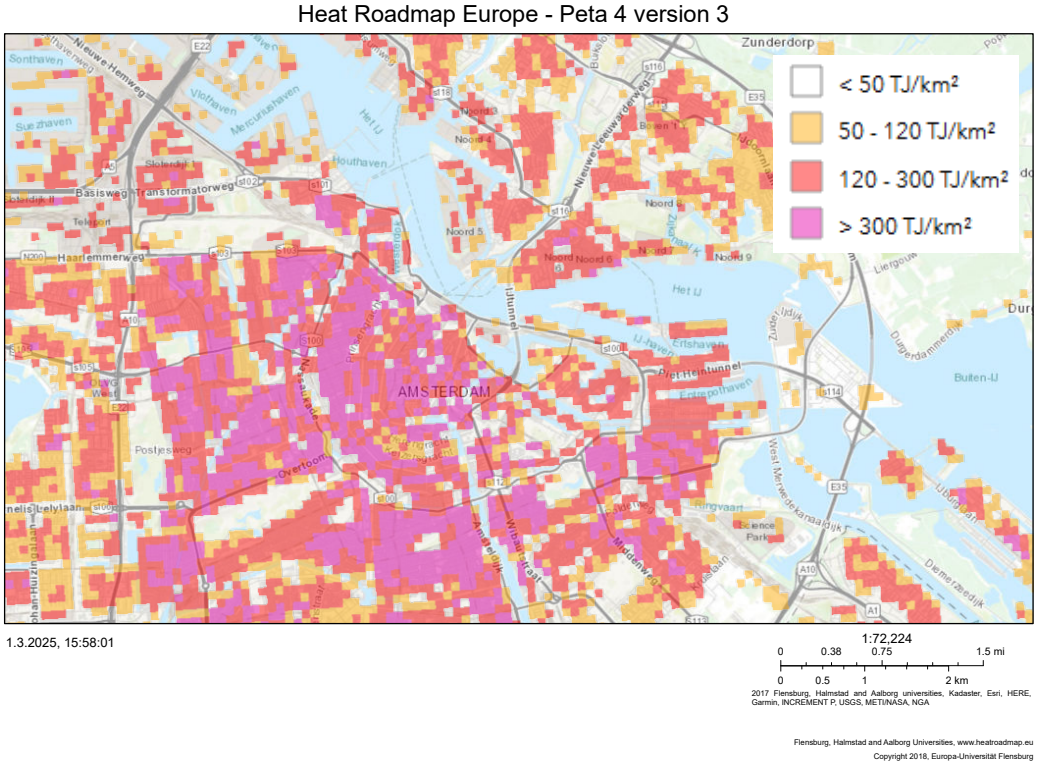


Figure 3.1: Heat demand density in Amsterdam in 2015 adapted from [37]

3.2.2. Potential Heat Sources

In the city centre of Amsterdam, there are no large (waste) heat sources such as a datacentre or a geothermal well. Furthermore, waste heat could only be collected from small sources such as supermarkets. Due to the lack of data and clarity about the contracting of waste heat sources (conversation with P. Voskuilen, September 2024) sources such as small supermarkets are not considered in this research. Another possible heating and cooling source are aquathermal applications. Aquathermal applications extract heat from or store thermal energy in surface water, wastewater or drinking water [38]. Thus, in Amsterdam the canals could be used as an additional heat source or thermal storage. However, there are strict regulations about how much the temperature of water bodies can be changed. This means that if one cluster uses a canal or other water body within the network, the usability of that water body for other clusters is limited. Therefore, the aforementioned potential sources are excluded from the analysis in this work.

A promising heat source for the city centre of Amsterdam are Photovoltaic Thermal Systems (PVT) [39]. PVT systems generate both electricity and thermal energy, which optimizes the utilization of the limited space in densely populated urban areas [39]. Furthermore, PVT systems can contribute to 5GDHC net-

works not only through supplying additional thermal energy but also by supplying electricity for the heat pumps in the network. For the 5GDHC system for the city centre of Amsterdam in this work PVT systems are assumed as the main heat source next to the synergies between cooling and heating demands.

3.2.3. Heat Storage Solutions

In general, buildings in moderate climates have a cooling surplus in winter and a heating surplus in summer [40]. Therefore, seasonal storage is a key component of a low temperature district heating network for the city of Amsterdam. Additionally, seasonal storage can elevate the overall efficiency and decrease the levelized cost of energy of the system [41].

One promising storage solution for low temperature heating and cooling networks are Aquifer Thermal Energy Storage (ATES) systems. ATES systems extract and inject groundwater to extract and store heat in the subsurface [40]. These systems were already identified as suitable for the city centre of Amsterdam in previous work [27]. Therefore, ATES systems are assumed to be chosen the heat storage solution for the city centre of Amsterdam within this research project.

3.2.4. Stakeholder Analysis

In this section, the roles and interests of different stakeholders within a potential district heating and cooling network in the city of Amsterdam are analysed. Based on this analysis, requirements for the network but also specifically for the division of the city into cluster networks are derived.

The **residents of the city centre of Amsterdam** are the prosumers in the fifth generation district heating and cooling network. Their key interest is a reliable and affordable heating solution that meets their demand. Thus, a potential clustering method should ensure viable cluster networks with sufficient supply and demand. To ensure the reliability and the affordability, each network should be compact, i.e. short pipe lengths, which reduces costs and heat losses. The compactness of the network is especially relevant since the recommended topology of a 5GDHC network is a ring structure meaning every building is connected with every other building in the same cluster. Moreover, from the residents' point of view, an important requirement for a heating solution is also the usability [22]. Additionally, residents might care about the sustainability of their heating system like whether the heat is generated with renewable sources or not.

A significant share of buildings in the city centre of Amsterdam is owned by housing cooperations. **Housing cooperations** and other **building owners** make the decision about connecting to the 5GDHC network and the involved investment for equipment such as heat pumps and compatible radiators or floor heating. Important factors that influence their decision are the usability of the system, investment costs and the expected reliability and potential independent alternatives [22]. Furthermore, examples for alternative solutions are stand-alone Aquifer Thermal Energy Storage (ATES) systems or Air-Source Heat pumps. Moreover, when planning a district heating and cooling system for Amsterdam, it is important to ensure that connecting to the network is attractive to the building owners. As connecting to the network will also require an investment from the building owners into equipment such as a heat pump for their houses. Therefore, warranties and guaranteed servicing and maintenance is a key requirement (conversation with Nathan Industries, December 2024). This ensured servicing of the equipment is also important to residents as it directly affects the reliability of the heating system.

The **municipality** develops the long-term heating strategy for Amsterdam. In addition, the municipality has to develop specific neighbourhood execution plans [3]. Both these neighbourhood execution plans and the long-term strategy needs to adhere to affordability and feasibility principles, which are core principles defined by the Dutch government for the heating transition [3]. The feasibility principle also relates to feasible connections between buildings within the spacial constraints. The canals in the city centre of Amsterdam limit the feasible and affordable connections between buildings. Connecting buildings along streets is less challenging, more affordable and spatially efficient compared to connections across streets or canals. Therefore, a suitable clustering method should generate a partition of the city that follows the street structure for a feasible and affordable implementation of the networks. For a potential communal heating solution like a district heating and cooling network for the city the municipality will collaborate

with external companies and the other stakeholders in the planning and implementation process. Involving all stakeholders is important to the municipality to ensure a successful implementation of new heating solutions. Therefore, a suitable clustering method should be easy to explain and to understand. This way the clustering method can be utilized to engage different stakeholders in the planning process. Moreover, key requirements for a potential 5GDHC system in a densly populated urban area from the municipality point of view are starting small for faster implementation, spatial integration and reliability (conversation Mimi Eelman, Gemeente Amsterdam, December 2024). Due to the spatial limitations in densely populated city centres and the importance of the affordability of the solution it is important the cluster networks are compact, which entails buildings in a network should be in close proximity to one another.

Since heat pumps are key elements of fifth generation district heating and cooling networks [15], the **Electricity grid operator** is also an important stakeholder in the system. As the operator of the connected network, their main interest is to limit the stress on the grid caused by the heat pumps and avoid congestion [42]. Therefore, the cluster networks should have a good energy balance to reduce the electricity requirement of the heat pumps. An analysis of the implications for the electricity grid is out of scope for this project.

To conclude, seven key requirements for the clustering method and the generated partition can be derived from the actor analysis:

1. The Energy Balance in every cluster should be sufficient.
2. The resulting cluster networks should be compact.
3. The cluster method should follow the street structure.
4. The cluster networks should be small in terms of number of buildings.
5. The division should have similarly sized clusters.
6. The logic behind the clustering method should be easy to understand.
7. The clustering method should be easy to use.

In the next step optimization metrics to evaluate the performance of partitions are derived based on the here identified requirements. Requirements such as an easy-to-use and easy to explain method are not utilized for measuring the partition performance but rather assessed during the application of different clustering methods.

4 Methodology

4.1. Optimization Metrics

Based on the previously defined requirements for clustering buildings in a densely populated urban area optimization metrics are defined to measure the quality of a cluster network. Based on the results of these metrics for all clusters in a partition, the performance of the partition as a whole is assessed.

4.1.1. Cluster Size

One of the requirements derived from the problem and actor analysis is that the clusters should be small to support a modular approach. This means that the number of buildings that will be connected in a cluster network should be limited. A simple metric to account for this is the cluster size, which in this case is defined as the amount of buildings in a cluster. To calculate the cluster size CS of a cluster C the amount of buildings in this cluster is counted.

$$CS = |C| \tag{4.1}$$

4.1.2. Average Distance between Buildings in a Cluster

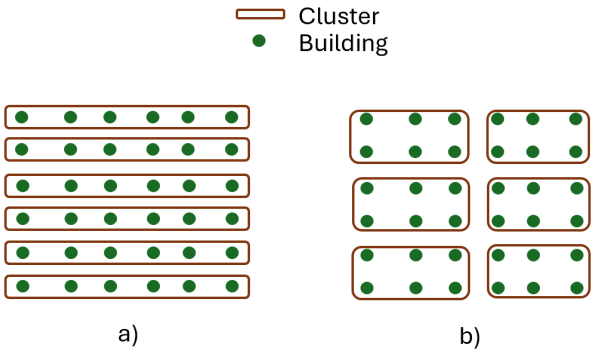


Figure 4.1: Two Partitions a) and b) of a set of buildings with different Average Distances within Clusters

In addition, a clustering method should lead to compact cluster networks where the total pipe length is short. Figure 4.1 illustrates two different partitions of buildings into clusters in an example area. An implemented cluster network connects all buildings with each other and follows a ring topology. Partition b) would be preferred over partition a) in the figure since the cluster networks are more compact and the total pipe length would be shorter. A suitable metric to express the compactness requirement is the average distance between two buildings in a cluster (d_{avg}). The d_{avg} is calculated per cluster C by calculating the Euclidean distance between all sets of two distinct buildings i and j in C , summing up those distances and dividing the sum by the number of building pairs in the cluster:

$$d_{avg} = \frac{1}{|C|(|C|-1)} \sum_{i \in C} \sum_{j \in C} d(i,j) \quad \text{where } i \neq j. \tag{4.2}$$

A more realistic way to calculate the distance between two buildings in a cluster would utilize the street network in the calculation. However, this would lead to a high computation time, which would make the assessment inefficient. Therefore, the Euclidean distance, which is a simpler and more efficient measure for the distance between buildings is preferred. This also ensures the applicability of the here defined metric to datasets of various sizes. In general, the distance measured with the street network and the Euclidean distance are highly correlated [43], which indicates that the Euclidean distance is a suitable simplification.

4.1.3. Block Completeness Coefficient

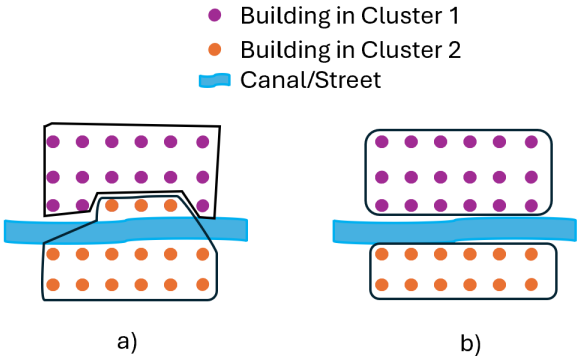


Figure 4.2: Two examplatory Partitions a) and b) of a set of buildings divided by a canal or street

Following the street network is an important requirement for a cluster, since it indicates the feasibility of implementing the necessary infrastructure. Therefore, buildings that are in a continuous area which is surrounded by streets or other rights-of-way should be in the same cluster. Such a set of buildings can also be referred to as urban block or building block. The defining characteristic of those blocks is that buildings within them are not interupted by public rights-of-way. Therefore, these building blocks directly relate to the street layout of the respective area and the completeness of these blocks in a cluster is an indicator for the alignment with the street structure. Figure 4.2 shows two different ways to divide a hypothetical small area with two urban blocks separated by a street or canal (blue) into two clusters. Partition a) in figure 4.2 connects a few buildings of one block over a canal to another block. This division does not follow the underlying street structure. The connection of a few buildings over a canal or street is typically longer and especially in the case of a bridge over a canal more difficult to install. Moreover, if one were to connect buildings over a canal or a street the question arises why connect only a few buildings if there are more buildings in the respective urban block that would only require a short connection. In contrast to figure 4.2 a), figure 4.2 b) is better aligned with the street structure. Therefore, a suitable metric should indicate that partition b) is preferable in figure 4.2. To quantify whether a cluster has complete building blocks and thus a good alignment with the street structure, the block completeness coefficient is defined. The coefficient is calculated as the number of buildings in a cluster divided by the sum of the number of buildings in each block, where at least one building is in the respective cluster (see equation 4.3). The denominator can be described as the number of buildings that would be in a cluster if all building blocks in the cluster are completely in the cluster.

$$BCC = \frac{|C|}{\sum_{B\,:\,B\cap C\neq\emptyset} |B|} \tag{4.3}$$

In equation 4.3, C denotes the cluster, B is a building block that is at least partially in cluster C as specified by the condition $B\,:\,B\cap C\neq\emptyset$. The resulting BCC is a value between 0 and 1. Moreover, values close to 0 indicate that the cluster consist of buildings belonging to certain building blocks, where the majority of buildings from the same building block are not in the cluster. In contrast to that a BCC close to 1 implies that the building blocks of the cluster are almost completely in the cluster. The BCC can also be written in percentages for easier interpretation.

4.1.4. Demand Fulfilment Coefficient

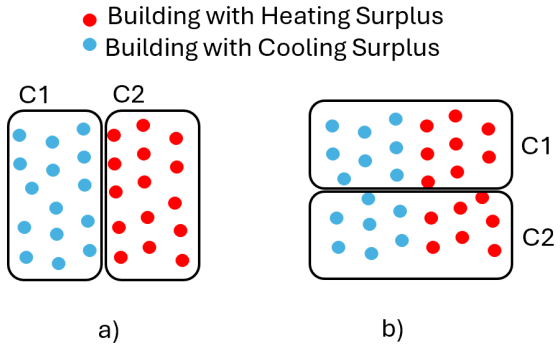


Figure 4.3: Example Partitions of Buildings with different Heating and Cooling Demands

One key requirement for the clustering method is that it should generate partitions, where each cluster has a good energy balance based on heating and cooling demands and on additional thermal sources or storage. Figure 4.3 shows two different partitions of a set of buildings into clusters. In this direct comparison, partition b) would be preferred as both clusters in b) have buildings with a heating surplus that can supply the buildings with a cooling surplus and vice versa. Thus, an optimization metric that can indicate that preference for matching buildings with a cooling surplus with buildings with a heating surplus is needed.

Wirtz et al. (2020) developed a metric called the Demand Overlap Coefficient (DOC) that indicates the efficiency of a 5GDHC network [44]. Moreover, the DOC of a cluster network describes how much of the total heating and cooling demand is balanced between buildings [44]. The DOC of a cluster C over a time T that is divided into discrete intervals t is calculated as [44]:

$$DOC = \frac{2 \cdot \sum_{t \in T} \min \left\{ \sum_{b \in C} \dot{Q}_{\text{h,dem},b,t}, \sum_{b \in C} \dot{Q}_{\text{c,dem},b,t} \right\}}{\sum_{t \in T} \sum_{b \in C} \left(\dot{Q}_{\text{h,dem},b,t} + \dot{Q}_{\text{c,dem},b,t} \right)} \tag{4.4}$$

Here, $\dot{Q}_{\text{h,dem},b,t}$ is the heating demand of a building b, $\dot{Q}_{\text{c,dem},b,t}$ is the cooling demand of a building b and B denotes all buildings in the cluster network. The DOC is a value between 0 and 1, with 0 indicating that none of the heating demand can be balanced with the cooling demand and 1 meaning that the heating and cooling demand profiles match perfectly [44].

In previous work a metric inspired by the DOC called the Demand Fulfilment Coefficient (DFC) is derived [27]. In contrast to the DOC, the DFC also includes external sources such as external thermal generation or storage [27]. This is relevant for this research since PVT systems on the rooftops of each building are considered. The thermal generation of a building changes the impact of it on the overall energy balance within the cluster network and should therefore be considered when the energy balance is quantified. Thus, within this work, the DFC is used as an optimization metric and calculated as follows for a cluster C over a time T that is divided into discrete intervals t:

$$DFC = \frac{\sum_{t \in T} min \left\{ \sum_{b \in C} \dot{Q}_{\text{h,dem},b,t}, \sum_{b \in C} (\dot{Q}_{\text{c,dem},b,t} + \dot{Q}_{\text{pvt},b,t}) \right\}}{\sum_{t \in T} max \left\{ \sum_{b \in C} \dot{Q}_{\text{h,dem},b,t}, \sum_{b \in C} \dot{Q}_{\text{c,dem},b,t} \right\}} \tag{4.5}$$

Here, $\dot{Q}_{\text{pvt},b,t}$ denotes the thermal generation of the PVT system of building b in time interval t. Like the DOC, the DFC is a value between 0 and 1, where 0 means that no part of the heating demand can be

fulfilled by either the cooling demand or the thermal generation in the cluster and no part of the cooling demand can be fulfilled by the heating demand. A DFC of 1 implies that both the heating and the cooling demand are completely fulfilled within the cluster. Thus, the higher the DFC the better the performance of the cluster.

4.1.5. Partition Performance Assessment

With the previously introduced optimization metrics the performance of an individual cluster with respect to the related requirements can be quantified. A good partition is characterized by desirable values of the optimization metrics for all clusters and little variation in performance between clusters. Little variation regarding the metrics between clusters is important since the heat transition should be fair meaning that all cluster networks should be performing similarly well on the identified metrics. Therefore, to evaluate the performance of a partition the optimization metrics are calculated for all clusters. The average, standard deviation, minimum and maximum of the optimization metrics are then used to quantify the performance of the partition. A well performing partition has close to optimal average values, a low standard deviation and a low spread, which is the difference between the maximum and minimum value, for all optimization metrics.

4.2. Design Choices

The implementation of a fifth generation district heating and cooling network should follow a modular approach, where small cluster networks are implemented step-by-step. For this modular approach so-called mini networks with a size from 2-50 connections or small networks, which usually have 51-1500 connections [45], are suitable. The decision on the optimal size faces a trade-off between quick implementation and better utilization of synergies by including more connections. According to the Gemeente Amsterdam, 200 connections are required for a viable network in the city centre of Amsterdam (personal correspondance Mimi Eelman). This corresponds to 50 buildings needed, assuming 4 connections per building, which corresponds to the average for buildings in the city centre of Amsterdam (personal correspondance Voskuilen). Therefore, the target size is set to 50 buildings. Based on this target number the desired number of clusters can be calculated by dividing the total number of buildings by the target cluster size. Since mini networks can also be smaller than 50 buildings [45] the lower limit of buildings per cluster network is set to 10 buildings in a network. The upper limit of the cluster size is assumed to be 100 buildings to limit the implementation time and upfront investment costs of the individual networks.

4.3. Clustering Algorithms

In this section the considered clustering algorithms are explained. In general, only clustering metrics that divide data based on a single objective, which is the distance between data points, are considered to ensure the comparability with the clustering methods developed in previous work [26, 27]. Nevertheless, multiobjective clustering approaches could also be tested in future research. Measuring the distance between two points in a cluster can be done with the use of the Euclidian method which has been described in depth in section 4.1.2 . This distance can also be calculated more realistically utilizing the street network in the calculation. However, this would lead to a high computation time and would complicate the application of the clustering methods. By using the Euclidean distance the clustering algorithms also directly optimize the average distance between buildings in a cluster d_{avg}.

4.3.1. DBSCAN

One commonly applied example of density-based clustering methods is the DBSCAN [46]. Clusters are formed starting from a randomly chosen data point and adding points that are in pre-defined radius ε. If there are at least $minPts$ points within the radius from the initial point including the point itself, then the initial point is a core point of this cluster. The clusters are expanded by subsequently adding points within a ε radius of any core points in the cluster. If a point is not within the ε distance of a core point and does not qualify as a core point of a cluster itself, the point is not assigned to a cluster but instead labeled as noise. Noise points are not density reachable from core points [47]. In general, a high ε leads to larger clusters and less unclustered points (noise), whereas a low ε corrresponds with smaller clusters and more unclustered points (noise). A high $minPts$ parameter causes fewer clusters and more unclustered points (noise), while a small $minPts$ parameter leads to less unclustered points (noise) and more clusters.

One advantage of the DBSCAN is that it can detect non-convex shapes [48]. Furthermore, the creation of clusters based on the density of building points can be an advantage for clustering buildings since it is desireable that the clusters follow the street structure. One prominent issue with DBSCAN found in the literature is finding suitable values for the input parameters ε and $minPts$ [46]. To determine the parameters two approaches are tested within this work. The first approach is a simple tuning function that tests parameters within predefined ranges and stops when the amount of clusters is within a tolerance region of the desired value. Starczewski et al. (2020) developed another method to choose input parameters for DBSCAN [46]. Moreover, they calculate the ε parameter based on the k_{dist} function which calculates the distance between an element of a dataset and its k_{th} nearest neighbour [46]. The $minPts$ parameter is determined based on the dimension of the dataset and a factor that is defined during the analysis of the k_{dist} function [46].

4.3.2. Single Linkage Clustering

The Single Linkage Clustering (SLC) Algorithm starts with every point as its own cluster. Iteratively, the two closest clusters are merged until there is one cluster with all points. In case of quantitative data, the closeness of two clusters is their minimum distance to each other, which is the distance between the closest points from the two clusters. Like in this study, the euclidean distance is commonly used in literature to measure the distance within the SLC algorithm [49]. For qualitative data on the other hand similarity measures are used [49].

A drawback with Single Linkage Clustering is the so-called Chaining effect. This chaining effect describes elongated chain-like clusters which arise due to clusters being connected based on single points despite the rest of the points being further away [49]. Moreover, for a specified amount of clusters isolated points could become single point clusters while other clusters have a large amount of points. Thus, Single Linkage Clustering is senstitive to outliers [48]. On the other hand, SLC has the advantage that it follows the street structure of a city when clustering buildings, since the closest connection between two buildings is added in each step of the iteration when generating a partition. Another advantage of Single Linkage Clustering is that it recognizes non-convex shapes [48]. Furthermore, Single Linkage Clustering is efficient for large datasets [48].

4.3.3. K-Means

The K-Means algorithm divides data in a pre-defined number of clusters n. In the beginning the algorithm randomly chooses n initial centroids. Then every point is assigned to the closest centroid based on a distance metric [50]. The centroids are then iteratively recalculated by calculating the mean of all points in a cluster until the difference between the new and previous centroids is smaller than a predefined threshold [48]. Moreover, the objective function used by the K-Means algorithm is the Within-Cluster Sum of Squares [50], where K is the pre-defined number of clusters, c_k is the centroid of cluster k, x_i is a data point and S_k is the set of data points in cluster k:

$$J = (\sum_{k=1}^{K} \sum_{x_i \in S_k} \|x_i - c_k\|^2)^{1/2} \tag{4.6}$$

One drawback when using K-Means for clustering buildings is that it can only recognize convex and isotropic shapes [48]. Thus, K-Means does not recognize the street structure when clustering buildings.

4.3.4. K-Medians

The K-Medians algorithm is a variant of the K-Means algorithm [47]. In literature it is also referred to as KMedoids algorithm. The main difference is that the K-Medians algorithm uses the median of all points in a cluster to recalculate the centroids in the iteration steps [51]. Moreover, the centroid in case of the K-Medians is always a data point [47].

In comparison to the K-Means, the K-Medians is more robust, since outliers have a lower impact within the iterations [30] [47]. Therefore, the K-Medians algorithm is promising in the context of finding a robust

method for clustering buildings for low temperature district heating and cooling networks. A drawback of the K-Medians is that it is more computationally complex compared to the K-Means [47].

4.3.5. Methods with Optimization Steps

The here considered algorithms DBSCAN, SLC, K-Means and KMedian all optimize the partition based on a distance metric, namely the Euclidean distance, rather than optimizing all identified optimization metrics. Moreover, buildings that are considered close in distance are together in a cluster. However, the algorithms all operate differently and exhibit different shortcomings. For example, K-Means and K-Medians do not recognize the street structure, which leads to an expected lower Block Completeness Coefficient (BCC) in terms of the optimization metrics. Moreover, the SLC algorithm can create an unbalanced partition in terms of cluster size, due to the chaining effect and some buildings being further away from their direct neighbours than others. Thus, algorithms not necessarily produce the best results measured by the optimization metrics. Therefore, the performance of a partition measured by the optimization metrics is improved with optimization steps based on the shortcomings of the algorithms and partition requirements. All modifications are applied to all tested algorithms to ensure the comparability of the methods.

The different optimization steps, the affected metric, conditions for the improvement and related partition requirements are shown in table 4.1. First the clustering algorithms are applied, which optimize the partition based on the Euclidean distance. This directly relates to the average distance between two buildings of the same cluster d_{avg}, which reflects the compactness and affordability requirements for a good partition. In this first step no conditions related to other requirements are enforced, which implies the prioritization of this step. The order of the optimization steps implicates a prioritization as conditions are defined to preserve the effect of the previous optimization steps. A short average distance between buildings in a cluster is prioritized because the cluster networks should consist of buildings in close proximity to another to ensure a short pipe length. This ensures both the feasibility of the implementation with respect to spacial constraints and the affordability of the network.

Table 4.1: Overview Clustering and Optimization Steps

Step	Optimization	Metric	Conditions	Related Partition Requirements
1	Clustering Algorithms	Average Distance between to buildings in a cluster d_{avg}	None	Compactness, Affordability
2	Size Optimization Step	Cluster Size CS	*Average Distance*: Splitting based on Algorithms, Merging with closest cluster	Modular Approach, Evenly sized small Networks
3	Energy Optimization Step	Demand Fulfilment Coefficient DFC	*Average Distance*: Switching Radius *Cluster Size*: only move buildings bigger to smaller cluster	Energy Balance in Each Cluster
4	Block Completeness Optimization Step	Block Completeness Coefficient BCC	*Average Distance*: common blocks *Cluster Size*: one for one switch *Energy Balance*: Maximum allowed change 10%	Alignment with Street Structure

In the next step the variation of the cluster sizes is reduced to ensure small networks that can be imple-

mented step by step to transform the heating system of a densely populated urban area. This step is executed next to ensure the feasibility of the modular approach and limit the required individual investments for all cluster networks. For this Size Optimization step, first all too large clusters are identified. It is assumed that clusters should have a maximum of 100 buildings. Then the respective clustering algorithm is applied on each too large cluster, until all clusters have a maximum of 100 buildings. Afterwards, the cluster with less than 10 buildings are determined. Each building in those clusters is then added to the cluster with the closest centroid. The upper limit is not strictly enforced during the merging process, which ensures that buildings can only added to neighbouring clusters. Thereby, the negative effect on the average distance between two buildings in a cluster d_{avg} is limited.

Next, measures to improve the energy balance of the cluster networks as measured by the Demand Fulfilment Coefficient are applied. The energy balance is prioritized third in this methodology, since the compactness and cluster size have a stronger impact on the implementation of the cluster networks with respect to the modular approach and the critical spatial limitations in densely populated urban areas. Nevertheless, an efficient utilization of the synergies between heating and cooling demands and distributed heat sources is important for the operation of the cluster networks. In this third optimization step, clusters are iteratively improved through switches based on the net heat demand of the individual buildings and clusters. The net heat demand of a building b is defined as follows:

$$\text{NHD}_b = Q_{h,b} - Q_{c,b} - Q_{pvt,b} \tag{4.7}$$

Here, $Q_{h,b}$ is the heating demand of the building [kWh], $Q_{c,b}$ is the cooling demand of the building [kWh] and $Q_{pvt,b}$ is the generation of the PVT system of the building [kWh]. The net heat demand can be calculated for different time intervals. Since for the analysis within this research infinite storage is assumed buildings are switched based on the yearly net heat demand to improve the yearly energy balance. This can be easily adapted to another time resolution for future applications. To switch buildings between clusters, first all clusters with a thermal energy surplus, in other words a negative net heat demand of the entire cluster, need to be identified. For this purpose, the net heat demand NHD_C of a cluster C is calculated as the sum of all net heat demands of buildings in this cluster:

$$\text{NHD}_C = \sum_{b \epsilon C} \text{NHD}_b \tag{4.8}$$

Then, clusters with a thermal energy deficit, so a positive cluster NHD, are determined. The Energy Balance of these deficit clusters can be improved by donating buildings with a heat deficit to a surplus cluster or receiving buildings with a heat surplus from a surplus cluster. For each deficit cluster, the closest building from a surplus cluster within a pre-defined switching radius, which ensures the negative impact on the d_{avg} average and standard deviation is limited, is determined. An additional rule to preserve the previously improved average and standard deviation of CS for the partition is that buildings should only be moved from a larger cluster to a smaller cluster. Hence, if the deficit cluster has more buildings than the surplus cluster, the closest building with a heat deficit from the deficit cluster is reassigned to the surplus cluster. Similarly, in case the deficit clusters has less buildings than the surplus cluster the closest building with a surplus, so a negative net heat demand is reassigned to the deficit clusters. Once a surplus cluster has no longer a thermal energy surplus it is removed from the switching possibilities. If a deficit cluster is either no longer a deficit cluster or there are no more buildings within the switching radius the iteration moves on to the next deficit cluster.

For the energy optimization step the switching radius, in which switches to improve the energy balance are allowed, needs to be defined. Moreover, the larger this switching radius, the more possibilities there are to improve the clusters. However, the further away a building from the cluster it is reassigned to is, the higher the negative impact on the d_{avg} average and standard deviation of the partition. Additionally, the energy improvement step can negatively impact the performance measured by the average and standard deviation of the BCC if it splits up blocks between clusters. Thus, there is a trade-off when deciding on a suitable switching radius for the energy optimization step. Within this work multiple switching radii

are tested and different optimization metric values depending on the switching radius are visualized in a graph. Based on these graphs the switching radius is selected.

Finally, the performance of all algorithms based on the Block Completeness Coefficient (BCC) is improved. As summarized in table 4.1 buildings are switched between two clusters that share two or more blocks if the Block Completeness Coefficient of the two clusters is improved by the switch. With the condition of having at least two blocks in common it is ensured that switches only happen in close proximity, which prevents significant losses of the performance measured by the d_{avg}. Since one building is always swapped with a building from another cluster the cluster sizes stay the same. An additional condition is that the energy balance of the two clusters between which the buildings are swapped does not change by more than 10%. This ensures that a potential negative effect on the average and standard deviation of the demand fulfilment coefficient (DFC) is limited.

4.4. Robustness Assessment

This section is dedicated to the methods on how the robustness of the different clustering methods is assessed within this work. Based on these methods the main research questions and the sub questions 4 and 5 will be answered.

4.4.1. Scenarios

Which buildings will participate and which will not participate in a low temperature district heating and cooling network is highly uncertain. Assessing the robustness of the performance of the different clustering methods can increase certainty in the planning process and aid decision-making. According to Kotireddy et al. (2019) the performance of a design and its robustness should be equally considered when choosing the best design [52]. To account for uncertainty regarding the participation of buildings in the low temperature district heating and cooling network scenarios with different participation rates are considered in this analysis. The starting point for the scenarios is the initial partition of the clustering methods in case of a participation rate of 100 %. For each run of a scenario that assumes a certain percentage of participation, buildings are randomly removed from the initial partition. The selection of Non-Participants for the scenarios is randomized, since it is not known who will participate and removing a building can have very different impacts based on its net heat demand, location and other characteristics. Moreover, the non-participants are most likely not spread evenly throughout the area and pre-defined clusters, since social factors such as whether neighbours participate and the availability of alternatives for the specific building can play a role in the decision. An analysis of these factors are out of scope for this project. To make the outcomes of the analysis more representative each non-deterministic scenario is tested in 100 separate runs.

In addition, the probabilities for different participation rates are unknown. However, Rysanek et al. (2013) illustrate how a non-probabilistic approach can be useful in case of multiple scenarios to support the selection of a robust design [53]. For this research the following scenarios are considered:

- 100 % participation
- 95 % participation
- 80 % participation
- 70 % participation
  This scenario is closest to the expected participation in the city centre of Amsterdam (personal correspondance Maeva Dang from the AMS institute).
- 50 % participation
  This scenario is included to represent a participation rate below the expected rate.
- Non-Participants Utility buildings
  Next to the base case this is the only other deterministic scenario. It is included since utility buildings might be more likely to implement their own independent ATES system or a comparable alternative (personal correspondance Voskuilen).

4.4.2. Minimax Regret Method

The minimax regret method is a commonly used approach to assessing robustness considering different scenarios [52]. Following the minimax regret method, a good design or in this case a good clustering method should perform close to optimal performance in each case [52]. This also means that a certain range of deviation from optimal performance is accepted with this method [54]. Therefore, the minimax regret method is a less conservative approach to risk [52]. In this method the robustness indicator is the maximum performance regret [52]. For the minimax regret method so-called performance indicators are used, which are metrics that are selected by the user to measure the performance of the design in the different scenarios.

In this research, the average and standard deviation of the optimization metrics Average Distance between buildings in a cluster (d_{avg}), Block Completeness Coefficient (BCC) and Demand Fulfilment Coefficient (DFC) are used as performance indicators of the different clustering methods. The cluster size is excluded since randomly removing buildings from the initial partition has the same effect on the average and standard deviation of CS for all clustering methods. To apply the minimax regret method the performance indicators $PI_{x,s}$ of each clustering method x in each scenario s are calculated in the first step [54]. The regret $R_{x,s}$ is then the absolute difference between the performance of the best performing method PI_s^* and the performance of the clustering method x $PI_{x,s}$ in scenario s:

$$R_{x,s} = PI_s^* - PI_{x,s} \tag{4.9}$$

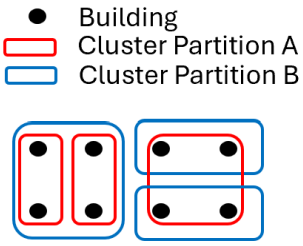
This way the regret of each clustering method is calculated for each scenario and each performance indicator. In case of the average BCC and the average DFC the best performing clustering method in a scenario is the one with the highest value. For the average d_{avg} the method with the lowest value in the scenario is the one that performs the best. For the standard deviation of all three metrics the best performing method in a scenario produces the lowest value. In the next step, the maximum regret is calculated for each method over all scenarios. The methods can then be ranked for each performance indicator from the most robust design with the lowest maximum regret to the least robust design with the highest maximum regret. Furthermore, for the minimax regret method the most average run is chosen per non-deterministic scenario. This most average run is determined based on the value of all performance indicators in the different runs. Additionally, the runs furthest away from the average per scenario are also screened for notable differences in performance and the ranking of the clustering methods. For the robustness analysis in this project all three considered performance indicators are weighted equally, which can be adjusted for future applications depending on the prioritization of requirements.

4.4.3. Identification of Viable Clusters

To implement small networks based on the most robust clustering methods in densely populated urban areas, it is important that the resulting networks are viable. If a cluster is viable across different participation scenarios it is robust with respect to participation uncertainty. Such a cluster can serve as a starting point for the implementation of the cluster networks. While a cluster might be viable for a specific subset of non-participants in a scenario, this might not be the case for another randomly chosen subset of buildings to remove. Therefore, in this analysis a cluster is considered viable in a scenario with a certain participation rate when it is fulfils the below defined viability criteria for all runs of this scenario. This can give further insights into the effect of non-participation of buildings on the clustering solution of different methods. A more robust method should lead to clusters that are viable in different scenarios. The viability criteria are for a cluster are defined based on the optimization metrics except the cluster size, since the cluster size will always reduce if buildings are removed from a cluster. Thus, the following criteria need to be fulfilled for a viable cluster network:

- lower limit of the BCC for feasible implementation of infrastructure
- lower limit of the DFC for the Energy Balance
- upper limit of the d_{avg} for low costs and heat losses

4.4.4. Partition Overlap



Characteristic	Pairs
Together in A and B	4
Apart in A and B	16
Together in A but not B	4
Together in B but not A	4
Total	28

Figure 4.4: Example counting building pairs for partition overlap

If it would be known which buildings do not participate before a clustering method is applied, buildings might be grouped differently. To examine this, first the respective share of buildings participating in the scenario are randomly selected. The clustering methods are then applied to the resulting dataset of buildings to generate new partitions. Afterwards, the overlap between the original partition with 100% participation with the scenario partitions is quantified. Moreover, external validity indices are commonly used to compare the agreement between clustering solutions [55]. One prominent validity index is the Rand Index (RI) which is based on counting object pairs or in this case building pairs [55]. The Rand Index is a number between 0 and 1, with 0 indicating no overlap between the partitions and 1 meaning the partitions are identical. It is calculated by dividing the number of 'agreeing' object pairs with the total number of object pairs [56]. An 'agreeing' object pair between two partitions is defined as two objects that are either in the same cluster in both partitions or in different clusters in both partitions [56]. An example for counting the building pairs to determine the overlap between two partitions is shown in figure 4.4. In this example the number of 'agreeing' building pairs is 20, which is the sum of building pairs that are together in a cluster in partition A and partition B. For the example the partition overlap as measured by the RI is therefore $\frac{20}{28} = 0.7143$. Moreover, a modified version of the Rand Index, which is corrected by chance and called the Adjusted Rand Index (ARI) is commonly used [57] [55]. This correction by chance is applied by substracting the expected Rand Index from both the numerator and denominator in the formula for the Rand Index.

Assuming two partitions $U = \{u_1, u_2, ..., u_R\}$ and $V = \{v_1, v_2, ..., v_S\}$ of a set of n objects (here buildings) are to be compared. Then n_{ij} denotes the number of objects that are in both cluster u_i of partition U and cluster v_j of partition V. Similarly, $n_{i\cdot}$ represents the total number of objects in cluster u_i in partition U and $n_{\cdot j}$ the total number of objects in cluster v_j in partition V. Then the Adjusted Rand Index can be calculated as follows:

$$ARI = \frac{\sum_{i,j}\binom{n_{ij}}{2} - \sum_{i}\binom{n_{i\cdot}}{2}\sum_{j}\binom{n_{\cdot j}}{2}/\binom{n}{2}}{\frac{1}{2}\left[\sum_{i}\binom{n_{i\cdot}}{2} + \sum_{j}\binom{n_{\cdot j}}{2}\right] - \sum_{i}\binom{n_{i\cdot}}{2}\sum_{j}\binom{n_{\cdot j}}{2}/\binom{n}{2}} \tag{4.10}$$

In contrast to the RI, the ARI can have values between -1 and 1 due to the correction with the expected RI [58]. Similar to the RI, a higher ARI indicates a higher overlap between the partitions U and V with 1 suggesting identical partitions. An ARI of 0 indicates an overlap as expected by chance, whereas a negative ARI means that the overlap is lower than what would be expected by chance.

5

Case Study Data

In this chapter, an overview of the data used in this research is provided. Additionally, the input data is analysed, and the preprocessing steps are presented.

5.1. Overview Input Data

The case study within this project focuses on the city centre of Amsterdam. All buildings and flats are available with their location, postcode, living area, roof area and unique ID in a dataset. Within this work, buildings are clustered, so it is assumed that each building is connected as one unit, so individual flats of buildings cannot be added to a cluster. Buildings that lack a building ID, an archetype or unheated buildings, which have neither heating nor cooling demand throughout the year, are removed from the dataset. To enable the calculation of the BCC, data that includes the polygons of the building blocks in the city centre of Amsterdam is used to assign each building a building block based on their location. The underlying data can be seen in figure A.1 where each building block is coloured in a different colour.

For the potential 5GDHC networks, PVT systems are assumed as the only external heat source. Previously, ATES was determined as a suitable seasonal storage for the city centre of Amsterdam [27]. However, since there is no data available for the ATES storage capacity, for the here presented analysis infinite storage is assumed. While this is a strong assumption, determining the ATES size for each cluster in the different scenarios exceeds the scope of this project. Due to this assumption, the analysis in this work will be done based on the yearly data. This also reduces computational time compared to a monthly, daily or hourly analysis. Moreover, in previous work, an approach to sizing ATES systems for each cluster was presented [27]. In this previous approach, the clusters were determined first, and the ATES systems per cluster were sized afterwards in a second step. Thus, a similar procedure could be applied on a promising partition identified within this work.

5.2. Heating and Cooling Data

For each archetype, such as corner row house or flat from a certain time period, hourly heating and cooling demand per m^2 for different types of heating networks is provided by AMS. There is data for 27 archetypes of buildings, that have been defined based on characteristics, such as for what the building is used for. In reality, there are most likely more individual heating and cooling patterns. For this project means that less expected synergies between heating and cooling demands than in reality, since those synergies rely on the difference in demand patterns. Since this research focuses on a low temperature district heating and cooling network, the respective dataset is used. The underlying heating and cooling data for low temperature networks assumes a certain standard of retrofitting for each of the considered buildings, so reduced heat losses. It is important to note that these retrofitting measures have not yet been implemented. To calculate the yearly heating and cooling demand of a building, first yearly heating and cooling demand for each archetype is calculated as the sum of the hourly demand. Then the total floor area is calculated as the sum of the area of all flats in the building. Finally, the total floor area is multiplied with the yearly heating and cooling demand per m^2 to determine the yearly heating and cooling demand of each building.

Figure 5.1 illustrates the number of buildings over the yearly heating and cooling demand used in this

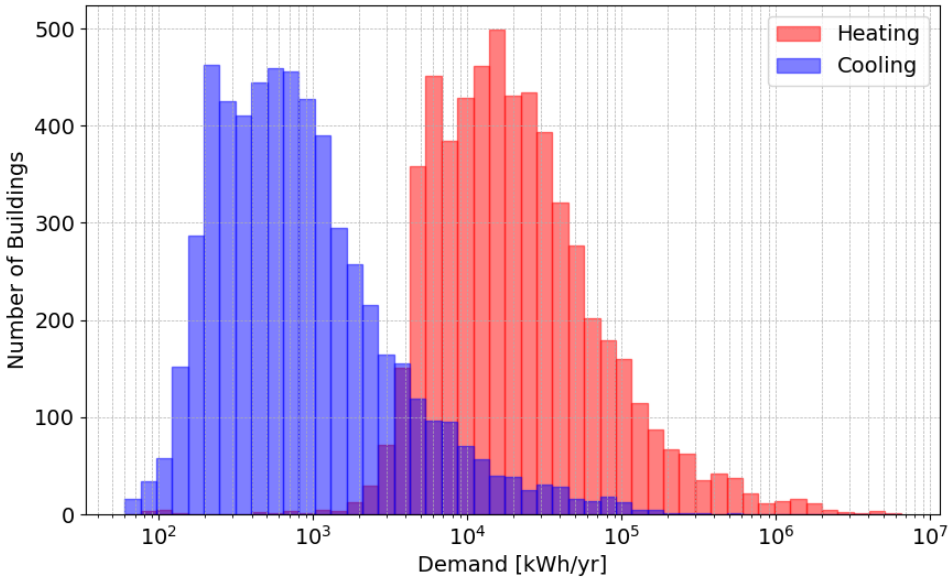


Figure 5.1: Histogram presenting the distribution of the yearly heating and cooling demand

study. The shape of the curves indicate a similar behaviour for heating and cooling of all buildings. Considering most of the buildings are residential buildings, one could conclude that the ratio of cooling to heating demand is similar for most buildings based on the observations. In general, all considered buildings require at least some cooling and heating throughout the year. The yearly heating demand (red) exceeds the yearly cooling demand (blue) by about two orders of magnitude. This suggests that additional heat sources are required for the implementation of a 5GDHC system, independent of available storage. Furthermore, for all buildings considered in this case study, the heating demand exceeds the cooling demand for every hour of the year. The total net heat demand, so the difference between the yearly heating and the yearly cooling demand, in the centre of Amsterdam is 345 GWh/yr.

5.3. PVT Data

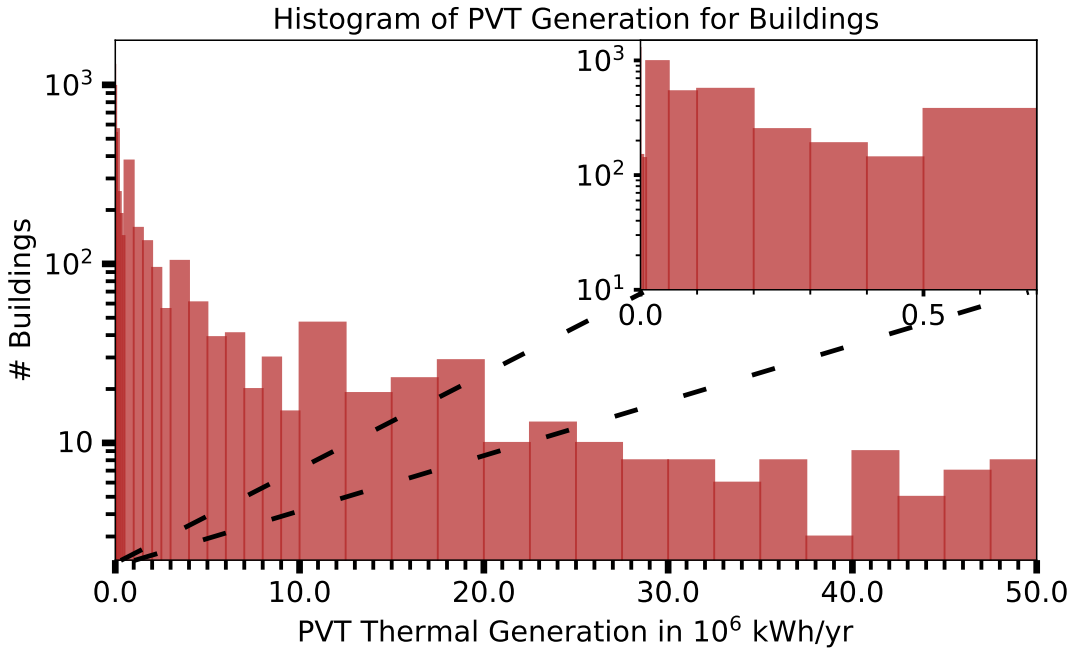


Figure 5.2: Histogram presenting the distribution of the yearly PVT Generation

The data on the thermal generation of the hypothetical Photovoltaic thermal collectors on each building is supplied by the AMS institute and is based on the potential analysis of the external partner PVWorks. Moreover, the data includes values for the yearly thermal generation of each building's PVT system. Since the PVT potential of each building is given in the format thermal energy output in $kWh/(year \cdot module)$, it is multiplied with the respective number of modules for each building. The distribution of the yearly PVT generation over the buildings is illustrated in figure 5.2. The figure shows that yearly PVT generation in the centre of Amsterdam is distributed unevenly between all 5808 considered buildings. For example, generate the five buildings with the highest share of thermal energy generation from PVT 56% of the total thermal generation from PVT. Moreover, the total PVT generation in the centre of Amsterdam according to the data is 72.4 TWh/year. This value appears too high. Early findings from the Simply Positive project, a joint effort of the AMS Institute, TU Delft and PVWorks estimated the PVT potential for the centre of Amsterdam as around 1 PJ $= 277.78$ GWh [39]. To achieve a more realistic analysis, the data is scaled with the factor $4 \cdot 10^{-3}$ under the assumption that the spread of the PVT generation over the buildings is correct, but the total is not. The total scaled PVT generation in the centre of Amsterdam then amounts to 289 GWh/yr. Thus, if all buildings would be connected in a system with infinite storage, the PVT generation would be sufficient to meet 83.8 % of the net heat demand under this assumption.

5.4. Additional Assumptions

For the identification of viable clusters within the robustness assessment, viability criteria are defined for the application of the method to the centre of Amsterdam. These criteria are chosen to investigate the viability of clusters generated by different methods across the considered scenarios. For future applications of this approach, these can be modified based on the requirements of the respective case. For the case study in the centre of Amsterdam, the upper limit of d_{avg} is chosen as 150m, since this is close to the average of the d_{avg} of the different clustering methods in the 100% scenario. This is based on the assumption that a cluster with an average d_avg in the base case is viable. Currently, the availability of additional heat sources despite the assumed PVT systems is not certain. Therefore, the lower limit for the DFC is chosen as 90% to limit the required additional heat supply for those clusters. Focusing on the clusters with a high DFC reduces implementation barriers regarding the additional required heat supply.

Finally, the lower limit for the BCC is chosen at 70 %, which is based on an average cluster for all clustering algorithms in the scenario with 100% participation. To sum up, for this project, the three criteria for a viable cluster are defined as follows:

- $BCC \geq 70$%
- $DFC \geq 90$%
- $d_{avg} \leq 150$m

6

Algorithm Tests

In this chapter, the general usability of the considered algorithms and the size optimization step are tested by applying them to different densely populated European city centres.

6.1. Set Up

To test the general usability of the clustering algorithms, building data within a 2km radius from a central point for each city is extracted using the OSMnx package in python [59]. The chosen cities are Barcelona, Amsterdam, Paris, Rome, Munich, Istanbul and Venice. Since data regarding the building blocks, heating and cooling demand and PVT generation is not available, only the optimization metrics average distance between two buildings in the same cluster d_avg and cluster size CS are used to evaluate the performance of the different methods. Within these tests, both the basic algorithms and the algorithms with the size optimization step are evaluated. Both the energy and block completeness optimization step cannot be tested since the required data is not available in OpenStreetMap.

6.2. European City Test

6.2.1. SLC Results

Table 6.1: SLC Performance Base and optimized for Cluster Size Barcelona

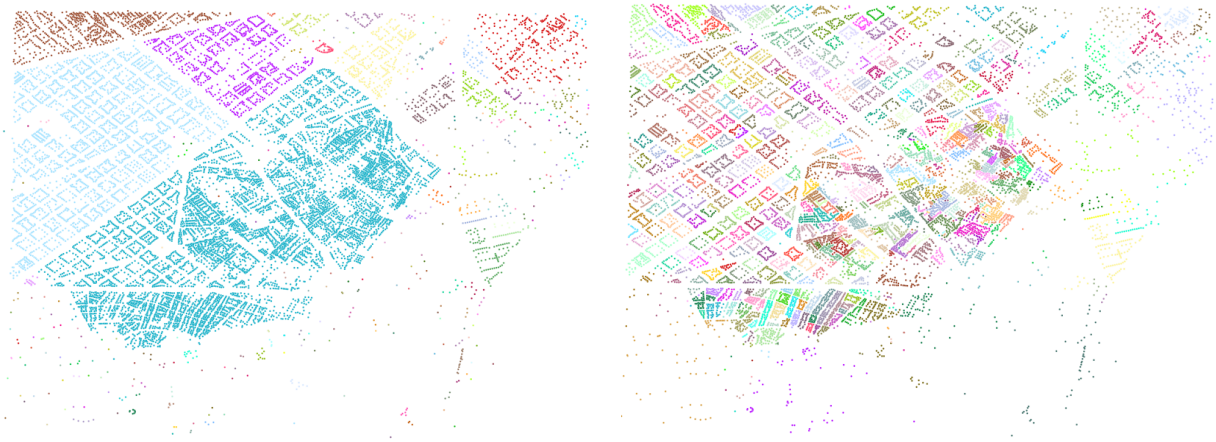
Method	**Avg.** CS	**Std.** CS	**Min** CS	**Max** CS	**Avg.** $d_{avg}[m]$	**Std.** $d_{avg}[m]$	**Spread** $d_{avg}[m]$
Base	129.42	828.94	1	8105	81.63	140.03	1097.49
Size Opt.	40.85	24.65	10	118	85.32	56.30	437.80

Table 6.1 shows the optimization metrics results of the SLC algorithm before and after applying the size optimization step for Barcelona. In this example 13889 buildings are clustered in the centre of Barcelona. The results show that after clustering buildings with the SLC algorithm the majority of them are in the same cluster (see Maximum Cluster size 8105 in 6.1). This also leads to a high variance in cluster size and average distance between buildings. Moreover, the average value of the average distance between two buildings in the same cluster is relatively low, since the SLC leads to 278 clusters in total out of which a few are very large and most are small with less than 10 buildings. Therefore, the minimum of the d_{avg} is 0 or close to 0 if single building clusters are not considered. As seen in the table the spread, so the difference between the maximum and minimum d_{avg}, is large. This indicates that the initial partition by the SLC does not perform well measured by the compactness and size requirements for a good partition. This observation is made for all city centres tested. It can be explained by the chaining effect rooted in the working principle of the SLC. Additionally, each building is represented by its centroid meaning that larger buildings appear further away than they actually are in reality. This leads to larger buildings often being in single building clusters after the initial application of the SLC. Therefore, additional improvement steps such as the size optimization step are required when a method for clustering buildings based on the SLC is used and a partition with small and compact clusters is desired.

An example for the metric results for the SLC in combination with the size optimization step can be seen in

table 6.1. This size optimization step results in 340 clusters that are more evenly sized with an average of 40.85 buildings per cluster, which is close to the ideal cluster size 50, and a significantly reduced standard deviation in cluster size. The spread of cluster sizes is also significantly reduced as seen by the minimum and maximum CS. Additionally, while the average of the average distance between two buildings in a cluster has slightly increased the standard deviation of this metric is significantly reduced after the size optimization step (see table 6.1). Also for the d_{avg} the spread is significantly lower, which indicates a better performance of the partition.

Figure 6.1 shows the inner centre of Barcelona where each building is represented as a point and colored according the cluster it belongs to in the SLC partition. Furthermore, figure 6.1 (a) shows the partition after applying the SLC algorithm and (b) shows the clustered buildings after the partition in (a) is improved through the size optimization step. In figure (a) the chaining effect of the SLC is recognizable for example on the right side of the figure are a few smaller clusters that are shaped like a line. The bigger clusters also surround very small clusters like for example the teal blue cluster surrounds a purple single building cluster. This is caused by the working principle of the SLC in combination with the simplification that buildings are represented with their centroid only, which leads to larger buildings appearing further away from their neighbours than they are in reality. This figure also illustrates the effect of the size optimization step on the partition generated by the SLC. Hence, the size optimization step is important for fulfilling the requirements of small and compact clusters. A similar improvement of the SLC performance through the size optimization step can be observed for all seven tested European city centres. Thus, the size optimization step improves the performance of the SLC algorithm significantly, measured by the defined performance indicators, which are average and standard deviation of the metrics d_{avg} and cluster size.



a) Base b) Size Optimization Step

Figure 6.1: SLC Partition Barcelona

6.2.2. DBSCAN Results

Table 6.2: DBSCAN Optimization Metric Results Barcelona

Approach	**Avg.** CS	**Std.** CS	**Max** CS	**Avg.** d_{avg}[m]	**Std.** d_{avg}[m]	**Max** d_{avg}[m]	**Noise**
Tuning	157.49	603.07	5087	108.20	131.13	800.48	502
K-distance	348.95	1351.47	8125	154.66	213.03	1099.80	280

The input parameters ε and $minPts$ are selected separately for each city data set according to the two aforementioned approaches (see 4.3.1). The main observed drawback of the tuning function is that it is unclear whether the parameters are suitable for the data, since the criteria is the number of clusters. In some cases this leads to less meaningful clusters that are either too fragmented or too merged. Table 6.2 shows the results for clustering 13889 buildings in the center of Barcelona using the two approaches for parameter selection. For Barcelona the tuning function found $minPts = 4$ and $\varepsilon = 46.09$ for the

most promising result measured by the optimization metrics. The k-distance approach found $minPts = 4$ and $\varepsilon = 63.6$. The minimum CS is 4, which is equivalent to the $minPts$ parameter in both cases and therefore excluded from the table. Consequently, the minimum d_{avg} is approximately 10m in both cases and excluded to improve the readability of the table.

As seen in the table 6.2 the k-distance approach creates less noise, so less unclustered buildings, than the tuning function approach. For all tested cities the tuning function leads to more unclustered buildings than the k-distance approach. The reason is that for all cities the k-distance approach found a higher ε parameter, which leads to larger clusters and less noise. On the other hand, the tuning function leads to a lower average distance and a lower variance in the average distance for all cities (example Barcelona in table 6.2). Similarly, the average cluster size and its variance is lower for the tuning function approach compared to the k-distance approach (see table 6.2). This difference in performance is also caused by the larger ϵ parameter leading to larger less compact clusters, in case of the k-distance approach. Thus, while the tuning function performs better measured by the optimization metrics, the k-distance approach excludes less buildings from the partition. Even though for some buildings that are outliers an individual system might be preferable, it is not known if individual alternatives are feasible and affordable for those buildings at this stage in the planning process. Additionally, since the focus of this analysis is densely populated urban city centres, there are in general few outliers compared to more rural areas. Also many larger buildings are labelled as outliers due to building centroids being used for the clustering. Therefore, excluding buildings from the 5GDHC network is undesirable at this stage in the planning process.

In some cities the tuning function leads to unusable results as seen in figure 6.2. Here, the combination of a high $minPts = 8$ and a high variance in density throughout the city leads to the majority of buildings being labelled as noise (light green points). This further illustrates that deciding on the parameters to achieve a certain number of clusters as it is done within the tuning approach is not robust and cannot be used for all cases. Therefore, the k-distance approach appears to be more applicable to different datasets, even though controlling the number of clusters and cluster size is not possible. This observation also has implications for the size optimization step. For the size optimization step, the too large clusters should be split by applying the DBSCAN algorithm on the buildings in each cluster separately. Subsequently, new input parameters need to be determined for each cluster separately. Therefore, the the size optimization step might not be possible if the k-distance approach creates one large cluster and many small clusters every time and the tuning function leads to unusable results for some subsets of buildings. Additionally, the iterative splitting of clusters with DBSCAN is computationally expensive.

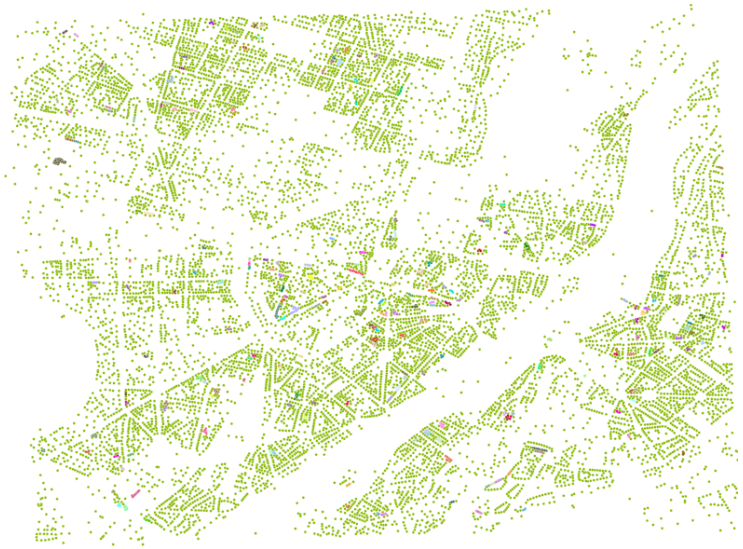


Figure 6.2: Munich: DBSCAN with Parameters based on tuning function

When attempting the size optimization step, the hypothesis that the size optimization step is not possible

in case of the DBSCAN algorithm is confirmed. No suitable input parameters can be found within the tested parameter ranges by applying the tuning function for the majority of clusters for all tested European city centres. Furthermore, the k-distance function does not generate useful sub-clusters since for most almost all points are in one cluster. This also indicates, that the DBSCAN algorithm is not a robust algorithm for finding clusters for a low temperature district heating and cooling network, since a robust algorithm should also be applicable to areas of various sizes and numbers of buildings. Additionally, these tests show the typical complexity of choosing the right parameters for the DBSCAN algorithm as mentioned in literature [46].

Finally, the tests on European city centres show that the DBSCAN is not easy to use or easy to understand for stakeholders and thus does not meet the identified requirements (see 3.2.4). In conclusion the DBSCAN algorithm is not robust and therefore not a good starting point for clustering buildings for district heating and cooling networks in densely populated urban city centres. Therefore, the DBSCAN is excluded from further analysis.

6.2.3. K-Means Results

Table 6.3: K-Means Performance Base and optimized for Cluster Size Barcelona

Method	**Avg.** CS	**Std.** CS	**Min** CS	**Max** CS	**Avg.** $d_{avg}[m]$	**Std.** $d_{avg}[m]$	**Max** $d_{avg}[m]$
Base	50.14	22.76	5	107	97.18	27.04	223.95
Size Opt.	51.06	19.33	10	98	97.63	36.67	378.44

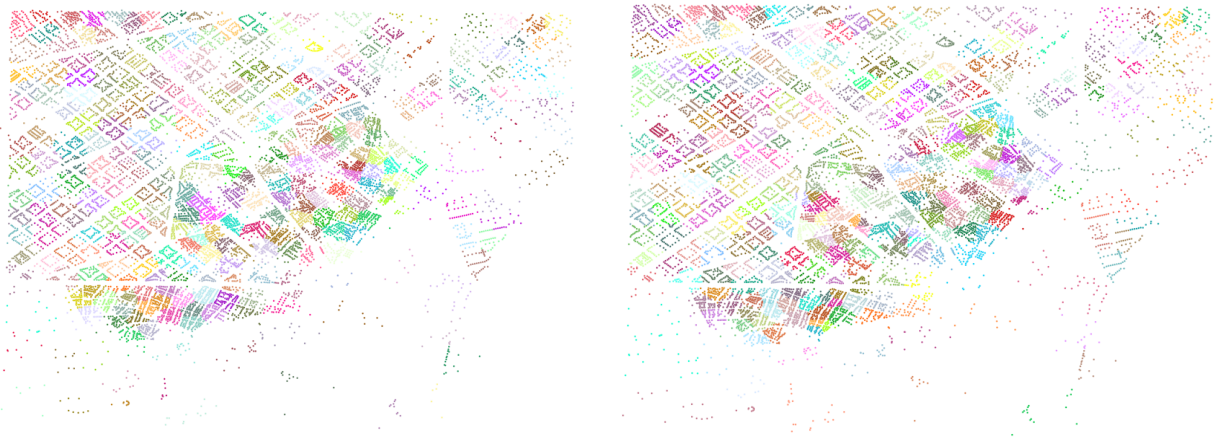
Compared to the two previous algorithms the K-Means without additional steps produces already a partition with small and compact clusters. Example results for the city centre of Barcelona can be seen in table 6.3. Table 6.3 includes the results before and after the size optimization step. Since the K-Means algorithm already produces relatively balanced clusters in terms of size and average distance, the impact of the size optimization step is lower then what is observed for the SLC. When comparing the visualizations of the partition before and after the size optimization step in figure 6.3, the difference is hardly visible. However, the size optimization step still reduces the variance in cluster size (see table 6.3). In terms of the average distance between two buildings in the same cluster, the K-Means performs similarly well with and without the size optimization step (see table 6.3). However, the deviation and the maximum of the average distance between buildings has increased through the optimization step (see table 6.3). For the other tested European city centres the size optimization step has little impact on the average distance between buildings in a cluster. The negative impact observed in the Barcelona case is most likely due to the initial partition generated by the K-Means already having compact clusters. This is also indicated by the minimum and maximum cluster size as they are close to the selected boundaries 10 and 100. In general, the observed impact of the size optimization steps on the average distance metric is negligble, whereas there are small improvements in terms of cluster size variance.

6.2.4. K-Medians Results

Table 6.4: K-Medians Performance Base and optimized for Cluster Size Barcelona

Method	**Avg.** CS	**Std.** CS	**Min** CS	**Max** CS	**Avg.** $d_{avg}[m]$	**Std.** $d_{avg}[m]$	**Max** $d_{avg}[m]$
Base	50.14	27.58	7	192	104.51	60.99	497.57
Size Opt.	47.57	21.24	11	100	101.95	37.33	511.20

Similarly to the K-Means the K-Medians without additional steps already results a good performance measured by the resulting cluster size and average distance d_{avg}. An example of metric results for the application of the K-Medians algorithm on the selected European city centres can be seen in table 6.4. It can be seen that the size optimization step mostly influences the cluster size. Similar to the K-Means results for Barcelona, in case of the K-Medians the maximum distance between buildings slightly increased. However, the deviation and average of the average distance between buildings decreased after the size optimization step of the K-Medians. In general, the size optimization step has little influence on the performance of the K-Medians algorithm except for some reduction of variance of the metric results.



a) Base b) Size Optimization Step

Figure 6.3: K-Means Partition Barcelona

6.2.5. Comparison

The DBSCAN is unsuitable because of the challenges regarding the parameter selection and high share of outliers that affect the usability in the context of low temperature district heating networks. For all cities, the initial partition generated by the DBSCAN includes highly varying cluster sizes similar to the initial SLC partitions. However, the DBSCAN leads to a signficant amount of unclustered buildings (noise) which corresponds to the one building clusters that appear with the SLC algorithm. Still, the amount of unclustered buildings in the DBSCAN partition always exceeded the amount of one building clusters with the SLC. This shows that in addition to not meeting the identified requirements the other clustering algorithms also have more promising first results than the DBSCAN.

Table 6.5: Performance of Clustering Methods including Size Optimization Step Paris

Algorithm	**Avg.** CS	**Std.** CS	**Min** CS	**Max** CS	**Avg.** $d_{avg}[m]$	**Std.** $d_{avg}[m]$	**Max** $d_{avg}[m]$
SLC	48.64	28.18	10	108	88.62	30.47	245.14
K-Means	50.23	15.07	10	100	84.22	15.83	204.90
K-Medians	45.66	21.72	10	100	89.71	29.40	287.10

With the size optimization step, the SLC based method performed similarly to the K-Means and K-Medians algorithm, especially considering the average values of d_{avg} and cluster size. An example of the optimization metric results of the three methods can be seen in table 6.5, where the results for Paris are shown. For Paris the K-Means performs clearly best with the lowest d_{avg} average and standard deviation as well as the least variation in cluster sizes (see table 6.5). Determining the second best method is less clear since K-Medians and SLC with size optimization perform similar measured by d_{avg} and cluster size. Overall, the K-Means algorithm performed best for all cities measured by cluster size and Average Distance between buildings of a cluster. While after the size optimization step all three SLC, K-Means and K-Medians have similarly good values for the average of the performance metrics, the K-Means always has the lowest variance, lowest maximum value and highest minimum value. Thus, the K-Means algorithm appears to be the most promising after the initial algorithm tests on European city centres. However, when comparing figures 6.3 b) and 6.1 b), the partition of the SLC-based method appears to follow the street structure better than the K-Means-based method. This is also observed in the other European city centre tests and will be further analysed within the case study in chapter 7, where the Block Completeness Coefficient (BCC) can be utilized to capture this phenomenon.

7

Case Study Amsterdam

In this chapter, the results for the case study of the city centre in Amsterdam are presented and analysed. For the base case all buildings in the centre, where sufficient data specified in chapter 5 is available, are considered which implicates a participation rate of 100 %.

7.1. Base Algorithms Performance

a) SLC

b) K-Means

c) K-Medians

Figure 7.1: Clustered Buildings with basic Algorithms

Table 7.1 shows the results for the CS and d_{avg} after applying the clustering methods to the dataset of

the centre of Amsterdam. For the SLC algorithm, both the standard deviation of the cluster size CS and the standard deviation of the average distance between buildings d_{avg} is very high compared to the other algorithms. Figure 7.1 shows the partitions generated by the algorithms SLC (a), K-Means (b) and K-Median (c). In the figure, each building is represented by a point and the different colours represent different clusters. Figure 7.1 further illustrates how the SLC leads to some large and many small clusters, which has also been observed in chapter 6. This is due to the SLC connecting the two closest unconnected buildings iteratively until the desired number of clusters is reached. The dark purple cluster in between the large red and pink clusters in figure 7.1 a) illustrates the chaining effect of the SLC described in chapter 4. Even though some buildings in this cluster are far apart they are connected through many links in between. In contrast to the SLC, the K-Means and K-Medians generate clusters which are more similar in size and diameter (see figures 7.1 (b) and (c)). This is also indicated by the comparatively low standard deviation of both the cluster size and d_{avg} in table 7.1. Measured by cluster size and d_{avg} the K-Means performs best, since it leads to the lowest average, standard deviation and spread for both metrics. Additionally, the cluster sizes of the K-Means spread between 12 and 88, which already fulfils the lower and upper limit for the size optimization step. Thus, the size optimization step is not needed for the K-Means. So, the initial results regarding CS and d_{avg} for the case study meet the expectations.

Table 7.1: Cluster Size and Average Distance Results Amsterdam

Algorithm	**Avg.** CS	**Std.** CS	**Min** CS	**Max** CS	**Avg.** $d_{avg}[m]$	**Std.** $d_{avg}[m]$	**Max** $d_{avg}[m]$
SLC	117.12	436.14	2	2576	133.50	191.50	1185.81
K-Means	50.05	17.44	12	88	108.41	22.81	223.83
K-Medians	50.05	29.94	6	146	120.07	53.11	322.16

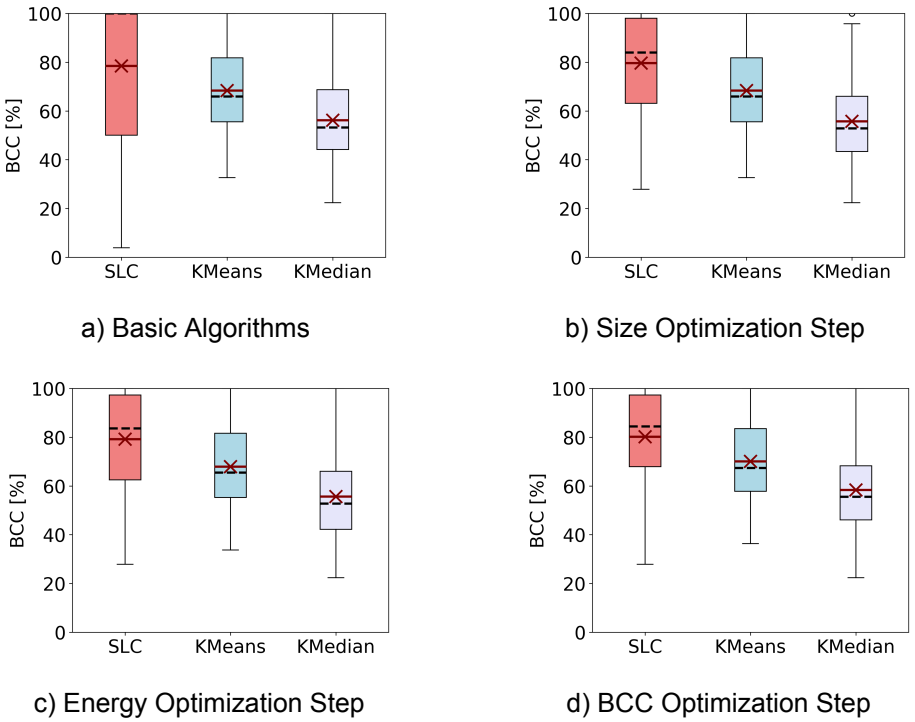
Figure 7.2 a) shows box plots of the BCC of the different Algorithms without any additional optimization steps. In the figure, the red line with the cross is the average and the black dashed line is the median. The average block completeness of the SLC is higher compared to the other algorithms, which indicates that the SLC follows the street structure better on average. This is consistent with the observations in chapter 6. In figures 7.1 (b) and (c) the partitions by the K-Means and K-Medians algorithm also connect buildings to clusters over canals and streets while neighbouring buildings of the same block are in a different cluster. However, looking at the BCC results in figure 7.2 the SLC has the highest standard deviation and spread in regard to the BCC. K-Means has an approximately 10 % lower average BCC but also a lower standard deviation and spread. The K-Medians results in the lowest average BCC and approximately the same standard deviation as the K-Means.

The box plots of the DFC for the initial algorithm application can be seen in figure 7.3 (a). In terms of the DFC the SLC algorithm leads to the highest average but also the highest spread and standard deviation. This observation can be explained by the imbalance in cluster size in combination with the uneven distribution of thermal energy generated by PVT systems throughout the city. For example, if a building with a high thermal energy surplus is connected with more buildings in a large cluster, it can balance the deficit of more buildings improving the cluster DFC. In addition, some buildings with a heat surplus might be in small clusters, which are weighted the same as larger clusters in the calculation of the average DFC. K-Means and K-Medians have approximately the same DFC standard deviation, while K-Means leads to a 2 % higher average and a slightly smaller spread.

7.2. Clustering Methods with Modifications

7.2.1. Cluster Size Optimization Step

As seen previously in chapter 6 the size optimization step significantly reduces the standard deviation and spread of cluster size and the d_{avg} of the SLC partition. Thus, the size optimization step is necessary for the SLC to meet the requirements of small and compact clusters. The resulting d_{avg} of the SLC algorithm with size optimization is visualized in figure 7.4 (a). Compared to before the size optimization step shown in table 7.1, the d_{avg} standard deviation and spread is significantly reduced in figure 7.4 (a). The effect on the BCC can be seen in figure 7.2 (b) in comparison to figure 7.2 (c). While the average is 1.1 % higher the more significant change is in the spread and the standard deviation, which reduced to 19 %. In the



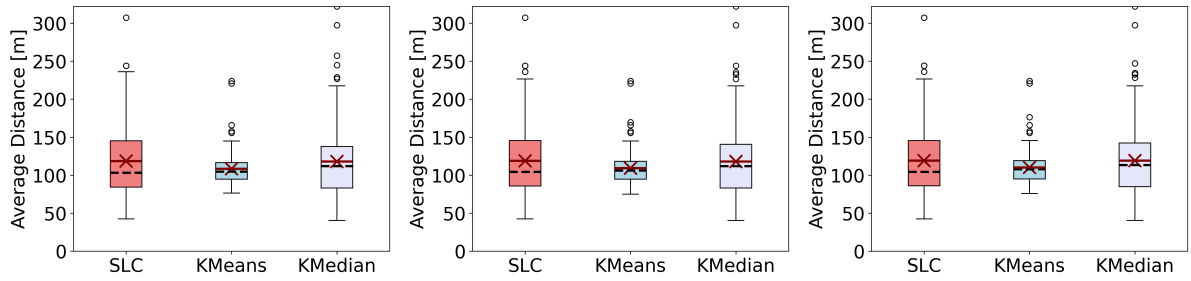
a) Basic Algorithms b) Size Optimization Step

c) Energy Optimization Step d) BCC Optimization Step

Figure 7.2: Block Completeness Coefficient of different Methods

case study, the K-Means already produces a partition with cluster sizes between 12 and 88 buildings. This means that the size optimization step does not change the partition. Therefore, the performance of the K-Means according to the optimization metrics stays the same. For the K-Medians the size optimization step reduces the standard deviation of the cluster size and d_{avg}. The influence of the size optimization step on the DFC and BCC of the K-Medians is negligible.

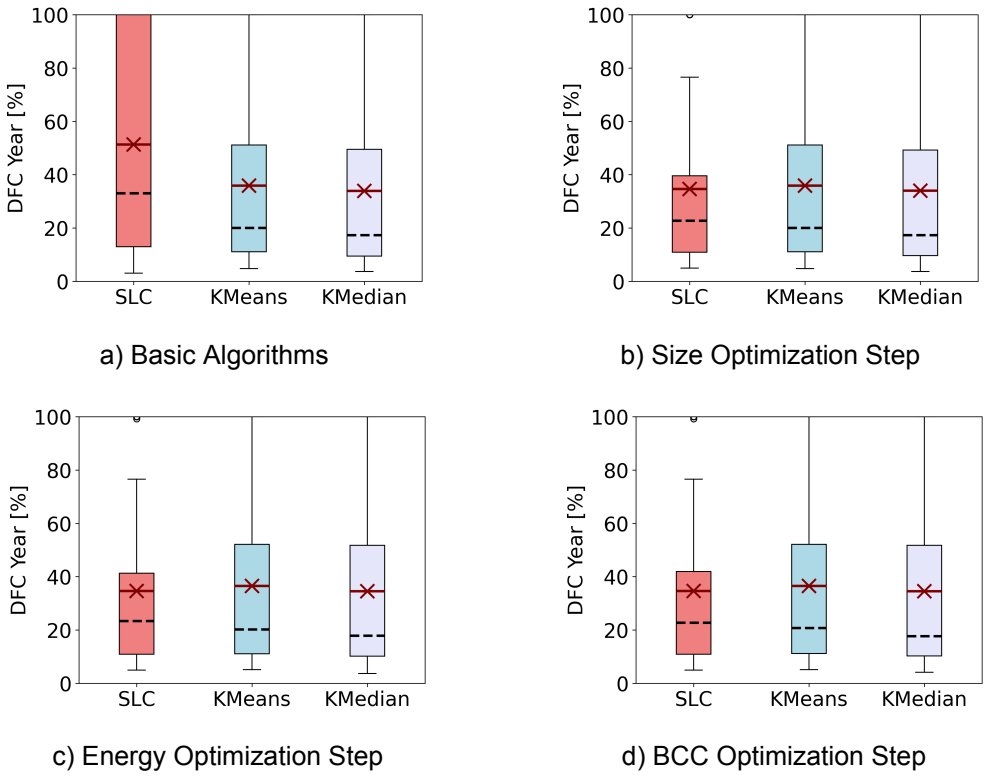
7.2.2. Energy Balance Optimization Step



a) Size Optimization Step b) Energy Optimization Step c) BCC Optimization Step

Figure 7.4: d_{avg} for clustering methods with modifications

First, the effect of different switching radii on the optimization metrics for all clustering methods is investigated. Here, radii between 0, which corresponds to the solution after the size optimization step, and 1000 m are tested in 50 m intervals. It could be observed that the DFC and the d_{avg} increases with the switching radius, while the BCC is reduced for all clustering methods. Thus, there is a Trade-Off when choosing a suitable switching radius between the improvements in the DFC and the negative effect on the other optimization metrics. The graphical visualization of the DFC depending on the radius shows gradual improvements with an increasing radius but with a rather flat slope and no saturation point. Additionally, the d_{avg} and BCC graph indicate no clear point from which onwards there is rapid growth



a) Basic Algorithms

b) Size Optimization Step

c) Energy Optimization Step

d) BCC Optimization Step

Figure 7.3: Demand Fulfilment Coefficient for different Methods

or decline. With no clear indication from the graphical analysis, a 50 m radius is selected for the energy optimization step since the negative effect on the d_{avg} and BCC is small and the percentage improvement in DFC exceeds those negative effects.

The DFC of the partitions by the different clustering methods after both modification steps is visualized in figure 7.3 (c). For all three clustering methods the increase in average DFC is below 1%, the standard deviation is almost the same and the minimal DFC has slightly increased after the energy optimization step. The changes in d_{avg} are negligible for all three methods as seen in figure 7.4 (c). Also the BCC has decreased less than 0.5 % in average for all three methods, which is shown in figure 7.2 (c). In general, it can be observed that the energy optimization step with a selected radius of 50 m has little influence on the performance of the individual algorithms. The final resulting partitions of the case study assuming 100 % participation are visualized in figure 7.5. These partitions are assumed for the 100% scenario which serves as the base case in chapter 8.

7.2.3. Block Completeness Optimization Step

Figure 7.2 d) shows the boxplots of the BCC for the different clustering methods. The results for the K-Means and the K-Medians after the block completeness step appear similar to the results before this modification, which can seen in figure 7.2 c). The main difference is a 3 % increase in the average BCC for both the K-Means (final BCC = 70.0 %) and the K-Medians (final BCC = 58.3 %). For both methods the spread and deviation of the BCC only changes marginally with this modification step. For the SLC the increase in average is 1% (final BCC = 80%) and the standard deviation decreases by approximately 0.8 %. The decrease in standard deviation of the BCC for the SLC is also indicated by the reduced length of the box in figure 7.2 d), which shows the middle 50 % of all clusters are in a smaller range compared to c). The effect of the BCC optimization step on the d_avg results of the different algorithms is negligible for all three clustering methods as seen when comparing figure 7.4 b) and c). Furthermore, the cluster size is unaffected since the BCC optimization step entails exclusively one-for-one switches between neighbouring clusters. The DFC results for the SLC, K-Means and K-Medians methods after the BCC optimization step (figure 7.3 d)) are almost identical to the results before (figure 7.3 c)). This little

impact of the BCC optimization step on the other optimization metrics can be explained by the defined conditions for switching buildings in this step.

a) SLC

b) K-Means

c) K-Medians

Figure 7.5: Buildings clustered by clustering methods after modifications

7.3. Evaluation Case Study

The case study illustrates the behaviour of the considered clustering methods and the effect of the modification steps. The final partitions by the SLC, K-Means and K-Medians method are shown in figure 7.5. In this figure, the difference in cluster shapes between the different algorithms is visible. The partition based on the SLC consists of clusters of varying sizes, that well aligned with the canal and street structure (figure 7.5 a). In contrast to the SLC, the K-Means leads to more round, elliptical clusters, that are similar in size (figure 7.5 b). The K-Medians partition also exhibits elliptical shaped cluster, which appear distorted compared to the K-Means clusters (figure 7.5 c). These distortions are related to the main difference between the K-Means and the K-Medians algorithm, which is that the cluster centroid points are actual data points in case of K-Medians. In this case study the K-Means leads to the most promising partition for a 5GDHC network in the centre of Amsterdam. However, the partition based on Single Linkage Clustering has the best alignment with the street structure, which suggests the feasibility of the connections in the partition. K-Medians exhibits disadvantages in the case study which are also previously identified in chapter 6. This is most likely rooted in the centre of Amsterdam being densely populated, meaning few outliers which could affect the performance of the K-Means.

Until now it is assumed that all buildings in the centre of Amsterdam will participate in the 5GDHC network. In the next step the effect of uncertainty regarding participation is analysed for the case study in the centre of Amsterdam within the robustness assessment in chapter 8.

8

Robustness Assessment

In this chapter the robustness of the clustering methods SLC, K-Means and K-Medians is assessed for the case study in the city centre of Amsterdam based on the minimax regret method. The DBSCAN is excluded since it is determined to be unfit in chapter 6. For this analysis, the partitions after the Block Completeness Optimization Step in chapter 7 are the starting point. For each run of a scenario with a certain participation rate buildings are removed randomly from this partition. Next, the viability of clusters generated by the clustering methods in the different scenarios is analyzed. Finally, it is examined how much the partitions would change if the non-participants would be known prior to the clustering of the buildings.

8.1. Minimax Regret Analysis

In this section the regret for all six performance indicators, namely the average and the standard deviation of the optimization metrics d_{avg}, BCC and DFC, is calculated and analyzed for the three considered clustering methods. For each scenario, a representative run is chosen that performs closest to average for all six performance indicators.

8.1.1. Average and Standard Deviation of the Average Distance

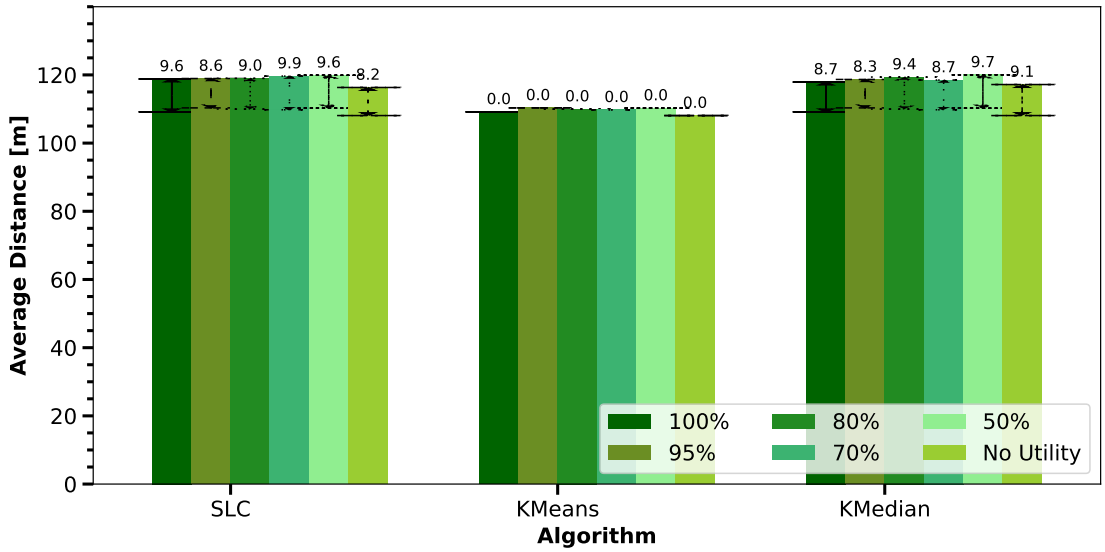


Figure 8.1: Average d_{avg} of different clustering methods in scenarios with regret visualization

First the regret regarding the average and standard deviation of d_{avg} are analyzed. Figure 8.1 shows the average d_{avg} for different methods in the representative run of the considered scenarios, as well as the respective calculated regret. The K-Means based method performs best in all scenarios and thus has a maximum regret of 0m regarding the average distance between buildings. Therefore, the regret of the

other clustering methods is calculated as the difference between their performance and the performance of the K-Means as the best performing method in each scenario. This leads to a maximum regret of 11.8m for the SLC, which occurs in the 50% participation scenario and corresponds to a performance 10.9% worse than the K-Means in that scenario measured by the d_{avg}. Moreover, the maximum regret of the K-Medians also occurs in the 50% participation scenario and is slightly lower at 11.1m, which means the K-Medians performs at most 10.3% worse than the K-Means algorithm measured by the average distance between buildings. In general, the difference between the K-Means as the best performing method and the other clustering methods does not change significantly in the different scenarios (see figure 8.1). Additionally, for all three compared methods the average d_{avg} stays almost constant in the different scenarios.

Table 8.1: Regret regarding the standard deviation of d_{avg} for the clustering methods for the most average run in each scenario

	SLC	K-Means	K-Medians
100% Participation	25.6m	0m	25.5m
95% Participation	24.3m	0m	23.8m
80% Participation	25.7m	0m	26.3m
70% Participation	24.6m	0m	25.6m
50% Participation	23.5m	0m	24.8m
No Utility	25.0m	0m	23.6m

Table 8.1 shows the standard deviation regarding the d_{avg} for representative runs of each scenario. It should be noted that for all three clustering method the d_{avg} standard deviation stays almost constant over the scenarios with variations of less than 2m, which appear also within the different runs of one scenario. Therefore, these variations are most likely not caused by difference in participation rates. The best performing clustering method according to the minimax regret method for the standard deviation of the d_avg is the K-Means with a maximum regret of 0m. The SLC method leads to a maximum regret of 25.7m and the K-Medians has a maximum regret of 26.3m in terms of the standard deviation of the d_{avg}. The observations show that the different participation rates have little impact on the d_{avg} performance of the clusters in the partitions by the three clustering algorithms. A possible explanation is that the city centre of Amsterdam is densely packed with buildings and has little outliers in terms of distance to neighbouring buildings, that the effect of each building on the d_{avg} of a cluster is small. Hence, in terms of d_{avg} of the clusters in the partition all three methods are robust. The working principle of the K-Means that leads to the partition with the most compact cluster in the 100% scenario, which also causes the K-Means to perform best in the other scenarios in terms of compactness.

8.1.2. Average and Standard Deviation of the Demand Fulfilment Coefficient

Table 8.2: Regret regarding the average DFC of the clustering methods for the most average run in each scenario

	SLC	K-Means	K-Medians
100% Participation	2.0%	0%	2.0%
95% Participation	2.5 %	0%	2.0%
80% Participation	1.5 %	0%	2.3%
70% Participation	0.9 %	0%	1.4%
50% Participation	0 %	0.1%	0.8%
No Utility	3.5 %	0%	3.0 %

For the average Demand Fulfilment Coefficient (DFC) the regret of each method in the different scenarios can be seen in table 8.2. Similar to the d_{avg}, the K-Means is the best performing method across all tested

scenarios and therefore has a maximum regret of 0.1 % in terms of the DFC. The second best performing method is the K-Medians with a regret of 3.0 % in terms of the DFC, which occurs in the deterministic scenario with no utility buildings. The SLC has the highest maximum regret with 3.5 % regarding the DFC. Still the SLC outperforms the K-Medians in the 50 %, the 70% and 80% participation scenarios. In the 50% scenario, the SLC is the best performing method. In general, it can be observed that the difference in average DFC between the clustering methods is lower for lower participation rates. Furthermore, the DFC exhibits variations for all clustering methods in the different scenarios, for example the K-Means has a DFC of 36.5 % in the 100 % participation scenario and 30.3 % in the most average run of the 50 % participation scenario. In the different runs of each scenario the DFC also varies due to the strong influence of certain buildings on the energy balance within their cluster. Therefore, the lower the participation rate the higher the variation among the different runs in each scenario.

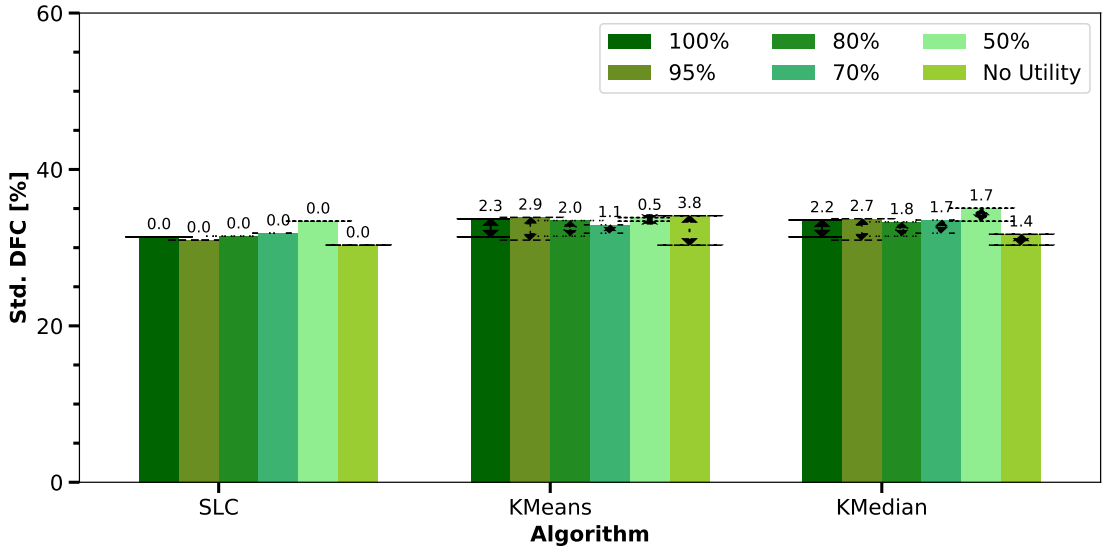


Figure 8.2: Standard deviation of the DFC of different clustering methods in scenarios with regret visualization

Figure 8.2 shows the standard deviation of the DFC for a representative run of each scenario and the respective regret for all three clustering methods. For this performance indicator the SLC has the lowest maximum regret with 0 %, whereas K-Means has a maximum regret of 2.9% and K-Medians has a maximum regret of 2.7%. No general trend for the standard deviation of the DFC in different participation scenarios can be observed. Additionally, the difference between the three clustering methods is small in all scenarios.

8.1.3. Average and Standard Deviation of the Block Completness Coefficient

Figure 8.3 illustrates the average Block Completeness Coefficient (BCC) of the clustering methods in the scenarios and the respective regret of each method in each scenario. Measured by the average BCC the SLC method performs best in all scenarios and therefore has a maximum regret of 0%. The K-Means has a maximum regret of 11.7 % regarding the BCC, which occurs in the 80 % scenario. Furthermore, the maximum regret of the K-Medians is 24.1 %, which occurs in the no utility scenario. Figure 8.3 also indicates a general trend of the average BCC increasing slowly when the share of participants decreases. This can be explained by the calculation of the BCC for which non-participating buildings are not considered as a part of the respective block, since these buildings cannot be included in a meaningful way. This means that in scenarios with less participating buildings fewer buildings are needed for a 'complete' block. Even though the SLC is deemed most robust in terms of the BCC according to the minimax regret method, it is noted that the standard deviation of the BCC is slightly higher for the SLC compared to the other methods. For example the standard deviation of the BCC for the SLC method is 18.5 % in the base case, while for the K-Means the standard deviation of the BCC is 17.3 %.

The regret regarding the standard deviation of the BCC for the different clustering methods is presented

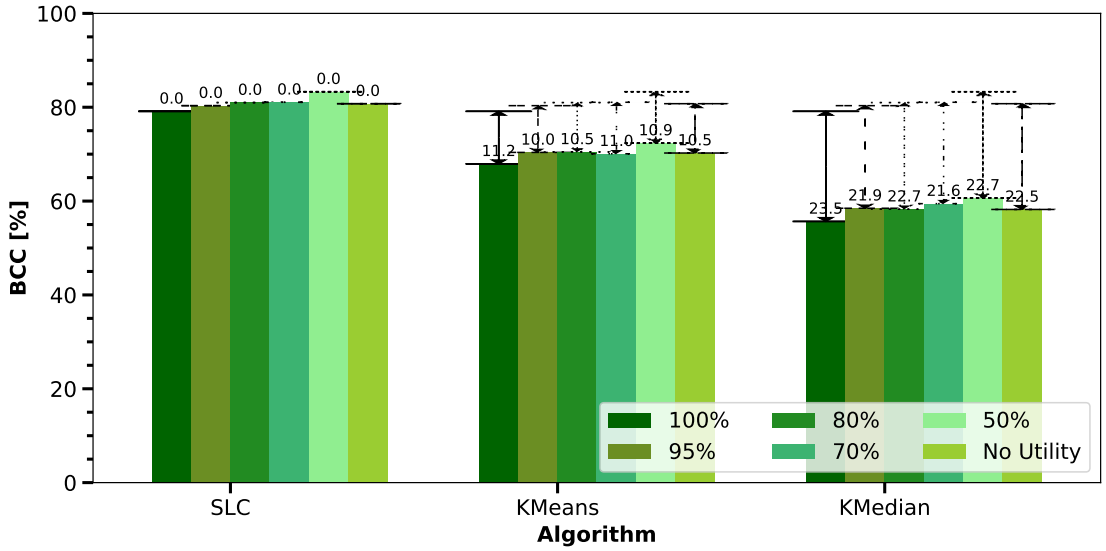


Figure 8.3: Average BCC for different clustering methods across Scenarios with regret visualization

in table 8.3. It can be seen that it varies in the different scenarios which clustering method performs best. The maximum regret regarding this performance indicator is 1.4% for the SLC, 1.8% for the K-Means and 2.1% for the K-Medians. This indicates that the SLC is most robust in terms of the standard deviation of the BCC. However, looking at the different scenarios and scenario runs, all three clustering methods perform similarly and it varies which clustering method performs best with respect to this performance indicator. In general, the differences between the clustering algorithms in terms of the standard deviation of the BCC are small for all scenarios.

Table 8.3: Regret regarding the standard deviation of the BCC for the clustering methods for the most average run in each scenario

	SLC	K-Means	K-Medians
100% Participation	1.2%	0%	0.1%
95% Participation	0.9 %	0.1%	0%
80% Participation	0.7 %	0.3%	0%
70% Participation	1.4 %	1.8%	0%
50% Participation	0.0 %	0.1%	2.1%
No Utility	1.1 %	0%	0.4%

8.1.4. Remarks Minimax Regret and Scenario Results

All in all, the results from the different scenarios and the regret calculation for the minimax regret method are consistent with the performance of the clustering methods in the previous chapters. According to the minimax regret method, the K-Means is the most robust clustering method since it has the lowest regret for 4 out of 6 performance metrics. In general, the different shares of participation have little impact on the partition performance for all three tested clustering methods, which indicates all three methods are robust in their performance. While the minimax regret method gives insights into the overall partition performance, it is also of interest how the viability of clusters is affected by the different scenarios. This is further analyzed in the following section 8.2.

8.2. Cluster Viability Analysis

In this section, the viability of clusters in the different scenarios for all clustering methods is analyzed. A viable cluster in the case study needs to fulfil the three criteria regarding d_{avg}, BCC and DFC as defined in 5.4. If a cluster fulfils these criteria in all runs of a scenario it is viable in this scenario. This also means

that for a specific subset of non-participants, there might be more viable clusters as presented in this analysis. However, this analysis aims at identifying robust clusters for the different clustering methods, which is why the focus is on clusters that are viable in all runs of a scenario. Additionally, the chosen viability criteria affect the partitions of the clustering methods differently. For example is the amount of viable SLC clusters more affected by the d_{avg} criteria, since the SLC leads to more elongated clusters with a higher average distance between buildings. K-Means and K-Medians on the otherhand are more sensitive towards changes in the BCC criterium, since the algorithms lead to more elliptical clusters around a centroid, which means that more building blocks are divided in different clusters.

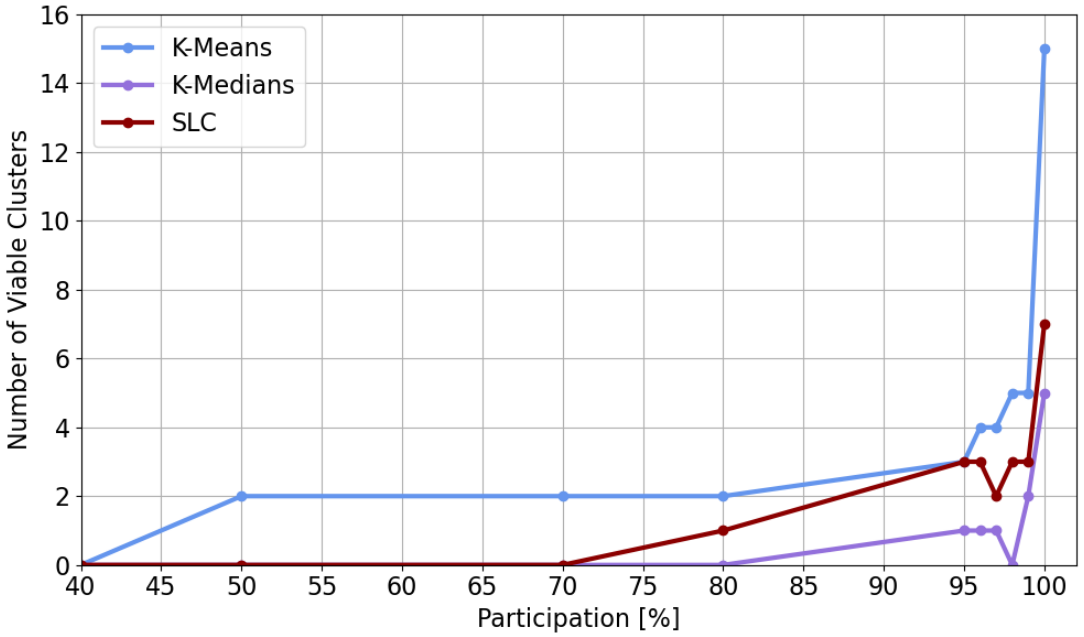


Figure 8.4: Number of viable clusters for different participation rates

Figure 8.4 illustrates the number of viable clusters for each clustering method over different participation rates. For all scenarios, K-Means has the most clusters that fulfil those criteria. Moreover, the number of viable clusters decreases with the participation rate for all clustering methods. Hence, the participation rate impacts the viability of the individual cluster networks in the partition. Here, the question arises if the clustering methods would cluster the buildings differently if the non-participants would be known beforehand. This is analysed in section 8.3. For the graph in figure 8.4 additional scenarios (96%, 97%, 98% and 99%) are included to see what happens between a participation rate of 95% and 100%, since there is a big drop in viable clusters for all three clustering method. It can be observed that between the 100% and 99% participation point the number of viable clusters decreases significantly for all three clustering methods. This occurs because a cluster is only viable in a scenario if it fulfils the criteria in all runs of the scenario. The clusters that depend on a few buildings in terms of for example the required DFC are no longer viable if those buildings do not participate. An additional 40% scenario is also added in the figure 8.4 to examine if the two K-Means cluster stay viable in this scenario. This is not the case, which shows that there is a minimum required participation rate to be viable for all clusters in this analysis. Table 8.4 includes only the viable clusters of the originally defined scenarios and especially also the scenario with no utility buildings. Based on the table it seems that the utility buildings are not required participants for most clusters in the partitions of the three clustering methods, since only for K-Means there are two viable clusters less compared to the case with 100% participation.

The viable clusters in the 100 % participation scenario are depicted in figures 8.5 and 8.6 (a). For all three clustering methods viable clusters appear more concentrated on the right slightly separated neighbourhood in the eastern part of the centrum. One reason for this is that there are a lot of buildings with high PVT potential compared to their own heating demand in that area. Figure 8.6 (a)-(d) shows viable

Table 8.4: Number of Viable Clusters of Clustering Methods in Scenarios

Scenario	SLC	K-Means	K-Medians
100 % Participation	7	15	5
95 % Participation	3	3	1
80 % Participation	1	2	0
70 % Participation	0	2	0
50 % Participation	0	2	0
No Utility	7	15	3

K-Means clusters in different scenarios. Furthermore, the pink and green K-Means clusters in figure 8.6 (c) on the eastern outer part of the centrum are the only clusters that are viable in all runs of all scenarios. Thus, these clusters are the most promising starting clusters for a 5GDHC network in the centre of Amsterdam. A closer look into these clusters also shows that there are a number of buildings with a heat surplus. Moreover, the PVT production is spread over buildings so that the cluster still has a heat surplus even without multiple buildings with a surplus. The building blocks are also smaller in that area compared to the inner part of the centre. In addition, the buildings in those clusters are in close proximity to eachother without outliers. Therefore, the clusters are robust regarding different levels of participation. In clusters that are only viable in case of 100% participation the viability with respect to the energy balance often depends on a few key buildings with a high heat surplus because of their PVT generation. If these important buildings participate the cluster networks are also viable for other levels of participation.

(a) SLC (b) K-Medians

Figure 8.5: Viable Clusters 100 % Scenario

When comparing subfigure a) and d), which are both deterministic scenarios, two clusters that are viable in the scenario with no utility buildings but not viable in the 100 % participation scenario can be observed in figure 8.6. These clusters are the pink cluster in the north of the centrum and the yellow cluster that is close to the central point of the centrum (see figure 8.6 (d)). So, some clusters become viable through the non-participation of certain buildings. One possible reason with respect to the viability criteria is the left out buildings have a heat deficit considering their own heating demand and PVT production and there is not enough surplus heat in the cluster. The non-participation of a building with a heat deficit has a positive effect on the DFC of the cluster. Another possible reason is that the left out buildings are on average further away from the rest of the buildings in the cluster, so without those buildings the d_{avg} decreases. Similarly, leaving out certain buildings could increase the Block Completeness Coefficient of a cluster.

(a) 100 % Participation

(b) 95 % Participation

(c) 50 % Participation

(d) No Utility Buildings

Figure 8.6: Viable K-Means Clusters in different Scenarios

8.3. Partition Overlap

Table 8.5 shows the average overlap of different scenario partitions with the 100% participation partitions for the different clustering methods measured by the Adjusted Rand Index (ARI). The ARI for the three clustering methods with partition improvement steps exhibits little variance throughout the different scenarios. For all scenarios, the SLC based method generates the most similar partition to the partition of the 100 % scenario. Moreover, the K-Means scenario partitions has the second highest overlap compared to the base case, while the K-Medians partitions in the scenarios differ the most from the K-Medians partition in the 100% participation scenario.

Scenario	SLC	K-Means	K-Median
ARI 95%	0.798	0.625	0.409
ARI 80%	0.697	0.593	0.406
ARI 50%	0.743	0.629	0.421
ARI No Utility	0.835	0.663	0.419

Table 8.5: ARI scores of scenario partitions for different clustering methods after modifications

9 Discussion

Within this thesis, a method to compare the performance and robustness of different clustering algorithms has been developed. A key aspect of the presented robustness analysis is the application of the minimax regret method, which is a scenario based approach. To obtain the scenario solutions for each participation rate, buildings are left out in 100 random separate runs. From these 100 runs the average run measured by the optimization metrics, that are also used as performance indicators for the minimax regret method, is selected as the representative run of the scenario. Since it is the average over all optimization metrics and clustering methods, it is not necessarily the most representative result for each individual metric or each clustering method in the scenario. Nevertheless, the performance ranking of the clustering methods for all chosen optimization metrics is the same for all runs of all scenarios. Thus, while the numerical value of the regret might not be representative of all the possible constellations for one scenario, the general indication of which algorithm performs best appears valid. Still, while the 100 runs per scenario give a good indication, 1000 runs per scenario should be evaluated to ensure statistical significance. Unfortunately, 1000 runs per scenario exceed the time constraints of this project. Nevertheless, the sensitivity analysis A.4 in the Appendix shows that there is little variation for the different metrics across scenario runs, which indicates that the selected average run is suitable to represent the scenario results. Additionally, the presented method to analyse the robustness of clustering methods in the context of district heating and cooling networks is easy to use, adaptable and expandable to the specific requirements of different areas and clearly indicates the robustness of the different approaches.

The Euclidean distance is utilized as a distance metric for all considered clustering algorithms and the d_{avg} metric in this study. In reality, not all buildings can be connected via the shortest distance but rather along the street network or bridges. Therefore, the question arises whether utilizing a graph of the street network with buildings as nodes would produce a more adequate result. The main drawback with a street network based approach is an increased computation time. On the other hand, it is unclear how much can be gained to switching to a network based approach, since the calculation of the distance with the street network is strongly correlated with the Euclidean distance. Nevertheless, the utilization of the Euclidean distance is a simplification within the clustering approaches that should be acknowledged. The impact of this simplification also depends on the street layout of the considered area. For city centres with a square or grid-like street layout, the Manhattan metric should be considered as an alternative for a better alignment with the street network without significant computational drawbacks.

In the here presented methodology, buildings are clustered based on their centroid. This leads to smaller buildings appearing closer to neighbouring buildings than larger buildings. Since buildings are not connected strictly from centroid to centroid in reality, one could consider developing an approach where buildings are represented with more points. This might prevent larger buildings from having an unrealistic negative impact on compactness measures or wrongly appearing as outliers.

There are also significant limitations considering the input data of the case study used in this work. First, the heating and cooling demand data of all considered buildings is a forecast based on the same profile scaled according to building archetype and floor area. Even if the majority of buildings are residential,

shops and buildings used for other purposes should show different patterns. Hence, there is no real opportunity to match heating and cooling demands, which limits the energy optimization step. Furthermore, according to the input data, the heating demand always exceeds the cooling demand for all buildings at all times, which appears unlikely. In fact, the AMS institute will be working on an improved forecast for the cooling data.

The energy optimization step depends only on the difference between the thermal generation of the PVT and the net heat demand of the buildings. However, the PVT data appears flawed as the total thermal generation exceeded initial predictions and the total heating demand by a factor greater than 1000. Additionally, the five buildings with the highest thermal generation through PVT generate 56% of the total thermal PVT generation in the city centre of Amsterdam. This suggests that not only the order of magnitude, but also the distribution of the PVT data across buildings, is not correct. Looking at the roof area of the buildings, an approximately linear increase of PVT generation with the roof area is expected, with some deviation due to shadows, roof angle and orientation. However, in the data at hand the PVT generation almost exponentially increases with the roof area, which indicates a calculation error. If this is indeed the error in the PVT data, the correction factor $4 \cdot 10^{-3}$ would have reduced the generation of smaller producers too much and the generation of bigger producers not enough. A more realistic distribution of the PVT data would entail an increased generation of buildings with little and a lower generation of buildings with a lot of PVT. This would lead to a more even spread of the PVT generation over the buildings, which could improve the average and reduce the spread of the demand fulfilment coefficient of the individual cluster networks. A more evenly distributed PVT generation could also increase the robustness of the cluster networks, since it reduces the dependence on individual buildings with a high PVT production. Thus, the PVT data requires further investigation and presents a significant limitation, especially considering the values of the Demand Fulfilment Coefficient. The unbalanced PVT data also affects the energy optimization step and explains the little switches possible for the different radii. More accurate PVT generation data and cooling data could increase the meaningfulness of the energy optimization step, since more switches may be possible. This could also lead to a more clear choice of the switching radius.

Then there is the lack of data regarding possible ATES systems, leading to the strong assumption of infinite storage. This simplification and the fact that the PVT data is only available on a yearly scale, resulted in the yearly demand and supply being used during the energy optimization step. For a future application of this method, an hourly scale could provide more accuracy regarding the energy balance and DFC values. Also, the availability of data on the ATES potential throughout the area could enable more accurate results and a more effective energy optimization step.

Finally, even though the investigated clustering algorithms and modifications are tested on limited data, the results clearly indicate advantages of the K-Means in the context of low temperature district heating and cooling grids. The K-Means consistently shows the best performance by the chosen optimization metrics in the initial algorithm tests, the case study and the different scenarios. One of the reasons the K-Means performs so well in the case study and densely populated city centres in general is that there are typically little outliers in these areas. The buildings are mostly evenly distributed and in proximity to one another, which means that the advantage regarding the robustness against outliers of the K-Medians does not apply. For areas with almost no outliers, the fact that the K-Medians uses a data point as a centroid can be a disadvantage, since it leads to more distorted elliptical clusters with larger average distances and variations regarding size and compactness compared to K-Means. For other areas with more outliers, the K-Medians might outperform the K-Means and ensure more evenly sized and compact clusters. In general, the SLC has the advantage in terms of block completeness and therefore alignment with the street structure. In cases where the spaces in between blocks are obstacles that are hardly possible to connect buildings over, the SLC will deliver the most suitable partition. The energy balance of the resulting partition of the different clustering algorithms strongly depends on how the heating and cooling demands and additional heat sources are spread throughout the city.

10

Conclusion

This project has investigated the knowledge gaps regarding the comparison of clustering algorithms and the effect of uncertainty regarding participation for dividing buildings into clusters for a 5GDHC network in densely populated urban city centres. This assessment has been carried out in the context of advancing the heating transition to fulfil the European Climate Goals. This work has focused specifically on a case study in the city centre of Amsterdam, where the municipality wants to phase out natural gas. This ambition faces significant challenges due to the strict spatial limitations and the current building stock being poorly insulated.

10.1. Research Questions

To address the knowledge gaps, the following research question, derived in Chapter 2, serves as guidance through this work.

Research Question

What is the best clustering method in terms of robustness for a fifth generation district heating and cooling (5GDHC) network in densely populated areas?

Within this research four different clustering methods, namely Single Linkage Clustering (SLC), Density-Based Spatial Clustering of Applications with Noise (DBSCAN), K-Means and K-Medians, are investigated regarding their performance and robustness. For this purpose, a methodology for comparing the performance and assessing the robustness of clustering methods for 5GDHC networks for densely populated urban areas has been developed.

A good partition that enables a modular approach is characterized by small and compact clusters for quick implementation, short pipe lengths and lower initial investment costs. Therefore, the average distance between two buildings in a cluster (d_{avg}) and the cluster size were utilized to measure the performance of the clustering methods. New energy infrastructure is predominantly implemented following the street layout. Hence, another optimization metric to measure the performance is the Block Completeness Coefficient (BCC). Furthermore, each cluster network in a partition should have a good energy balance within each cluster, the Demand Fulfilment Coefficient (DFC) is part of the optimization metrics. Since the investigated clustering algorithms, all cluster buildings based on their Euclidean distance to each other their performance in terms of the optimization metrics can be improved by the developed size and energy balance optimization step.

Initial tests on various densely populated city centres showed that the DBSCAN is unsuitable as a universal approach to clustering buildings for 5GDHC networks in densely populated urban areas. The DBSCAN is difficult to use due to the uncertainty regarding the selection of suitable parameters. In addition, the initially generated partitions require significant improvement steps that were not possible based on DBSCAN. The most promising algorithm after the application on European city centre is the K-Means algorithm. The K-Means based method consistently performed best for all city centres and also had

the lowest variance in the optimization metrics. For the Single Linkage Clustering the size optimization step improved the performance of the partition significantly for both the algorithm tests on European city centres and the application on the case study data. Within the case study, the SLC based method is advantageous regarding the BCC, which indicates a better alignment with the street structure. For all other optimization metrics, the K-Means performed best in the case study.

The scenario-based robustness assessment showed that SLC, K-Means and K-Medians are robust in their performance measured by the optimization metrics for different participation scenarios. However, due to the best performance considering the average distance between two buildings, the Demand Fulfilment Coefficient and the lowest variance in performance across all metrics, the K-Means is identified as the most robust method. The K-Means partition also leads to the highest amount of viable clusters across different scenarios and thereby delivered promising starting clusters for the implementation of a district heating and cooling network. The assessment also indicates shortcomings of the K-Means regarding the alignment with the street layout. Moreover, the scenario-based assessment shows that the participation rate impacts the viability of the cluster networks for all clustering methods.

Despite the simplifications made and concerns regarding the data quality, the developed method delivers a good indication of the robustness and comparative performance of different clustering algorithms. The methodology is easy to use and adaptable to for example additional sources and storage.

10.2. Scientific Contribution

This research contributes to the topics of fifth generation district heating and cooling networks and the comparison of clustering methods for specific datasets. Firstly, it delivers valuable insights regarding the effect of participation uncertainty and the general performance of clustering solutions, which adds to the literature regarding the design of 5GDHC networks. Secondly, it adresses the knowledge gap regarding the comparison of clustering methods for the application on datasets specific to urban energy planning. Within the project the advantages and drawbacks of the different clustering algorithms for their application on urban datasets are identified, which can be a valuable input for future research. Finally, this research suggests several additional steps to improve the clustering algorithms for urban energy planning.

10.3. Societal Contribution and Policy Recommendation

The findings within this work contribute to the planning of novel heating solutions for densely populated urban centres, which are a great challenge within the heat transition. This supports the efforts to reach national and international climate goals. By adressing participation uncertainty in the planning of fifth generation district heating and cooling networks, implementation barriers are reduced. In addition, within this work alternatives to current planning tools for 5GDHC networks are suggested.

In the Netherlands, the current state-of-the-art tool for planning urban energy systems VESTA Mais exhibits drawbacks for planning low temperature district heating and cooling networks. These drawbacks include that the solution space is limited based on a profitability assessment, which could lead to promising solutions being overlooked. Vesta Mais also leaves out buildings based on this profitability assessment, regardless of whether alternative solutions are available or affordable for the respective buildings. Here, the clustering methods selected within this work have the advantage that all possible solutions are considered, and all buildings are included. Moreover, in Vesta Mais the demand and supply are not utilized in tandem to balance buildings with each other or additional heat supply sources. In contrast to this tool, the clustering methods in this research consider demand and supply simultaneously within a designated step to improve the energy balance within clusters. Therefore, the presented clustering approaches, especially K-Means as the best performing and most robust method, are promising alternatives to be used for planning 5GDHC networks.

To implement low temperature district heating and cooling networks successfully, all stakeholders should be involved in the process. A successful involvement of stakeholders such as building owners can build

trust in the solution and positively affect the participation rate. The developed methodology for assessing the performance and robustness of clustering methods can also be utilized by for example the municipality to illustrate how a solution is selected or to compare different approaches. Thus, the proposed method can support engaging stakeholders in the planning process.

For the case study in the centre of Amsterdam, the most promising partition generated by the K-Means based approach was identified. The viability assessment indicated promising K-Means starting cluster networks. Before implementing a 5GDHC network in these areas, the input data should be revised and a technical feasibility study should be performed. This should also include finding a suitable location and size for ATES systems, as well as possible network layouts in each cluster. One important assumption that needs to be considered is that the buildings should be retrofitted for the district heating and cooling network. Within the promising starting clusters, one could connect and retrofit buildings simultaneously for the implementation of the network. The starting point could be already retrofitted buildings or buildings that require less time or lower investments to be retrofitted. Then, buildings can be connected step-by-step while retrofit measures are applied to all buildings willing to participate in the cluster.

The assessment also revealed that the viability of the cluster networks depends on the participation in the network. For the clusters that are viable in the 100% scenario but not in scenarios with lower participation, there are key buildings that need to participate for the cluster to be viable. In most cases, the key buildings are characterized by a thermal generation through PVT that exceeds their own heating demand. For these buildings, independent systems with their own storage are a promising alternative. Therefore, the municipality of Amsterdam should focus on convincing those buildings to participate to advance the implementation of low temperature district heating and cooling networks in Amsterdam.

11

Recommendations

This chapter provides a brief overview of recommendations for the future continuation of this research project.

The assessment in this thesis suggested the K-Means as the most robust and best performing method for clustering buildings for 5GDHC networks. One drawback of the K-Means that was identified is the misalignment with the street network. But certain connections might be not feasible due to obstacles such as the canals in Amsterdam. Here, the feasibility of the resulting cluster networks with respect to the implementation of necessary infrastructure could be investigated. This could give further insight into the severity of this drawback, which will most likely depend on the street structure and spatial limitations for new infrastructure. Moreover, the energy optimization step could be enhanced by integrating other waste heat sources, more reliable cooling data and data on possible ATES locations and capacities. This could also greatly improve the cluster viability analysis within this work. To aid the case in the centre of Amsterdam, data of the respective subsurface focusing on identifying obstacles for new infrastructure and possible locations for ATES systems could be collected and analysed.

One of the key takeaways of this study is that the participation or non-participation of buildings can affect the viability of the respective cluster network. Therefore, an interesting focus of future work could be the influence of social, economical and other factors on the participation of building owners in a low temperature district heating and cooling network. For example, a Threshold model could be utilized to assess the influence of social factors, such as whether neighbours participate. These insights in the participation decision of building owners might create more certainty and help to identify possible policy incentives to increase participation in the network.

The focus of the robustness of the methods in this research is the uncertainty regarding participation. However, there are other sources of uncertainty that could be addressed in future research, such as the availability and performance of additional heat sources in the network. Here, the effect of uncertainty regarding available renewable energy sources such as PVT and aqua thermal applications could be analysed in a scenario-based approach. The implementation of a low temperature district heating and cooling grid also impacts the electricity grid due to the required heat pumps. This interconnection with the electricity grid could be investigated further to identify associated risks and implementation barriers and develop mitigation strategies.

References

[1] European Commission. *The European Green Deal*. Accessed: 2024-09-10, available at https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal_en. 2019.

[2] Energie Beheer Nederland. *EBN Infographic 2024*. Accessed: 2024-09-10, available at https://www.energie-nederland.nl/en/topics/energy-transition/facts-figures/. 2024.

[3] Emma Koster et al. *The Natural Gas Phase-Out in the Netherlands*. Tech. rep. Commissioned by the European Climate Foundation. CE Delft, Feb. 2022. URL: https://ce.nl/wp-content/uploads/2022/04/CE_Delft_210381_The_natural_gas_phase-out_in_the_Netherlands_DEF.pdf.

[4] C. Duin et al. *Transitievisie warmte amsterdam*. Accessed: 2023-10-10. 2020. URL: https://openresearch.amsterdam/nl/page/63522/transitievisie-warmte-amsterdam.

[5] International Energy Agency. *Natural Gas - The Netherlands*. Accessed: 2025-04-07. URL: https://www.iea.org/countries/the-netherlands/natural-gas.

[6] Mathilde Fajardy et al. "An overview of the electrification of residential and commercial heating and cooling and prospects for decarbonisation". In: *Univeristy of Cambridge* (2020).

[7] AMS Institute. *Sustainable Heating Solutions for Amsterdam's Historic Buildings*. https://www.ams-institute.org/urban-challenges/urban-energy/sustainable-heating-solutions-for-amsterdams-historic-buildings/. Accessed: 2024-09-26. 2021.

[8] Jan H Miedema et al. "Opportunities and barriers for biomass gasification for green gas in the dutch residential sector". In: *Energies* 11.11 (2018), p. 2969.

[9] Kristian Gjoka et al. "A comparative life cycle assessment of fifth-generation district heating and cooling systems". In: *Energy and Buildings* (2024), p. 114776.

[10] René Verhoeven et al. "Minewater 2.0 project in Heerlen the Netherlands: transformation of a geothermal mine water pilot project into a full scale hybrid sustainable energy infrastructure for heating and cooling". In: *Energy Procedia* 46 (2014), pp. 58–67.

[11] Gemeente Amsterdam. " Nieuw Amsterdams Klimaat - Routekaart Amsterdam Klimaatneutraal 2050 ". In: *Gemeente Amsterdam—Ruimte & Duurzaamheid: Amsterdam, The Netherlands* (2020). Accessed: 2024-09-10, available at https://www.amsterdam.nl/bestuur-organisatie/volg-beleid/duurzaamheid/klimaatneutraal/.

[12] Henrik Lund et al. "4th Generation District Heating (4GDH): Integrating smart thermal grids into future sustainable energy systems". In: *energy* 68 (2014), pp. 1–11.

[13] B Roossien et al. *KoWaNet: Technical Design Framework for Heating and Cooling Networks*. https://shorturl.at/8FJQt. Accessed: October 8, 2024. 2020.

[14] Henrik Lund et al. "Perspectives on fourth and fifth generation district heating". In: *Energy* 227 (2021), p. 120520.

[15] Simone Buffa et al. "5th generation district heating and cooling systems: A review of existing cases in Europe". In: *Renewable and Sustainable Energy Reviews* 104 (2019), pp. 504–522.

[16] Stef Boesten et al. "5th generation district heating and cooling systems as a solution for renewable urban thermal energy supply". In: *Advances in Geosciences* 49 (2019), pp. 129–136.

[17] Kristian Gjoka et al. "Fifth-generation district heating and cooling systems: A review of recent advancements and implementation barriers". In: *Renewable and Sustainable Energy Reviews* 171 (2023), p. 112997.

[18] Akos Revesz et al. “Developing novel 5th generation district energy networks”. In: *Energy* 201 (2020), p. 117389.

[19] Anna Volkova et al. “5th generation district heating and cooling (5GDHC) implementation potential in urban areas with existing district heating systems”. In: *Energy Reports* 8 (2022), pp. 10037–10047.

[20] Marco Wirtz et al. “Survey of 53 fifth-generation district heating and cooling (5GDHC) networks in Germany”. In: *Energy Technology* 10.11 (2022), p. 2200749.

[21] Marco Wirtz et al. “5th generation district heating and cooling network planning: A Dantzig–Wolfe decomposition approach”. In: *Energy Conversion and Management* 276 (2023), p. 116593.

[22] Kaisa Kontu et al. “Multicriteria evaluation of heating choices for a new sustainable residential area”. In: *Energy and Buildings* 93 (2015), pp. 169–179.

[23] Pavel Berkhin. “A survey of clustering data mining techniques”. In: *Grouping multidimensional data: Recent advances in clustering*. Springer, 2006, pp. 25–71.

[24] Mahamed GH Omran et al. “An overview of clustering methods”. In: *Intelligent Data Analysis* 11.6 (2007), pp. 583–605.

[25] Mayra Z Rodriguez et al. “Clustering algorithms: A comparative approach”. In: *PloS one* 14.1 (2019), e0210236.

[26] Sarah van Burk. “Creating Clusters in a 5th Generation District Heating and Cooling Network An alternative to a neighbourhood-based approach”. https://resolver.tudelft.nl/uuid:059d58aa-9163-444a-94e2-a76e48626683. MA thesis. Delft University of Technology, 2023.

[27] Thom van den Akerboom. “Clustering buildings with ATES systems to improve Amsterdam’s Fifth Generation Heating and Cooling Network”. http://resolver.tudelft.nl/uuid:7bbaa640-b659-43c5-b031-a902376c91f6. MA thesis. Delft University of Technology, 2024.

[28] Jan Schiefelbein et al. “Clustering of Buildings within City Districts to reduce Runtime for Energy System Placement Optimization”. In: ().

[29] Parastoo Pilehforooshha et al. “A local adaptive density-based algorithm for clustering polygonal buildings in urban block polygons”. In: *Geocarto International* 35.2 (2020), pp. 141–167.

[30] J. Loustau et al. “Clustering and typification of urban districts for energy system modelling”. In: *Proceedings of ECOS* (2023). Accessed: 2025-01-30. DOI: 10.52202/069564-0288. URL: https://www.proceedings.com/069564-0288.html.

[31] Marek J Chalupnik et al. “Comparison of ilities for protection against uncertainty in system design”. In: *Journal of Engineering Design* 24.12 (2013), pp. 814–829.

[32] Maeva Dang et al. *Reduce, Reuse and Produce Potentials for De Waag on Nieuwmarkt*. Tech. rep. Accessed: 2025-01-23. AMS Institute and Delft University of Technology, 2023. URL: https://openresearch.amsterdam/image/2023/2/1/recommendations_de_waag_tud_ams.pdf.

[33] Planbureau voor de Leefomgeving. “FUNCTIONEEL ONTWERP VESTA MAIS 5.0”. In: (2021).

[34] Naud Loomans et al. “Exploring trade-offs: A decision-support tool for local energy system planning”. In: *Applied Energy* 369 (2024), p. 123527.

[35] Ed Nijpels. *Ontwerp van het Klimaatakkoord*. https://www.klimaatakkoord.nl/documenten/publicaties/2018/12/21/ontwerp-klimaatakkoord. Accessed: 2025-04-06. 2018.

[36] AMS Institute. *Urban Energy*. Accessed: 2025-04-01. 2024. URL: https://www.ams-institute.org/urban-challenges/urban-energy/.

[37] Heat Roadmap Europe. *Peta4 - Heat Roadmap Europe*. Accessed: 2025-03-01. 2018. URL: https://heatroadmap.eu/peta4/.

[38] STOWA. *Juridisch Kader Aquathermie 2019; Speelruimte voor de Praktijk*. Accessed: 2025-03-29. 2019. URL: https://www.stowa.nl/sites/default/files/assets/PUBLICATIES/Publicaties%202019/STOWA%202019-28%20Aquathermie%20juridisch.pdf.

[39] AMS Institute and TU Delft and PV Works. “Amsterdam could meet nearly half its electricity needs by better utilizing its rooftops”. In: *Simply Positive Project* (2025). Accessed: 2025-03-01. URL: https://www.tudelft.nl/en/tu-delft-urban-energy/news-events/amsterdam-could-meet-nearly-half-its-electricity-needs-by-better-utilizing-its-rooftops.

[40] Martin Bloemendal et al. “The effect of a density gradient in groundwater on ATES system efficiency and subsurface space use”. In: *Advances in Geosciences* 45 (2018), pp. 85–103.

[41] Xiang Li et al. “Techno-economic analysis of fifth-generation district heating and cooling combined with seasonal borehole thermal energy storage”. In: *Energy* 285 (2023), p. 129382.

[42] J. Schilling et al. *Handreiking Warmtenetten*. Accessed: 2025-02-02. 2020. URL: https://www.aedes.nl/verduurzaming/handreiking-warmtenetten-warmtenetten.

[43] Francis P Boscoe et al. “A nationwide comparison of driving distance versus straight-line distance to hospitals”. In: *The Professional Geographer* 64.2 (2012), pp. 188–196.

[44] Marco Wirtz et al. “Quantifying demand balancing in bidirectional low temperature networks”. In: *Energy and Buildings* 224 (2020), p. 110245.

[45] Nationaal Programma Lokale Warmtetransitie. *Handreiking mini- en kleinschalige warmtenetten*. Accessed: 2025-02-01. 2024. URL: https://www.nplw.nl/uploads/files/Warmtenet/NPLW-Handreiking-mini-en-kleinschalige-warmtenetten-wcag.pdf.

[46] Artur Starczewski et al. “A new method for automatic determining of the DBSCAN parameters”. In: *Journal of Artificial Intelligence and Soft Computing Research* 10 (2020).

[47] Peter Wittek. “Unsupervised learning”. In: *Quantum Machine Learning* (2014), pp. 57–62.

[48] F. Pedregosa et al. *Scikit-learn: Machine Learning in Python – Clustering Module*. Accessed: 2025-03-02. 2024. URL: https://scikit-learn.org/stable/modules/clustering.html.

[49] Angur Mahmud Jarman. “Hierarchical cluster analysis: Comparison of single linkage, complete linkage, average linkage and centroid linkage method”. In: *Georgia Southern University* 29 (2020), p. 32.

[50] Xuefeng Gao et al. “A new methodology for building energy performance benchmarking: An approach based on intelligent clustering algorithm”. In: *Energy and Buildings* 84 (2014), pp. 607–616.

[51] Martin Helm. *Use This Clustering Method If You Have Many Outliers*. Accessed: 2025-03-05. 2021. URL: https://towardsdatascience.com/use-this-clustering-method-if-you-have-many-outliers-5c99b4cd380d.

[52] Rajesh Kotireddy et al. “Integrating robustness indicators into multi-objective optimization to find robust optimal low-energy building designs”. In: *Journal of Building Performance Simulation* 12.5 (2019), pp. 546–565.

[53] AM Rysanek et al. “Optimum building energy retrofits under technical and economic uncertainty”. In: *Energy and Buildings* 57 (2013), pp. 324–337.

[54] Soheil Alavirad et al. “Future-Proof Energy-Retrofit strategy for an existing Dutch neighbourhood”. In: *Energy and Buildings* 260 (2022), p. 111914.

[55] Matthijs J Warrens et al. “Understanding the adjusted rand index and other partition comparison indices based on counting object pairs”. In: *Journal of Classification* 39.3 (2022), pp. 487–509.

[56] Lawrence Hubert et al. “Comparing partitions”. In: *Journal of classification* 2 (1985), pp. 193–218.

[57] José E Chacón et al. “Minimum adjusted Rand index for two clusterings of a given size”. In: *Advances in Data Analysis and Classification* 17.1 (2023), pp. 125–133.

[58] Michaela Hoffman et al. “A note on using the adjusted Rand index for link prediction in networks”. In: *Social networks* 42 (2015), pp. 72–79.

[59] Geoff Boeing. *OSMnx: Python for Street Networks*. https://pypi.org/project/osmnx/. Version 1.7.0. 2024.

A

Supporting Data

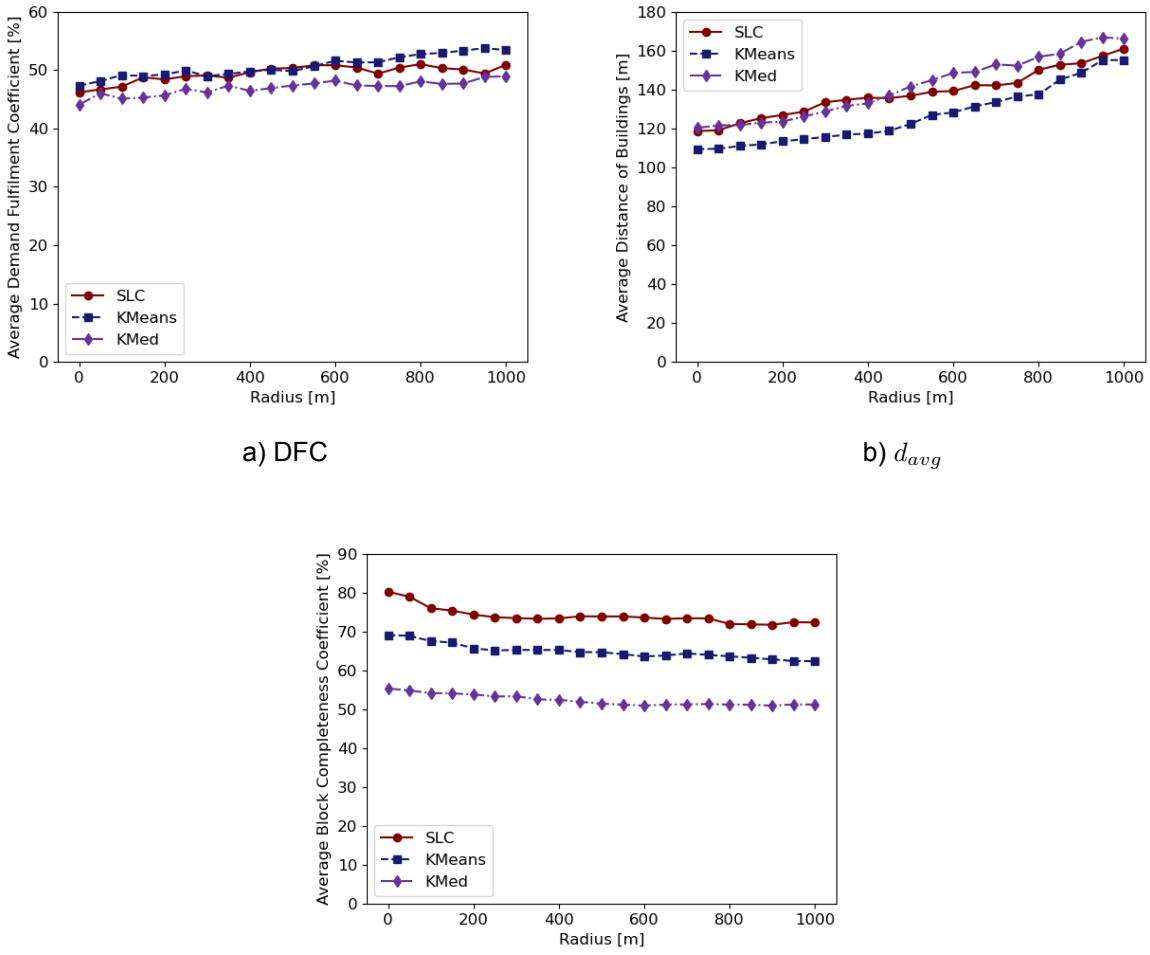
A.1. Underlying Data

Figure A.1 shows the building blocks in the centre of Amsterdam, where each building block has a distinct colour. The in the figure represented data is the base for the calculation of the BCC (see subsection 4.1.3).

Figure A.1: Building Blocks in the city centre of Amsterdam

A.2. Case Study Results

Figure A.2 illustrates the value of the optimization metrics over the switching radius in the energy optimization step. Since the increase in improvement regarding the DFC for increasing radii is small and the negative effect on the other metrics is negligible for the 50m radius, the 50m radius is chosen.



a) DFC

b) d_{avg}

c) BCC

Figure A.2: Average of the Optimization Metrics Depending on the Switching Radius

A.3. Cluster Viability

Figure A.3 shows the viable SLC clusters in the 95% participation and the case of no utility buildings.

a) 95 % participation b) No Utility

Figure A.3: Viable SLC clusters in different scenarios

A.4. Sensitivity Analysis Scenario Results

The here presented tables include a sensitivity analysis of the average DFC, average BCC and average d_{avg} for the different runs of the scenarios. It can be seen that for all three performance indicators, there is little variation across the different runs of a scenario.

Table A.1: Demand Fulfilment Coefficient Sensitivity Analysis K-Means

Scenario	Average	Standard Deviation	Minimum	Maximum
95 % Participation	36.1%	0.6%	34.3%	37.0%
80 % Participation	35.3%	1%	32.1%	37.4%
70 % Participation	34.8%	1.3%	31.2%	38.4%
50 % Participation	33.2%	1.4%	29.8%	38.1%

Table A.2: Demand Fulfilment Coefficient Sensitivity Analysis SLC

Scenario	Average	Standard Deviation	Minimum	Maximum
95 % Participation	34.3%	0.5%	32.2%	35.2%
80 % Participation	33.6%	1%	31.0%	37.5%
70 % Participation	32.9%	1.4%	29.6%	36.5%
50 % Participation	31.5%	1.4%	28.4%	35.8%

Table A.3: Block Completeness Coefficient Sensitivity Analysis KMeans

Scenario	Average	Standard Deviation	Minimum	Maximum
95 % Participation	69.1%	0.2%	67.8%	68.7%
80 % Participation	68.7%	0.3%	67.9%	69.5%
70 % Participation	69.3%	0.5%	68.0%	70.3%
50 % Participation	71.1%	0.7%	69.5%	73.5%

Table A.4: Block Completeness Coefficient Sensitivity Analysis SLC

Scenario	Average	Standard Deviation	Minimum	Maximum
95 % Participation	79.3%	0.2%	78.9%	80.0%
80 % Participation	80.0%	0.4%	79.0%	81.3%
70 % Participation	80.6%	0.5%	79.7%	81.8%
50 % Participation	82.2%	0.8%	80.5%	83.9%

Table A.5: d_{avg} Sensitivity Analysis KMeans

Scenario	Average	Standard Deviation	Minimum	Maximum
95 % Participation	109.1m	0.2m	108.6m	109.5m
80 % Participation	109.2m	0.4m	107.9m	110.4m
70 % Participation	109.1m	0.56m	107.7m	110.9m
50 % Participation	109.1m	0.9m	106.4m	111.6m

Table A.6: d_{avg} Sensitivity Analysis SLC

Scenario	Average	Standard Deviation	Minimum	Maximum
95 % Participation	118.8m	0.25m	118.2m	119.4m
80 % Participation	118.9m	0.46m	117.7m	119.8m
70 % Participation	118.7m	0.6m	117.2m	120.2m
50 % Participation	118.8m	1.0m	116.5m	121.4m