

Thesis

Clustering Dispersed Industrial Demand for Integration into the Dutch Hydrogen Backbone

Author

Thijs van Dijk
Student number: 4923596

Graduation Committee

Chair: Jan Anne Annema
First supervisor: Petra Heijnen
Second supervisor: Martijn Warnier

Institution

Delft University of Technology
Faculty of Technology, Policy and Management

Date

August 16, 2026

Project repository

github.com/vandijkthijs-source/Clustering-hydrogen-demand



Preface

This Master Thesis marks the end of an important phase of my life as a student. During my years of studying, I have learned a lot from lectures, fellow students and friends. Not only academically, but also through collaboration, discussions and the experiences that come with student life. I am grateful for this period and for the opportunity to conclude it with this thesis.

After completing my bachelor's degree in Mechanical Engineering, I wanted to explore more of the business and organisational side of technology. This led me to start the Master Management of Technology. During these two years, I learned to look at technological challenges from a broader perspective, including strategy, policy, economics and decision-making. At the same time, I gradually started to miss the more technical and analytical aspects of engineering. This thesis allowed me to bring both sides together.

Before starting this thesis, I worked as a data engineer for one year. I enjoyed the technical puzzling that came with that work, especially the process of structuring problems, working with data and improving models step by step. This made me increasingly interested in network theory and infrastructure modelling. With the help of Petra Heijnen and Martijn Warnier, this interest was connected to the topic of hydrogen backbone connections and Cluster 6. This resulted in a thesis topic that combined technical modelling with a relevant societal challenge: supporting the energy transition.

The beginning of the thesis process was not always easy. Finding a suitable topic took time, and it required several discussions before the final direction became clear. Looking back, this search was also valuable. It helped me understand what kind of problem I wanted to work on and what type of research suited me. Throughout the process, I learned to work more independently, to deal with uncertainty and to translate a broad problem into a structured research approach. I also learned how important it is to keep improving a model while staying critical of what the results can and cannot show.

I would like to thank Petra Heijnen and Martijn Warnier, my first and second supervisors, for their guidance, support and enthusiasm throughout the process. Their feedback helped me to keep moving forward and to improve both the model and the thesis. I also want to thank Jan Anne Annema for being the chair of my graduation committee and for his enthusiasm.

Finally, I want to thank my friends, family and fellow students who supported me during this thesis period. Being able to share thoughts, doubts and progress along the way helped me to stay motivated and bring this project to a good conclusion.

I hope you enjoy reading this thesis.

Executive Summary

The Netherlands is quickly changing towards reducing its greenhouse gas emissions to combat climate change. This comes with major challenges as the electricity grid starts reaching its limits (International Energy Agency, 2025b) and high-temperature applications are difficult to electrify. Industry is an important part of this challenge. The manufacturing sector alone accounts for approximately 33% of national greenhouse gas emissions (Statistics Netherlands (CBS), 2026). A large amount of these emissions are linked to the use of natural gas, for example for high-temperature heat and feedstock (producing ammonia, methanol or other basic chemicals). Energy efficiency and electrification can reduce part of the natural gas use but they are less suitable for processes that require very high temperatures or for feedstock. In these cases, green hydrogen may play a more targeted role.

To support this transition, the Netherlands is developing a national hydrogen backbone. This network, developed by Hynetwork, a subsidiary of Gasunie, is intended to connect the five largest industrial clusters in the country and provide them with access to a future hydrogen supply (Hynetwork / Gasunie, 2025). However, the current design mainly focuses on large, concentrated industrial areas. Smaller and geographically dispersed industrial sites, often referred to as Cluster 6, are not automatically included in this infrastructure. As a result, a significant part of Dutch industrial demand may remain disconnected from the emerging hydrogen network.

The importance of Cluster 6 is shown in the 2025 report of the Landelijke Cluster Energie Strategie (LCES). This report includes data from 330 Cluster 6 production sites, which together represent approximately 12.5% of total Dutch CO₂ emissions (Stichting Cluster 6, 2025). The report shows that many of these companies have clear ambitions to decarbonise. However, it also concludes that most planned decarbonisation projects are unlikely to be realised in time. Around 73% of the projects that depend on new or expanded energy infrastructure cannot be completed before 2030 because of limited availability of electricity, hydrogen and other energy networks (Stichting Cluster 6, 2025).

The main issue is therefore not a lack of willingness among companies to reduce emissions. Rather, many companies are unable to act because the required infrastructure is not available. Cluster 6 companies face several barriers, including electricity grid congestion, missing hydrogen infrastructure, high investment costs and uncertainty about future energy supply.

This problem is reinforced by the spatial design of the national hydrogen backbone. Because the backbone is primarily aimed at the five major industrial clusters, many Cluster 6 sites are located too far away to connect directly. Their dispersed locations make individual connections expensive and inefficient. At the same time, leaving these companies disconnected would create an uneven energy transition in which large industrial clusters gain access to hydrogen infrastructure while smaller industrial regions fall behind.

This study addresses that problem by developing a method to assess how Cluster 6 industries can be connected to the national hydrogen backbone. The method combines spatial demand clustering with network optimisation using the Optimal Network Layout Tool (ONLT) (Heijnen, 2025). By grouping nearby hydrogen demand and identifying cost-efficient network layouts, the study explores where collective connections to the backbone may be technically and economically feasible.

The results show that spatial clustering has a clear influence on the design and cost of hydrogen infrastructure for Cluster 6. When only the internal network costs are considered, the model favours a relatively detailed clustering configuration. In this case, the lowest cost is found at 70 clusters, all with their own connections to the hydrogen backbone.

However, this result changes when connection costs to the hydrogen backbone are included. Each additional cluster may require an additional T-junction, which makes configurations with many clusters more expensive. When these T-junction costs are added, the preferred configuration shifts towards fewer and larger clusters. When connection costs range from € 150.000 to € 500.000, 34 clusters becomes the most

relevant practical configuration. This configuration provides a balance between keeping internal network distances acceptable and limiting the number of required backbone connection points. When T-junction costs become very high, ranging from € 600.000 to € 1.000.000, the preferred configuration shifts further towards 25 clusters.

To support the feasibility of the clustering results, an actor analysis was carried out to create actor-informed classification of the company dataset. Each company was first linked to its sector association and then labelled as either an anchor demand firm or a marginal demand firm. Sectors with a stronger expected role for hydrogen, higher technical relevance and more potential to support collective infrastructure development were labelled as anchor firms. This includes specialty chemicals, metallurgy, glass and, conditionally, ceramics. Sectors where hydrogen demand is more uncertain, for example because electrification may be a more suitable option, were labelled as marginal demand firms. This includes food processing, paper and cardboard, waste processing, data centres and onshore oil and gas.

These labels were then used to assess whether the clustering configurations were supported by credible demand. The configuration with 34 clusters contains anchor-demand companies in almost all clusters, making it more credible from an implementation perspective. In contrast, the configuration with 70 clusters includes more clusters without any anchor-demand companies.

Overall, the results show that the connection of dispersed Cluster 6 demand depends on the balance between internal transport distances and the number of required backbone connections. A configuration with many clusters keeps companies close together and therefore reduces internal pipeline distances. However, it also creates many separate connection points to the hydrogen backbone. A configuration with fewer clusters limits the number of backbone connections, but requires hydrogen demand to be bundled over longer internal distances.

Contents

Preface	1
Executive Summary	2
Nomenclature	8
1 Introduction	10
1.1 Link to Management of Technology	11
1.2 AI Statement	12
2 Literature Review and Research Gap	13
2.1 Literature Review	15
2.2 Literature Gap	18
2.3 Research Objective and Research Questions	19
3 Actor Analysis	20
3.1 Who owns the problem?	20
3.2 Which actors matter?	20
3.3 Who controls the formal resources?	21
3.4 Why are Cluster 6 firms not homogeneous?	23
3.5 How does this affect the clustering model?	25
4 Methodology	27
4.1 Research Design	27
4.2 Actor-informed Demand Classification	27
4.3 Spatial Clustering Method	28
4.4 Network Modelling with the ONLT	28
4.4.1 Graph Representation	29
4.4.2 Optimal Network Layout Tool	29
4.4.3 Cost function	30
4.5 Evaluation and Comparison	30
4.6 Robustness	31
4.7 Assumptions and Limitations	32
5 Data Collection and Preprocessing	33
5.1 Identification of Cluster 6 companies	33
5.2 Geographic Data Collection	33
5.3 Sector Classification	34
5.4 Hydrogen Demand Estimation	35
5.5 Actor-informed Demand Labels	35
5.6 Hydrogen Backbone Data	36
5.7 Cost Conversion	37
5.8 Backbone Connection Cost	38
6 Model Setup and Configurations	39
6.1 Clustering Configurations	39
6.2 ONLT Implementation	39
6.3 Cluster-to-Backbone Connection	39
6.4 Configuration Design	40
6.5 Output Indicators	40

7	Results and Analysis	42
7.1	Baseline Network Costs	42
7.2	Effect of Adding T-junction Costs	43
7.3	Interim Finding	47
7.4	Anchor-site Composition of Selected Clustering Configurations	47
7.5	Robustness of Selected Clustering Configurations	49
7.6	Threshold Sensitivity of Selected Clustering Configurations	51
8	Discussion	55
8.1	Interpretation of the Main Results	55
8.2	The Role of T-junction Costs	55
8.3	Actor-informed Interpretation of the Results	56
8.4	Practical Implications	56
8.5	Validation of the Results	57
8.6	Limitations	57
9	Conclusions	59
9.1	Answering the Research Questions	59
9.1.1	How can dispersed industrial demand within Cluster 6 be spatially clustered into representative regional groups?	59
9.1.2	How do different clustering configurations affect the spatial structure, connection requirements and feasibility of Cluster 6 hydrogen networks?	59
9.1.3	What is the relationship between cluster configuration and total infrastructure cost, and at what point do additional clusters yield diminishing returns?	60
9.1.4	How does spatial clustering of dispersed industrial hydrogen demand affect transport infrastructure design and cost?	60
9.2	Final Conclusion	60
10	Recommendations	62
10.1	Recommendations for Practice	62
10.2	Recommendations for Future Research	62
A	Anchor-informed Comparison of K-means Seed Runs	64

List of Figures

1	Cluster 6 companies (Stichting Cluster 6, 2025)	10
2	Formal chart of the most important institutional relations for the Cluster 6 hydrogen-connection problem.	22
3	Power-interest grid for the Cluster 6 hydrogen-connection problem.	25
4	Examples of a simple graph and a weighted graph	29
5	Overview of the ONLT optimisation steps. Adapted from the ONLT user guide	30
6	Geocoded Cluster 6 company locations	34
7	Actor-informed classification of Cluster 6 demand nodes	36
8	Simplified representation of the national hydrogen backbone and Cluster 6 company locations	37
9	Total ONLT cost for clustering configurations between $k = 20$ and $k = 140$, excluding T-junction costs.	42
10	k=70 network configuration	43
11	Total cost including T-junction penalties for C_T values between 150,000 and 1,000,000.	44
12	Total cost curves for four representative T-junction cost assumptions: $C_T = 0$, $C_T = 150,000$, $C_T = 500,000$ and $C_T = 1,000,000$	45
13	Shift in the cost-minimising number of clusters under increasing T-junction penalties.	46
14	k=34 network configuration	47
15	Anchor and marginal sites within the $k = 34$ clustering configuration.	48
16	Anchor count per K-means cluster for the $k = 70$ clustering configuration.	49
17	Per-cluster stability for $k = 34$, measured using mean Jaccard similarity across K-means seed runs.	50
18	Per-cluster stability for $k = 70$, measured using mean Jaccard similarity across K-means seed runs.	51
19	Number of credible feasible clusters under different contribution and subsidy assumptions. . .	54

List of Tables

2	Positioning of the reviewed literature and identified research gap	14
3	Main actors in the Cluster 6 hydrogen-connection problem.	21
4	Key dependencies in the Cluster 6 hydrogen-connection problem.	23
5	Sector-level assessment of Cluster 6 firms' likelihood to coordinate for a collective hydrogen connection.	24
6	Branch associations representing Cluster 6 industries	33
7	Sector categories used for Cluster 6 company classification	35
8	configuration dimensions used in the model setup.	40
9	Link between output indicators and actor decision criteria.	41
10	Contribution levels used in the threshold sensitivity analysis.	53
11	Anchor-informed seed comparison for $k = 34$	64
12	Anchor-informed seed comparison for $k = 70$	65

Nomenclature

Abbreviations

Abbreviation	Meaning
ACM	Authority for Consumers and Markets
CAPEX	Capital expenditure
CBS	Statistics Netherlands
DSO	Distribution System Operator
FNLI	Federatie Nederlandse Levensmiddelen Industrie
HNS	Hynetwork Services
IEA	International Energy Agency
IPO	Interprovinciaal Overleg
KGG	Ministry of Climate Policy and Green Growth
KNB	Koninklijke Nederlandse Bouwkeramiek
LCES	Landelijke Cluster Energie Strategie
ONLT	Optimal Network Layout Tool
OPEX	Operational expenditure
OWE	Opschaling volledig hernieuwbare waterstofproductie via elektrolyse
RFNBO	Renewable Fuel of Non-Biological Origin
RVO	Netherlands Enterprise Agency
VA	Vereniging Afvalbedrijven
VNG	Vereniging Nederlandse Glasfabrikanten
VNCI	Vereniging van de Nederlandse Chemische Industrie
VNMI	Vereniging Nederlandse Metallurgische Industrie
VNP	Vereniging van Nederlandse Papier- en Kartonfabrieken

Symbols

Symbol	Meaning
$\mathcal{S}$	Set of industrial sectors; $s_i \in \mathcal{S}$ is the sector of company i
$\mathcal{K}$	Set of tested numbers of spatial clusters, $\mathcal{K} = \{20, \dots, 140\}$; $k \in \mathcal{K}$
$\mathcal{C}_k$	Set of clusters in configuration k , $\mathcal{C}_k = \{1, \dots, k\}$; cluster index $c \in \mathcal{C}_k$
$\mathcal{N}_c^{(k)}$	Set of company nodes assigned to cluster $c \in \mathcal{C}_k$ under configuration $k \in \mathcal{K}$
$\mathcal{V}_c^{(k)}$	Set of network nodes in cluster c under configuration k , including company, connection and intermediate nodes
$\mathcal{E}_c^{(k)}$	Set of pipeline edges in cluster c under configuration k ; $(u, v) \in \mathcal{E}_c^{(k)}$
n_i	Company node $i \in \mathcal{N}$, represented by $n_i = (x_i, y_i, D_i, s_i, a_i)$
x_i, y_i	Spatial coordinates of company $i \in \mathcal{N}$
s_i	Sector classification of company $i \in \mathcal{N}$, with $s_i \in \mathcal{S}$
a_i	Actor-informed demand label of company $i \in \mathcal{N}$,
G_i	Natural-gas demand allocated to company $i \in \mathcal{N}$
ρ_{s_i}	Hydrogen substitution rate for the sector $s_i \in \mathcal{S}$ of company i
D_i	Estimated hydrogen demand of company $i \in \mathcal{N}$
$D_c^{(k)}$	Aggregated hydrogen demand of cluster $c \in \mathcal{C}_k$ under configuration $k \in \mathcal{K}$
$A_c^{(k)}$	Number of anchor-demand companies in cluster $c \in \mathcal{C}_k$ under configuration $k \in \mathcal{K}$
$\mu_c^{(k)}$	Spatial centroid of cluster $c \in \mathcal{C}_k$ under configuration $k \in \mathcal{K}$
α	Cost coefficient in the ONLT edge-cost function
β	Capacity-cost exponent in the ONLT cost function; $\beta = 0.42$ in the baseline analysis

Symbol	Meaning
$C_{\text{ONLT}}(k)$	Total ONLT cost of all clusters in configuration $k \in \mathcal{K}$
γ	Conversion factor from ONLT cost units to euros, expressed in euros per ONLT cost unit
$C_{\text{euro}}(k)$	Euro-equivalent network cost for clustering configuration $k \in \mathcal{K}$
C_T	Assumed cost per T-junction
$C_{\text{total}}(k)$	Total network cost for clustering configuration $k \in \mathcal{K}$, including T-junction costs
D_{min}	Assumed minimum annual hydrogen-demand threshold for a collective connection
m_{H_2}	Annual hydrogen mass, expressed in kilograms
E_{H_2}	Energy equivalent of the hydrogen demand, expressed in terajoules

1 Introduction

Combating climate change requires a rapid reduction of greenhouse gas emissions. In the Netherlands, industry is an important part of this transition. Industrial companies use large amounts of energy for heat, feedstock and other production processes, and a significant share of this energy is still supplied by natural gas. Although electrification, energy efficiency and process innovation can reduce part of this demand, not all industrial processes can easily be electrified. This is especially the case for high-temperature heat and feedstock use (Detz et al., 2020).

Green hydrogen is therefore increasingly considered an option for industrial decarbonization. When produced through electrolysis using renewable electricity, hydrogen can replace fossil fuels in processes where direct electrification is technically difficult. For this reason, hydrogen is expected to become part of the future Dutch energy system.

To support the development of a hydrogen economy, the Netherlands is developing a national hydrogen backbone. This backbone is intended to connect the largest industrial clusters in the country and provide access to future hydrogen supply (Gasunie, n.d.). The current infrastructure planning mainly focuses on the five major industrial clusters, where large-scale demand is geographically concentrated. This is logical from an infrastructure perspective, because these clusters provide high demand volumes and clear connection points.

However, not all industrial demand is located within these major clusters. A substantial group of industrial companies is located outside the five main industrial areas. These companies are often referred to as Cluster 6. They include firms from sectors such as food processing, paper and cardboard, specialty chemicals, metallurgy, glass, ceramics, waste processing, data centres and onshore oil and gas. Together, these companies represent approximately 12.5% of Dutch emissions, but their locations are spread across the country rather than concentrated around a few large industrial areas (Stichting Cluster 6, 2025).

Figure 1 illustrates this dispersed spatial pattern. Unlike the five major industrial clusters, Cluster 6 companies are spread across different provinces and regions. This spatial dispersion is central to the infrastructure challenge addressed in this thesis. Individual sites may be too small or too remote to justify a direct connection to the national hydrogen backbone. At the same time, their combined demand may still be large enough to make collective infrastructure solutions relevant.

The dispersed character of Cluster 6 creates a trade-off in hydrogen infrastructure planning. If Cluster 6 demand is treated as a set of individual sites, the required infrastructure may become fragmented and expensive. If demand is grouped into regional clusters, fewer connections to the hydrogen backbone may be needed and demand becomes more visible to infrastructure planners. However, stronger clustering also increases the internal transport distances within each cluster. The challenge is therefore to find a balance between many small clusters with short internal distances and fewer larger clusters with fewer backbone connection points.

This problem is directly related to the current development of the Dutch hydrogen backbone. The backbone is being developed as a national transport network, but connection opportunities for dispersed industrial

De bedrijven van de 9 sectoren verspreid over Nederland



Figure 1: Cluster 6 companies (Stichting Cluster 6, 2025).

users depend on whether demand can be aggregated into credible regional connection points. Hynetwork and Gasunie therefore play an important role in determining how future hydrogen infrastructure may connect to industrial users outside the main clusters (Gasunie, 2024; Hynetwork Services, 2025).

At the same time, Cluster 6 firms are not homogeneous. Some sectors, such as specialty chemicals, metallurgy and glass, have a stronger expected role for hydrogen because of feedstock use or high-temperature processes. Other sectors may have more realistic alternatives, such as electrification. This means that the credibility of clustered hydrogen demand does not only depend on location and demand volume, but also on the type of companies included in each cluster.

This thesis investigates how spatial clustering of dispersed industrial hydrogen demand affects transport infrastructure design and cost. The focus is on Cluster 6 companies in the Netherlands and their possible connection to the national hydrogen backbone. The study combines company-level demand estimation, spatial clustering and network optimisation using the Optimal Network Layout Tool (Heijnen, 2026). By varying the number of clusters, the research evaluates how different spatial configurations influence network length, connection requirements and total infrastructure cost.

The relevance of this study is both scientific and practical. Scientifically, it contributes to the literature on hydrogen infrastructure planning by treating spatial clustering as a modelling choice rather than as a fixed input assumption. Many existing infrastructure studies optimise network design for predefined demand nodes, but give less attention to how the aggregation of demand itself affects the outcome. This thesis addresses that gap by systematically comparing different clustering configurations.

Practically, the study supports the discussion about how Cluster 6 companies could be included in future hydrogen infrastructure planning. The results can help identify where collective connections may be more promising and where infrastructure costs remain high. They can also support discussions between Cluster 6 firms, Stichting Cluster 6, Hynetwork, regional authorities and policymakers about how dispersed industrial demand can be made more credible and actionable.

The central research question of this thesis is:

How does spatial clustering of dispersed industrial hydrogen demand affect transport infrastructure design and cost?

To answer this question, the thesis is structured as follows. Chapter 2 reviews the academic literature on hydrogen infrastructure planning, spatial modelling, industrial hydrogen demand and network optimisation. It identifies the research gap and formulates the research objective and sub-questions. Chapter 3 presents the actor analysis, which explains why Cluster 6 hydrogen infrastructure is not only a technical design problem, but also a coordination problem between firms, infrastructure operators, regulators and public authorities. Chapter 4 describes the clustering method, demand configurations, network model and cost evaluation. Chapter 5 explains the data collection and preprocessing steps. Chapter 6 describes the model setup and configuration design. Chapter 7 presents the network and cost outcomes for different clustering configurations. Finally, the thesis discusses the implications of the findings and provides conclusions for hydrogen infrastructure planning for dispersed industrial demand.

1.1 Link to Management of Technology

This thesis is closely connected to the Management of Technology programme. It studies a technological transition that cannot be understood from a technical perspective alone. The development of hydrogen infrastructure for Cluster 6 involves technical network design, industrial demand uncertainty, investment decisions, policy choices and coordination between multiple actors.

The research combines these elements by linking spatial clustering and network optimisation with actor-informed interpretation. The technical part of the thesis focuses on modelling dispersed industrial demand

and evaluating infrastructure cost trade-offs. The management and governance part focuses on the question of whether clustered demand is credible from an implementation perspective, given the roles of firms, Hynetwork, Stichting Cluster 6, policymakers and regional authorities.

In this way, the thesis reflects the central idea of Management of Technology: technological systems are shaped not only by engineering possibilities, but also by organisational, economic and institutional conditions. The hydrogen backbone is therefore treated not only as an infrastructure system, but also as a coordination challenge in the Dutch energy transition.

1.2 AI Statement

During the writing and modelling process of this thesis, generative artificial intelligence tools were used as supporting tools. Their use was limited to activities such as improving language and readability, restructuring paragraphs, identifying inconsistencies in the text, clarifying methodological descriptions, and providing support during the debugging and documentation of code. AI tools were also used to explore alternative formulations and to improve the presentation of equations, assumptions and modelling steps.

Generative AI was not used to independently conduct the research, select data sources, define the research methodology or determine the analytical approach. All research decisions, including the selection and processing of data, the design of the clustering method, the implementation of the network model, the choice of assumptions and the interpretation of the results, were made by the author. AI-generated suggestions were treated as preliminary input and were only included after being reviewed, verified and, where necessary, revised.

Particular attention was given to the reliability of AI-generated content. Factual claims, calculations, references, equations and code suggestions were checked against the original data, academic literature and model outputs. AI-generated references or unsupported factual statements were not accepted without independent verification. The author therefore remained responsible for ensuring that the final thesis accurately represented the research process and its findings.

No confidential, personal or commercially sensitive information was shared with generative AI tools. The use of AI did not replace independent analysis, critical reflection or academic judgement. The author remains fully responsible for the content, accuracy, originality and academic integrity of this thesis.

2 Literature Review and Research Gap

This chapter analyses the current academic literature to identify a clear knowledge gap. Based on this gap, research questions are formulated, building on the problem statement introduced in the previous chapter.

Relevant studies were identified through academic databases and citation networks. The selection focused on hydrogen transport, infrastructure planning, and modelling approaches, including optimisation, spatial modelling, and system analysis. In total, 25 studies were reviewed.

Table 2: Positioning of the reviewed literature and identified research gap

Literature stream	Main contribution	Representative studies	Remaining gap
Hydrogen infrastructure optimisation	Develops cost-minimising models for hydrogen production, storage, transport and network design at regional, national or European scale.	(Beagle et al., 2023; Bique & Zondervan, 2018; Kotek et al., 2023; Kountouris et al., 2024; Namazifard et al., 2024; Parolin et al., 2022, 2024; Timalisina et al., 2025)	Most studies optimise a given spatial demand structure, but do not systematically test how different levels of spatial aggregation affect network cost.
Spatial and regional hydrogen planning	Shows that geography, clustering, local demand distribution and routing assumptions strongly influence hydrogen infrastructure design.	(Hammond et al., 2025; Lafeber, 2024; Mendler et al., 2025; Namazifard et al., 2024; Parolin et al., 2024)	Spatial detail is recognised as important, but the trade-off between highly disaggregated company-level demand and aggregated cluster-level demand remains underexplored.
Pipeline routing and graph-based network design	Uses graph theory, minimum spanning trees, Steiner trees and routing algorithms to determine cost-efficient pipeline layouts.	(Hammond et al., 2025; Lafeber, 2024; Oostdijk, 2025; Parolin et al., 2022)	Existing work focuses mainly on finding better routes for a fixed set of nodes, rather than analysing how the number and location of aggregated nodes shape total network cost.
Industrial hydrogen demand studies	Estimates potential hydrogen demand in hard-to-abate industries and shows that demand is spatially concentrated in industrial clusters.	(Franco & Giovannini, 2023; Lipiäinen et al., 2023; Liu et al., 2024; Neuwirth et al., 2024)	Demand is often estimated at plant, sector or regional level, but the methodological implications of aggregating industrial sites into clusters are not fully assessed.
Transport modes and hydrogen carriers	Compares pipelines, trucks, liquid hydrogen, ammonia, methanol, formic acid and other carriers across cost, efficiency and emissions.	(Busch et al., 2023; Cheng & Cheng, 2023; Kim et al., 2023; Lipiäinen et al., 2023; Yang et al., 2023)	These studies clarify technology choices, but usually do not address how spatial aggregation of demand changes the required network structure.
Policy, regulation and governance	Highlights the role of regulation, investment uncertainty, public support and institutional responsibilities in hydrogen infrastructure development.	(Beagle et al., 2023; Kotek et al., 2023; Kountouris et al., 2024; van ’t Veer et al., 2025; Vergara et al., 2024)	Policy studies identify uncertainty and coordination problems, but provide limited quantitative evidence on how aggregation choices affect infrastructure planning outcomes.

2.1 Literature Review

The reviewed literature can be grouped into six main categories:

- (1) hydrogen infrastructure planning and optimisation,
- (2) hydrogen transport modes and carriers,
- (3) industrial hydrogen demand and hard-to-abate sectors,
- (4) spatial, regional and cluster-based planning,
- (5) policy, regulation and governance, and
- (6) review and state-of-the-art studies.

Together, these strands show that hydrogen infrastructure planning is strongly shaped by spatial demand patterns, transport technology choices, industrial demand uncertainty and regulatory conditions. However, they also show that the effect of spatial aggregation itself is still insufficiently analysed.

1. Hydrogen infrastructure planning and optimisation

A large part of the literature focuses on optimisation-based hydrogen infrastructure planning. These studies generally aim to minimise total system cost by jointly considering hydrogen production, storage, transport and demand. Early work by (Bique & Zondervan, 2018) develops a national hydrogen supply chain model for Germany, optimising production, storage and transport decisions for 2030 and 2050. More recent studies extend this type of modelling with higher spatial detail, uncertainty analysis and sector coupling.

For example, (Beagle et al., 2023) develop the HOWDI model to analyse hydrogen distribution infrastructure in Texas under uncertainty. Their Monte Carlo analysis shows that policy incentives strongly affect whether hydrogen can be delivered at the target cost of 4 USD/kg. Similarly, (Namazifard et al., 2024) develop a MILP model for Belgium that allocates industrial hydrogen demand to clusters and evaluates network development under different supply configurations, including offshore green hydrogen, blue hydrogen, ammonia imports and decentralised inland electrolysis.

At the European scale, (Kotek et al., 2023) and (Kountouris et al., 2024) analyse future hydrogen backbone development. (Kotek et al., 2023) show that hydrogen blending in the natural gas grid is not a viable long-term transport solution and that dedicated or repurposed hydrogen pipelines are required. (Kountouris et al., 2024) identify major European hydrogen corridors and show that a self-sufficient green hydrogen system would require very large electrolysis capacity, while only moderately increasing total system cost.

Several studies also include long-term uncertainty and staged infrastructure development. (Timalsina et al., 2025) combine MILP with real options analysis and show that phased investment can reduce risk under uncertain demand. (Lafeber, 2024) develops a robust backtracking planning method for industrial port clusters, showing that early robust capacity investment can reduce long-term costs, but may also require higher upfront investment.

These optimisation studies provide important insights into cost-efficient hydrogen infrastructure design. However, most of them optimise the network for a predefined spatial demand structure. Demand nodes are usually fixed beforehand, either as regions, clusters or selected demand points. As a result, the spatial aggregation of demand is treated as an input assumption rather than as a variable that may itself influence infrastructure cost.

2. Hydrogen transport modes and carriers

A second strand of literature evaluates hydrogen transport modes and carrier options. These studies compare pipelines, trucks, liquid hydrogen, ammonia, methanol, formic acid, synthetic methane and other storage or transport pathways.

Pipelines are generally identified as the most suitable option for large-scale and long-distance hydrogen transport. (Cheng & Cheng, 2023) argue that existing natural gas pipelines in Canada could play an important role in hydrogen transport, either through blending or repurposing. However, they also emphasise

technical challenges such as hydrogen embrittlement, leakage, inspection requirements and end-use compatibility. (Lipiäinen et al., 2023) compare several hydrogen supply routes for a large steel mill and find that on-site electrolysis and hydrogen pipelines perform best in terms of cost, efficiency and emissions.

Other studies focus on alternative carriers. (Busch et al., 2023) analyse the role of liquid hydrogen in a future German energy system. They find that liquid hydrogen transport is mainly relevant when there is explicit liquid hydrogen demand, for example in aviation or high-purity industrial applications. (Kim et al., 2023) compare methanol and formic acid as green hydrogen carriers produced from captured CO₂. Their results show that methanol can be economically attractive compared to liquid hydrogen, while electrochemical formic acid may become environmentally favourable under optimistic renewable electricity assumptions. (Yang et al., 2023) provide a broader technical review of compressed hydrogen, liquid hydrogen, cryo-compressed hydrogen, LOHCs, metal hydrides, pipelines, trailers and shipping.

This literature shows that there is no single best transport technology. The preferred option depends on distance, scale, purity requirements, available infrastructure and end-use form. However, these studies mostly evaluate transport modes for a given spatial configuration. They provide limited insight into how the spatial aggregation of demand changes the relative importance of internal transport, backbone connections and economies of scale.

3. Industrial hydrogen demand and hard-to-abate sectors

A third strand focuses on the demand side of hydrogen, especially in hard-to-abate industries. These studies show that hydrogen demand is likely to emerge first in sectors where electrification is difficult or where hydrogen can replace fossil feedstocks.

(Franco & Giovannini, 2023) review hydrogen applications in hard-to-abate sectors and identify several routes for hydrogen use, including blended combustion, feedstock substitution, hydrogen as a reducing agent and energy storage. The steel sector is highlighted as one of the most promising applications, particularly through hydrogen-based direct reduction and electric arc furnaces. (Neuwirth et al., 2024) estimate the technical potential for hydrogen demand in German industry using a site-specific approach. They find that potential demand is very large and strongly concentrated in a small number of steel and chemical sites.

Industrial case studies further illustrate the importance of demand scale. (Lipiäinen et al., 2023) show that a large steel mill can justify pipeline-based hydrogen supply because of its high and concentrated demand. (Liu et al., 2024) analyse ammonia production in Shandong and show that hydrogen demand can be coordinated with renewable power generation and interregional transport to reduce system costs. (Lafeber, 2024) also emphasises that hydrogen demand in industrial clusters is uncertain because plant participation and technology adoption may develop differently over time.

These studies are important because they explain why industrial hydrogen demand is not evenly distributed. Demand is sector-specific, technology-dependent and spatially concentrated. However, the link between company-level industrial demand and aggregated infrastructure nodes is not fully developed. In many models, demand is already aggregated before optimisation, without systematically testing how this aggregation affects the resulting network.

4. Spatial, regional and cluster-based planning

A fourth strand explicitly addresses the spatial dimension of hydrogen infrastructure. These studies show that geography, clustering, routing and regional characteristics strongly affect infrastructure cost.

(Mendler et al., 2025) develop a spatially resolved modelling framework for hydrogen valleys. Their results show that spatial resolution has a significant effect on system cost: overly coarse spatial aggregation can underestimate costs by up to 10%. They also show that different hydrogen exchange mechanisms and import assumptions change the optimal infrastructure configuration. (Hammond et al., 2025) focus on pipeline routing and demonstrate that accounting for real geographical obstacles and cost surfaces can reduce routing costs by 20–40% compared to simplified graph-based approaches.

Regional optimisation studies by (Parolin et al., 2022) and (Parolin et al., 2024) analyse hydrogen delivery infrastructure in Sicily. These studies show that multi-modal transport systems combining pipelines, compressed hydrogen trucks and liquid hydrogen trucks can reduce delivered hydrogen costs. They also show that multi-sector demand improves infrastructure utilisation and lowers average delivery cost. (Namazifard et al., 2024) apply a national cluster-based approach to Belgium, showing that different supply configurations still produce robust backbone requirements for industrial clusters.

(Lafeber, 2024) provides a more local cluster-level approach by developing a robust network planning method for industrial port clusters. This work is particularly relevant because it considers local routing constraints, demand uncertainty and staged deployment over time.

Overall, this literature confirms that spatial detail matters. However, most studies either use a fixed regional structure, predefined clusters or a fixed set of demand nodes. The number of clusters and the level of aggregation are rarely varied systematically. Therefore, the literature does not yet provide clear guidance on how aggregated industrial demand should be represented when modelling dispersed hydrogen users.

5. Policy, regulation and governance

A fifth strand focuses on the policy and institutional conditions for hydrogen infrastructure development. These studies show that hydrogen networks are not only technical systems, but also depend on regulation, investment incentives and actor responsibilities.

(van 't Veer et al., 2025) analyse hydrogen distribution in the Netherlands, with specific attention to Cluster 6 industries. They show that regulatory ambiguity regarding hydrogen distribution infrastructure creates uncertainty for investors and industrial users. In particular, unclear responsibilities for distribution system operators and other actors make it difficult to develop infrastructure for industries outside the major industrial clusters.

At the European scale, (Kotek et al., 2023) and (Kountouris et al., 2024) highlight the need for coordinated infrastructure development. Their studies show that hydrogen backbone development depends on cross-border planning, investment timing, demand uncertainty and the balance between domestic production and imports. (Beagle et al., 2023) also show that policy incentives, such as tax credits, can strongly affect whether hydrogen infrastructure reaches cost targets.

This literature is important for the Dutch Cluster 6 context because dispersed industries may face additional coordination challenges compared to large industrial clusters. However, policy and governance studies generally do not provide quantitative network models that test how different spatial configurations affect infrastructure cost.

6. Review and state-of-the-art studies

Finally, several studies provide broader reviews of hydrogen technologies, applications and research trends. (Yang et al., 2023) review storage and transport technologies and provide a technical basis for comparing infrastructure options. (Franco & Giovannini, 2023) review hydrogen applications in hard-to-abate sectors and explain where hydrogen is most likely to be relevant. (Vergara et al., 2024) provide a bibliometric review of industrial hydrogen applications in Europe, showing that research activity has increased strongly and that Germany plays a leading role in both scientific output and industrial hydrogen relevance.

These studies are useful for framing the broader hydrogen transition. They identify key technologies, applications and research trends, but they do not develop spatially explicit infrastructure models. As a result, they support the motivation for hydrogen infrastructure research, but do not address the specific methodological issue of demand aggregation.

2.2 Literature Gap

The reviewed literature shows that hydrogen infrastructure planning is well developed in several directions, but that these directions remain only partly connected.

First, optimisation and network design studies provide detailed cost-minimising infrastructure models. However, they usually begin from a predefined set of demand nodes. The spatial representation of demand is therefore treated as an assumption rather than as a modelling choice that can influence the outcome.

Second, spatial and regional studies demonstrate that geography matters. Spatial resolution, routing constraints, regional demand distribution and connection points can significantly affect network cost. Nevertheless, these studies generally do not systematically vary the level of spatial aggregation. They often compare supply configurations, transport modes or policy assumptions, but not different numbers of industrial demand clusters.

Third, industrial hydrogen demand studies show that future demand is sector-specific and spatially concentrated. Demand depends on process temperature, feedstock requirements, technology readiness and competition with electrification. However, the step from individual industrial sites to aggregated demand nodes is usually not critically examined.

Fourth, transport technology studies provide important cost and efficiency insights, especially regarding pipelines, trucks and hydrogen carriers. Yet they rarely analyse how the spatial distribution of demand affects the need for internal cluster networks, backbone connections and capacity economies of scale.

Finally, policy and governance studies highlight uncertainty, coordination problems and regulatory barriers, especially for dispersed industrial users such as Dutch Cluster 6 companies. However, these studies do not quantify how different spatial clustering choices affect infrastructure design and cost.

This creates a clear research gap. Existing studies either provide detailed optimisation models without testing demand aggregation, or they recognise spatial complexity without making clustering a central object of analysis. No reviewed study simultaneously focuses on geographically dispersed industrial hydrogen users, varies the level of spatial clustering, and quantifies how this affects pipeline network design and infrastructure cost.

2.3 Research Objective and Research Questions

The objective of this study is to analyse how different levels of spatial clustering affect the design and cost of hydrogen infrastructure for dispersed industrial demand. The study compares configurations with many small clusters to configurations with fewer, larger clusters. Smaller clusters reduce internal transport distances, but require more separate connections to the national hydrogen backbone. Larger clusters reduce the number of backbone connections, but increase the distance over which demand must be bundled within each cluster.

By evaluating this trade-off, the study aims to identify which clustering configurations provide a suitable balance between network cost, connection requirements and the credibility of clustered hydrogen demand.

This leads to the following research question:

How does spatial clustering of dispersed industrial hydrogen demand affect transport infrastructure design and cost?

To answer this question, the following sub-questions are formulated:

- How can dispersed industrial demand within Cluster 6 be spatially clustered into representative regional groups?
- How do different clustering configurations affect the spatial structure, connection requirements and feasibility of Cluster 6 hydrogen networks?
- What is the relationship between cluster configuration and total infrastructure cost, and at what point do additional clusters yield diminishing returns?

Based on the identified research gap, this study adopts k-means clustering to aggregate dispersed industrial demand and the Optimal Network Layout Tool (ONLT) to evaluate the resulting hydrogen network layouts.

3 Actor Analysis

The development of hydrogen infrastructure for Cluster 6 companies is not only a technical and economic design problem but also a multi-actor problem. Decisions about whether, where and when a hydrogen connection is realised depend on the interaction between network operators, regulators, ministries, regional authorities and the industrial companies themselves. Understanding this actor network is therefore necessary to interpret the modelling results of this thesis in their wider policy context.

This section follows the logic of the actor analysis method described in Chapter 4 of *Policy Analysis of Multi-Actor Systems* (Enserink et al., 2022). The results are organised around five questions that are most relevant for this thesis: who owns the problem, which actors matter, who controls the formal resources and why Cluster 6 firms are not homogeneous.

3.1 Who owns the problem?

The problem owner in this analysis is Stichting Cluster 6. They represent the dispersed industrial sites outside the five main Dutch industrial clusters and brings together their shared interest in whether and how to access the future hydrogen infrastructure (Stichting Cluster 6, 2025).

The desired situation is timely and affordable access to renewable hydrogen for the Cluster 6 companies. The current or expected situation is that many Cluster 6 decarbonisation projects cannot be realised before 2030 because the required electricity and hydrogen infrastructure is not yet available (Hynetwork / Gasunie, 2025; Stichting Cluster 6, 2025).

The main dilemma is that Cluster 6 firms are hesitant to commit to investing in hydrogen alternatives without assurance for affordable and reliable infrastructure, while infrastructure planners are hesitant to extend the backbone without credible demand. This makes the problem not only a technical design problem but also a coordination problem between many different actors.

Three decision arenas are relevant for this problem. The first is the national roll-out of the hydrogen backbone, led by the Ministry of Climate Policy and Green Growth and Hynetwork. The second is the regional coordination of Cluster 6 sites. The third is the connection process, in which Hynetwork evaluates individual or collective connection requests.

This thesis mainly focuses on the second and third arenas. The clustering analysis is used to make dispersed Cluster 6 demand more visible and credible, so that it can be better assessed for collective connection planning.

3.2 Which actors matter?

Table 3 summarises the actors that are most relevant for the Cluster 6 hydrogen-connection problem. The selection combines institutional actors, market actors, regional actors and representative organisations.

Table 3: Main actors in the Cluster 6 hydrogen-connection problem.

Actor	Relevance for Cluster 6 hydrogen connections
Stichting Cluster 6	Represents dispersed industrial demand and acts as collective problem owner.
Cluster 6 firms	Provide potential hydrogen demand and decide whether to participate in collective connections.
Branch associations	Represent sector-specific interests and help coordinate firms within sectors.
Hynetwork Services	Develops and operates the national hydrogen backbone and evaluates connection requests.
Gasunie	Parent company of Hynetwork Services and owner of relevant infrastructure expertise.
Ministry of Climate Policy and Green Growth	Sets national hydrogen policy and provides the mandate for backbone development.
Authority for Consumers and Markets	Future regulator for hydrogen network access and tariffs.
Regional Distribution System Operators	Potential future actors in regional hydrogen distribution, although their current mandate remains uncertain.
Provinces / IPO	Relevant for spatial planning, permitting and regional coordination.
Netherlands Enterprise Agency	Administers subsidy instruments relevant for industrial decarbonisation.
Hydrogen producers	Provide future hydrogen supply and depend on credible aggregated demand.
European Commission	Sets the wider EU legal and regulatory framework for hydrogen markets, state aid and certification.

The actor inventory shows that Cluster 6 hydrogen connections depend on actors from different levels. Stichting Cluster 6 and the firms represent the demand side, while Hynetwork and Gasunie control the main infrastructure pathway. Ministries, regulators and regional authorities shape the policy, regulatory and spatial conditions under which collective connections may be realised.

3.3 Who controls the formal resources?

The formal institutional setting is shown in Figure 2. The figure does not show every relation in the hydrogen system. Its purpose is to show the formal relations that are most relevant for connecting Cluster 6 sites to the national hydrogen backbone.

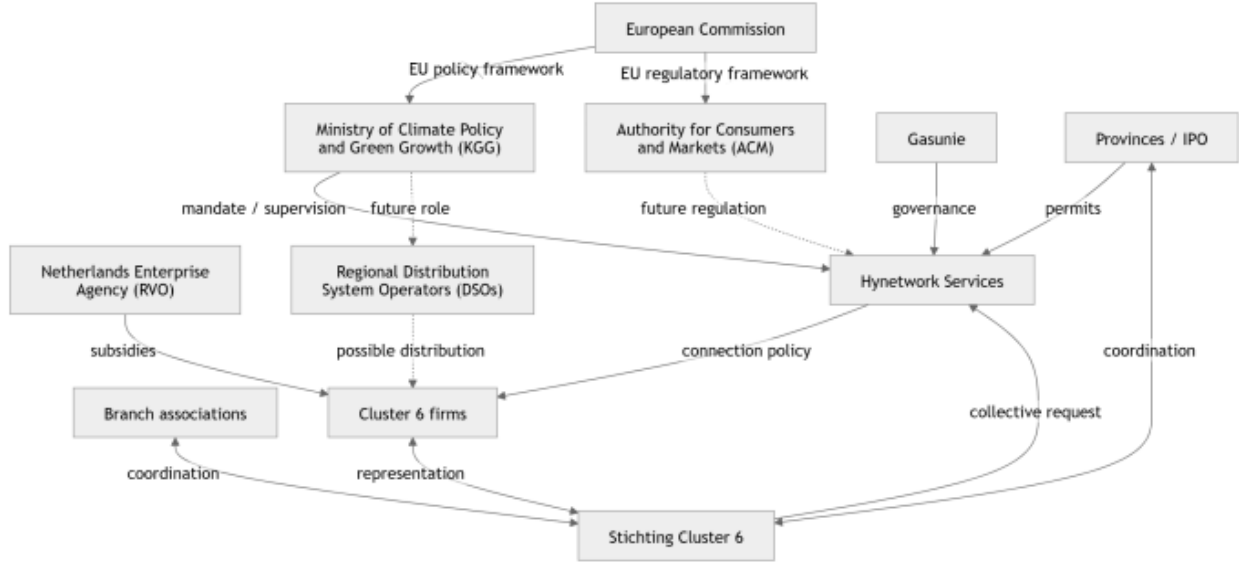


Figure 2: Formal chart of the most important institutional relations for the Cluster 6 hydrogen-connection problem.

Figure 2 shows that Stichting Cluster 6 can coordinate and aggregate demand, but does not control the main formal resources needed to realise a connection. The Ministry of Climate Policy and Green Growth provides the national policy framework and the mandate for the hydrogen backbone. Hynetwork Services has the central operational role, because any collective Cluster 6 connection must ultimately connect to the backbone that they develop and operate (Netherlands Court of Audit (Algemene Rekenkamer), 2025). The Authority for Consumers and Markets is relevant for future access and tariff regulation. Regional Distribution System Operators may become important for regional hydrogen distribution, but their formal role is still uncertain under current regulation (van 't Veer et al., 2025). Provinces influence spatial planning and permits, while the Netherlands Enterprise Agency supports industrial decarbonisation through subsidy instruments.

Table 4 summarises the most important resources and constraints.

Table 4: Key dependencies in the Cluster 6 hydrogen-connection problem.

Actor	Main resource or constraint	Implication for Cluster 6
Hynetwork Services	Backbone design, connection policy and technical expertise	Critical actor for any collective connection.
Ministry of Climate Policy and Green Growth	Policy mandate, public funding and regulatory framework	Determines the national priority given to Cluster 6.
Authority for Consumers and Markets	Future tariff and access regulation	Influences affordability and cost allocation.
Regional Distribution System Operators	Local network knowledge and possible future hydrogen distribution role	Important for regional connections, but current role is uncertain.
Provinces / IPO	Spatial planning, permits and regional coordination	Relevant for routing and regional feasibility.
Cluster 6 firms	Demand volumes, site locations and investment willingness	Collective demand is needed to make connections credible.
Stichting Cluster 6	Coordination capacity, sector data and representation	Can aggregate demand, but lacks formal decision-making power.
Hydrogen producers	Future hydrogen supply and offtake contracts	Need demand certainty before investing.
European Commission	EU legal framework, RFNBO rules and state-aid conditions	Sets background conditions for Dutch hydrogen policy.

The formal resource distribution creates an important dependency. Stichting Cluster 6 has information, representation and coordination capacity, but limited direct decision-making power. Hynetwork, Gasunie, the Ministry, the regulator and regional authorities control the resources that determine whether a collective connection can actually be realized. This means that Cluster 6 needs to make its demand visible and credible to actors that control the infrastructure decisions.

3.4 Why are Cluster 6 firms not homogeneous?

Treating Cluster 6 firms as one collective actor is useful for describing the overall problem but neglects an important problem. The willingness of a firm to participate in a collective hydrogen connection depends strongly on its sector.

Five sector-specific factors are especially relevant:

1. **Technical needs:** some sectors depend on high-temperature heat, feedstock use or process conditions that are difficult to electrify. Hydrogen is therefore more technically relevant for these sectors.
2. **Competing decarbonisation options:** for low-temperature processes, such as food processing and parts of paper production, direct electrification is often a credible alternative. This lowers the relative urgency of hydrogen.
3. **Sectoral organisation:** sectors represented by well-resourced branch associations, such as VNCI and FME/VNMI, can coordinate more easily than fragmented sectors.
4. **Geographic distribution:** some sectors are located near existing industrial infrastructure, while others are more dispersed.
5. **Capital intensity and investment horizon:** capital-intensive sectors with long investment cycles face stronger lock-in risks and therefore have stronger incentives to secure future hydrogen access.

Table 5 applies these factors to the main Cluster 6 sectors. The sector assessment is based on the LCES report (Stichting Cluster 6, 2025).

Table 5: Sector-level assessment of Cluster 6 firms' likelihood to coordinate for a collective hydrogen connection.

Sector (assoc.)	Hydrogen relevance	Competing options	Sectoral organisation	organisational capacity	Coordination willingness	H ₂ subst. rate
Specialty chemicals (VNCI)	High; H ₂ already used as feedstock; clear pathway	Limited for feedstock use; partial for heat	Strong (VNCI, well-funded, technically capable)		High. Strong incentives, organisational capacity and existing H ₂ expertise; likely first movers in a collective connection.	70%
Metallurgy (FME/VNMI)	High; H ₂ -DRI pathway mature; large per-site demand	Limited for high-temperature processes	Strong (FME federation; sector-specific lobby capacity)		High. Few but large sites; high per-site volumes make a collective connection attractive even with moderate dispersion.	60%
Glass (VNG)	Medium-high; oxy-H ₂ combustion demonstrated; burner adaptations needed	Limited; electric melting only partly applicable	Medium; small association and limited number of sites		Medium-high. Few large sites with strong technical need; willingness depends on whether neighbouring sectors join the same cluster.	50%
Ceramics (KNB)	Medium; pilots underway; full substitution requires kiln redesign	Limited	Medium; small but active association		Medium. Geographically dispersed; coordination depends strongly on regional aggregation with other sectors.	50%
Paper & cardboard (VNP)	Medium; high-pressure steam; H ₂ one of several routes	Strong; electrification credible for many processes	Medium-strong; concentrated sector with active association		Medium. Hydrogen competes with electrification; willingness depends on relative cost trajectories.	40%
Food processing (FNLI)	Low-medium; mainly peak heat; many low-temperature processes	Strong; electrification often preferred	Weak-fragmented; many heterogeneous firms		Low-medium. Many small sites with low per-site demand; electrification often more attractive; coordination costs high.	30%
Waste processing (VA)	Low; auxiliary heat use only; waste combustion limits substitution	Medium; own combustion sources	Medium; sector association with regulatory focus		Low. Limited material incentive; unlikely to drive collective connections but may join if geographically convenient.	20%
Data centres (NLDigital)	Very low; electricity-dominated; H ₂ mainly relevant for backup	Very strong; process is electricity-based	Strong, but oriented toward electricity-grid issues		Very low. Hydrogen is not a primary energy carrier for this sector; should not be included as core demand in clustering.	10%
Oil & gas onshore (Element NL)	Low-medium; partial process-heat substitution	Medium	Strong; well-organised and technically capable	well-organised and technically capable	Medium. Technical capacity is high, but strategic interest depends on the sector's own transition pathway.	30%

The sector analysis shows that Cluster 6 demand cannot be treated as equally reliable across all firms. Specialty chemicals, metallurgy, glass and ceramics are the most plausible anchor sectors for collective hydrogen connections. They combine higher hydrogen relevance, fewer credible alternatives and stronger incentives to coordinate. Food processing, paper, waste processing, data centres and onshore oil and gas are more uncertain. They may join a collective connection when geographically convenient, but they are less likely to drive one.

3.5 How does this affect the clustering model?

The actor analysis has direct consequences for the clustering model. First, it confirms that the Cluster 6 connection problem is a coordination problem. Individual firms are unlikely to justify expensive backbone connections on their own, while Hynetwork needs credible aggregated demand before extensions become attractive. The clustering model addresses this problem by turning dispersed company-level demand into regional demand clusters.

Second, the sector analysis justifies the use of actor-informed demand labels. Building on Table 5, Cluster 6 firms can be divided into two broad groups:

- **Anchor sectors:** specialty chemicals (VNCI), metallurgy (FME/VNMI), glass (VNG) and, conditionally, ceramics (KNB). These sectors combine relatively high hydrogen substitution potential, limited electrification alternatives and stronger coordination capacity.
- **Marginal sectors:** paper (VNP), food processing (FNLI), waste processing (VA), data centres (NLDigital) and onshore oil and gas (Element NL). These sectors face stronger alternatives, lower per-site hydrogen relevance or more uncertain participation.

This distinction matters for interpreting the network results. A cluster that contains at least one anchor-sector site is more institutionally plausible than a cluster that depends mainly on marginal-sector demand. Conversely, a low-cost network layout that relies heavily on marginal demand may be technically efficient but less robust from an actor perspective.

Figure 3 summarises the strategic position of the main actors. The problem owner has high interest but limited formal power. Hynetwork, Gasunie, the Ministry and the future regulator have more power because they control infrastructure, mandates, regulation or funding. The split between anchor and marginal sectors also changes the interpretation of Cluster 6 firms: anchor sectors have higher interest and are more likely to support collective connections, while marginal sectors are more uncertain.

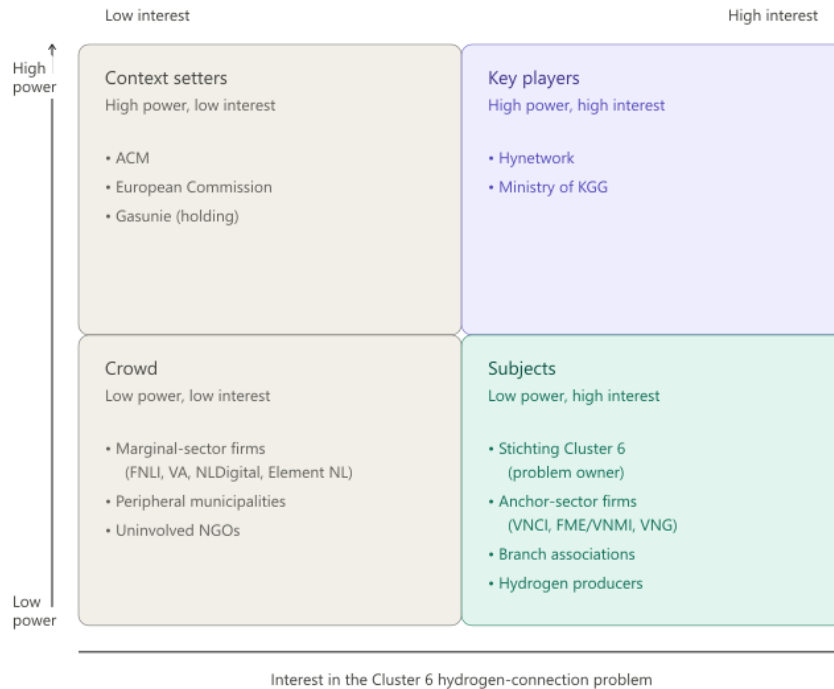


Figure 3: Power-interest grid for the Cluster 6 hydrogen-connection problem.

The modelling implications are therefore as follows:

- Clusters containing anchor-sector firms are more likely to represent credible collective connection opportunities.
- Marginal-sector demand should be treated with caution, because these firms may choose electrification or other decarbonisation routes instead of hydrogen.
- The results should be read together with the institutional setting: even technically attractive clusters require coordination with Hynetwork, regional authorities, regulators and potential funding instruments.

In line with Enserink et al. (2022), this actor analysis should be interpreted as a snapshot rather than a fixed reality. Actor positions may change as regulation develops, hydrogen costs change and regional infrastructure planning becomes more concrete. Nevertheless, the main conclusion is clear: spatial clustering is not only a technical aggregation method, but also a way to create credible collective demand in a fragmented institutional setting.

4 Methodology

This section describes how the research questions are answered. The methodology combines company-level data collection, actor-informed demand classification, spatial clustering, network optimisation and cost evaluations. The aim is to analyse how different ways of clustering dispersed Cluster 6 demand affect the design and cost of hydrogen infrastructure. In addition, the method assesses whether the resulting clusters contain credible hydrogen demand from an actor perspective.

4.1 Research Design

This study develops a modelling framework to analyse how different levels of spatial aggregation affect the connection of Cluster 6 companies to the national hydrogen backbone.

When companies are grouped into many small clusters, internal transport distances within each cluster are relatively short. This can reduce the cost of the internal network that connects companies within a cluster. However, more clusters also means that more separate connections to the hydrogen backbone may be required.

When companies are grouped into fewer larger clusters, the number of potential backbone connections is reduced. Demand is bundled into larger regional groups, which may make a collective connection more attractive. However, larger clusters also increase the internal distance over which demand has to be connected. This can increase internal network length and cost.

The methodology evaluates this trade-off by comparing a range of clustering configurations. The ONLT is selected because, unlike distance-based methods such as a minimum spanning tree, it accounts for both pipeline length and transport capacity and can introduce intermediate Steiner nodes to reduce costs. It therefore provides a suitable and computationally efficient method for comparing the internal network costs of many clustering configurations. Possible connection costs to the hydrogen backbone and actor-informed indicators are added afterwards.

4.2 Actor-informed Demand Classification

The actor analysis shows that Cluster 6 firms should not be treated as homogeneous hydrogen users. Some sectors have a stronger technical need for hydrogen and are more likely to support collective infrastructure development. Other sectors have more uncertain hydrogen demand because electrification or other decarbonisation routes may be more suitable.

For this reason, each company is labelled as either an anchor-demand or marginal-demand firm. This label is used after clustering to assess whether the resulting clusters contain credible hydrogen demand.

Let $\mathcal{N}$ denote the set of company nodes. Each company node $i \in \mathcal{N}$ is described as:

$$n_i = (x_i, y_i, D_i, s_i, a_i), \quad \forall i \in \mathcal{N}. \quad (1)$$

where x_i and y_i are the spatial coordinates of company i , D_i is the estimated hydrogen demand, s_i is the sector classification, and a_i is the actor-informed demand label.

For the actor-informed demand label, two labels are used:

- **Anchor demand:** sectors with relatively high hydrogen relevance and stronger coordination potential. This includes specialty chemicals (VNCI), metallurgy (FME/VNMI), glass (VNG) and ceramics (KNB).
- **Marginal demand:** sectors with lower hydrogen relevance, stronger electrification alternatives or more uncertain participation. This includes food processing (FNLI), paper and cardboard (VNP), waste processing (VA), data centres (NLDigital) and onshore oil and gas (Element NL).

4.3 Spatial Clustering Method

The company locations are grouped using the K-means clustering method (MacQueen, 1967). K-means partitions a set of observations into a predefined number of clusters. In this study, companies are clustered solely on the basis of geographic proximity, and each company is assigned to one cluster.

The algorithm begins with a predefined number of cluster centres, referred to as centroids. Each company is assigned to the nearest centroid based on geographic distance. The centroid of each cluster is then updated to the mean location of the companies assigned to it. This process continues until the cluster assignments stabilise or the maximum number of iterations is reached.

K-means is selected because it provides a consistent and computationally efficient way to vary the level of spatial aggregation. This is important because the clustering procedure is repeated for many values of k . The resulting centroid-based clusters can also be used directly as input for the subsequent network analysis.

Other clustering methods could produce different spatial configurations. Hierarchical clustering could illustrate how individual sites merge into larger regional groups, while DBSCAN could identify areas with high concentrations of industrial sites without requiring a predefined number of clusters. Demand-weighted or capacitated methods could include hydrogen demand directly, and network-constrained clustering could account for infrastructure corridors or distance to the hydrogen backbone.

These alternatives are not applied because the purpose of this study is not to compare clustering algorithms. Instead, the analysis focuses on how changing the number of spatial clusters affects infrastructure design and cost. K-means is therefore used as a transparent and consistent method for generating comparable clustering configurations. The implications of this methodological choice are discussed in the limitations and recommendations for future research.

The number of clusters is varied from $k = 20$ to $k = 140$. Lower values of k represent fewer and larger regional clusters, while higher values represent more numerous and smaller clusters.

Because K-means can be sensitive to the initial placement of the centroids, each selected clustering configuration is generated using multiple random initialisations. The resulting configurations are compared to assess whether the findings are stable or depend strongly on a specific initialisation.

For each value of k , the following attributes are calculated for every cluster:

- total estimated hydrogen demand;
- total anchor demand;
- total marginal demand;
- presence of at least one anchor-sector company;
- centroid location;
- number of companies in the cluster.

These cluster-level attributes are subsequently used as input for the network modelling and actor-informed evaluation.

4.4 Network Modelling with the ONLT

After the clustering configurations have been created, the Optimal Network Layout Tool (ONLT) is used to determine cost-efficient internal network layouts for each configuration (Heijnen, 2025). The ONLT is applied to estimate the cost of connecting the companies within each cluster.

The model does not determine the full national hydrogen backbone. Instead, it focuses on the internal network structure that would be needed to connect an individual cluster to a collective connection point to the hydrogen backbone. This makes it possible to calculate many configurations without the study becoming computationally intensive.

4.4.1 Graph Representation

Graph theory forms the basis for the network model. A network is represented as a graph $G = (N, E)$, where N is the set of nodes and E is the set of edges. In this study, nodes represent industrial demand locations or possible intermediate network points. Edges represent possible pipeline connections.

Each edge has a length and a required capacity. The length reflects the distance between connected nodes, while the capacity reflects the hydrogen flow that must be transported through that edge. The cost of an edge depends on both distance and capacity.

A simple graph only describes the connection between nodes. A weighted graph assigns a value to each edge, such as distance or cost. This study uses weighted graphs.

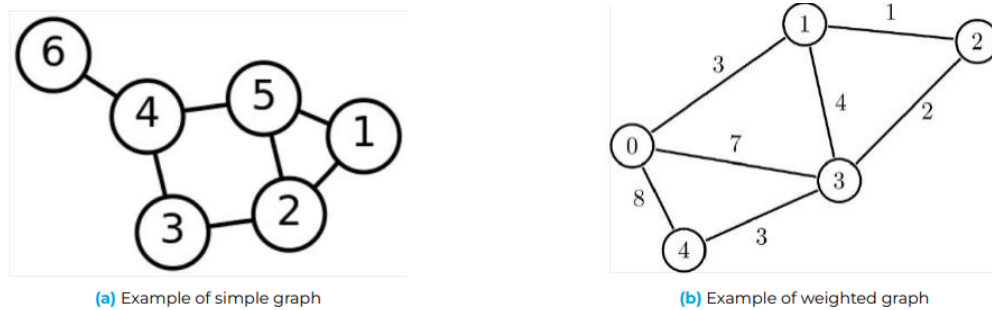


Figure 4: Examples of a simple graph and a weighted graph

A common graph-theory problem is finding the most efficient way to connect all nodes. A minimum spanning tree connects all demand nodes while minimising total network length. However, it does not account for differences in hydrogen flow, pipeline capacity or economies of scale. Consequently, the shortest network is not necessarily the least-cost network. The ONLT is therefore more suitable for this study because it optimises the network based on both pipeline length and required transport capacity. It can also reconfigure the initial spanning tree and introduce intermediate Steiner nodes when these reduce total costs. This produces a more realistic cost-based comparison of the different clustering configurations than a minimum spanning tree alone.

4.4.2 Optimal Network Layout Tool

The ONLT is a Python-based model that identifies near-optimal network structures by combining graph-theoretical principles with a cost-based optimisation approach. The model searches for a minimum-cost network that connects multiple demand nodes.

The ONLT determines the network layout through four sequential steps:

1. **Minimum spanning tree:** the initial network connects all nodes using the shortest total length.
2. **Minimum-cost spanning tree:** the network is reconfigured to reduce total cost while taking capacity into account.
3. **Steiner tree improvement:** additional intermediate nodes may be introduced when they reduce total network cost.

4. **Final improvement round:** earlier improvement steps are repeated until no further cost reduction is found.

These steps are illustrated in Figure 5, where the network develops from a distance-based solution to a cost-optimised configuration.

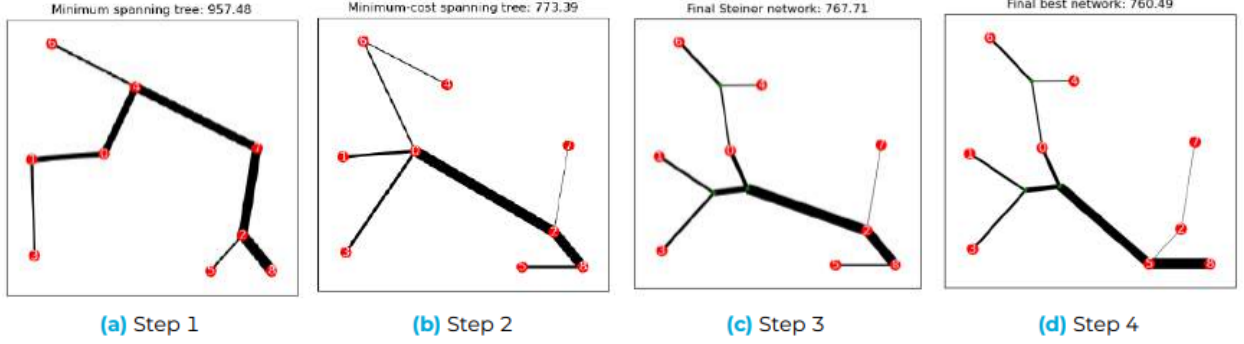


Figure 5: Overview of the ONLT optimisation steps. Adapted from the ONLT user guide

The ONLT produces tree-like network structures without redundancy. This means that each node is connected through a single path. The resulting layouts are useful for comparing network cost and capacity requirements, but they do not include backup routes or security-of-supply redundancy.

4.4.3 Cost function

The ONLT evaluates network layouts using a cost function that depends on pipeline length and required capacity. For each edge $(i, j) \in E_c$, where E_c denotes the set of edges in the network of cluster c , the cost is expressed as:

$$C_{ij} = \alpha \cdot L_{ij} \cdot Q_{ij}^{\beta}, \quad (2)$$

where α is the cost coefficient, L_{ij} is the pipeline length between nodes i and j , Q_{ij} is the required transport capacity, and β is the capacity-cost exponent.

The total ONLT cost for cluster c is calculated as:

$$C_{\text{ONLT},c} = \sum_{(i,j) \in E_c} \alpha \cdot L_{ij} \cdot Q_{ij}^{\beta}. \quad (3)$$

where E_c is the set of edges in the network of cluster c .

The value of β determines the strength of economies of scale in pipeline construction. If β is lower than one, additional capacity is relatively cheaper than building a completely separate pipeline. This makes bundled transport more attractive. Based on the cost relationship reported by Lafeber (2024), a value of $\beta = 0.42$ is adopted as the baseline parameter in this study.

4.5 Evaluation and Comparison

The clustering and network modelling results are compared across all values of k using an actor-informed evaluation framework. The actor analysis shows that a technically cost-efficient network configuration is not necessarily credible from an implementation perspective. The feasibility of a collective connection also depends on whether demand is sufficiently concentrated, whether firms with a strong interest in hydrogen are present and whether the required coordination with Hynetwork remains manageable.

These actor-related conditions are translated into measurable indicators. For each clustering configuration, the following indicators are evaluated:

- total ONLT network cost without T-junction costs, representing the cost of connecting firms within the clusters;
- total cost including T-junction costs, representing both internal network development and connection to the national hydrogen backbone;
- total internal network length, indicating the spatial effort required to aggregate dispersed demand;
- number of required backbone connections, reflecting the number of separate connection processes that must be coordinated with Hynetwork;
- total estimated hydrogen demand per cluster, indicating the scale and visibility of the collective demand;
- number of anchor-sector companies per cluster;
- share of cluster demand represented by anchor-sector companies;
- number of clusters containing at least one anchor-sector company;
- number of clusters that both contain anchor-sector demand and meet the specified demand threshold.

The first evaluation step compares internal network costs without T-junction costs. This shows how the level of spatial aggregation affects the cost and network length required to connect companies within each cluster.

The second step includes T-junction costs. The number of required backbone connections is not only a cost indicator, but also an actor-related coordination indicator. Each additional connection requires interaction between participating firms, Stichting Cluster 6, Hynetwork and relevant public authorities. Configurations with many clusters may therefore create a larger institutional and organisational burden, even when their internal networks are relatively short.

The third step evaluates the credibility of the clustered demand using the actor-informed company classification. Clusters containing anchor-sector companies are considered more plausible starting points for collective hydrogen infrastructure because these companies have a stronger technical need for hydrogen, fewer competing decarbonisation alternatives and a greater expected willingness to coordinate. Clusters that depend mainly on marginal-sector demand are interpreted more cautiously, since participation in a hydrogen network is more uncertain.

The comparison therefore does not identify a preferred configuration based on cost alone. Instead, it evaluates the balance between technical efficiency, the number of required institutional interfaces and the presence of credible collective demand. A configuration is considered more relevant when it limits internal network costs and backbone connection requirements while maintaining sufficient demand and anchor-sector participation within most clusters.

4.6 Robustness

Several checks are used to assess the robustness of the results.

First, the input data is checked manually. Company coordinates obtained through Google Maps are inspected to reduce the risk of assigning demand to head offices, administrative locations or unrelated sites. Sector classifications are checked against the Cluster 6 sector associations.

Second, the robustness of the clustering procedure is assessed. Since K-means depends on the initial placement of centroids, clustering configurations are repeated using different random initialisations. The resulting variation is used to determine whether the main cost trends and preferred clustering configurations remain stable.

Third, the sensitivity to T-junction costs is tested by varying the assumed connection cost. This shows whether the preferred clustering level is primarily driven by internal network costs or by the number of required backbone connections.

Fourth, demand threshold sensitivity is evaluated. This determines whether the selected clustering configurations remain relevant when different minimum demand requirements are applied.

Fifth, actor-informed robustness is assessed by examining the presence of anchor-sector companies within the selected clustering configurations. Clusters containing anchor-sector companies are interpreted as more credible from an implementation perspective than clusters that rely exclusively on marginal demand.

Together, these checks assess whether the main findings remain consistent under changes in clustering initialisation, backbone connection costs, demand thresholds and actor-informed cluster composition.

4.7 Assumptions and Limitations

Several assumptions are important when interpreting the results.

First, the study assumes that hydrogen can be supplied through the national hydrogen backbone and that pipeline connections are the relevant infrastructure option for the analysed clusters. Alternative supply options, such as local electrolysis, truck delivery or hydrogen carriers, are not included in the model.

Second, potential hydrogen demand is estimated using sector-level natural gas consumption and assumed hydrogen substitution rates. The resulting demand values should therefore be interpreted as sector-level approximations rather than company-specific demand forecasts.

Third, the spatial representation of demand is simplified. Companies are grouped using K-means clustering, while the aggregated demand of each cluster is represented by its centroid location. This does not account for all geographical, spatial-planning or routing constraints that may influence the location of an actual connection point.

Fourth, the network and cost modelling are simplified. The ONLT produces tree-like network structures, which are a plausible representation of cost-efficient regional connection networks during an early development stage. However, the model does not include alternative supply routes or network redundancy. It therefore does not assess resilience to pipeline failures, maintenance interruptions or other security-of-supply requirements.

Fifth, the euro-equivalent cost conversion is based on a fixed pipeline diameter. The resulting costs should be interpreted as indicative order-of-magnitude estimates rather than detailed engineering cost estimates.

Finally, T-junction costs and actor-informed demand labels are treated as scenario-based assumptions. They support the comparison of clustering configurations, but should not be interpreted as exact project costs or fixed predictions of company behaviour. Actual participation will depend on factors such as hydrogen prices, alternative decarbonisation options, infrastructure availability, regulation and investment decisions.

The results should therefore be interpreted as indicative infrastructure-planning insights rather than precise investment estimates. The method is intended to compare clustering configurations and identify relevant trade-offs, not to provide final engineering designs for hydrogen infrastructure.

5 Data Collection and Preprocessing

This chapter describes the collection and preprocessing of the data used as input for the clustering and network model. It explains how industrial demand, company characteristics, the hydrogen backbone and the relevant cost assumptions are translated into a consistent model dataset.

5.1 Identification of Cluster 6 companies

Industrial demand is first represented at the level of individual companies. The relevant companies are identified using the Cluster 6 inventory from the Landelijke Cluster Energie Strategie (LCES) (Stichting Cluster 6, 2025). This dataset provides an overview of industrial production sites outside the five major Dutch industrial clusters.

For each company, geographic coordinates are obtained using Google Maps. The location of each company is manually checked to ensure that the selected coordinate corresponds to an industrial site rather than a head office, administrative address or unrelated location. This step is important because the clustering and network optimisation depend directly on the spatial location of the demand nodes.

Each company is assigned to one of the Cluster 6 sector categories shown in Table 6. These sector labels are used for two purposes. First, they support the estimation of potential hydrogen demand. Second, they are used to classify companies as either anchor demand or marginal demand in the actor-informed analysis.

Table 6: Branch associations representing Cluster 6 industries

Association	Sector	Description
FNLI	Food processing	Federatie Nederlandse Levensmiddelen Industrie
VNP	Paper and cardboard	Vereniging van Nederlandse Papier- en Kartonfabrieken
VNCI	Specialty chemicals	Vereniging van de Nederlandse Chemische Industrie
FME / VNMI	Metallurgical industry	Base metals and foundries
VNG	Glass manufacturing	Vereniging Nederlandse Glasfabrikanten
KNB	Ceramics and building ceramics	Koninklijke Nederlandse Bouwkeramiek
VA	Waste processing	Vereniging Afvalbedrijven
NLDigital	Data centres and ICT	Digital infrastructure sector
Element NL	Oil and gas onshore	Formerly NOGEPA

5.2 Geographic Data Collection

After identifying the relevant companies, geographic coordinates are collected for each site. Coordinates are obtained using Google Maps by searching for the company name, combined where needed with additional location information such as municipality, industrial estate or production site name. Each location is manually checked to ensure that the selected point corresponds to an industrial facility.

The company locations are stored as geocoded point data in the input dataset, with latitude and longitude coordinates recorded for each company. These coordinates are converted into a GeoPandas spatial dataset using the WGS84 coordinate reference system (EPSG:4326). For spatial calculations, including clustering, distance calculations and network-layout preparation, the point geometries are reprojected to the Dutch national projected coordinate reference system, Rijksdriehoekscoördinaten / RD New (EPSG:28992). This ensures that distances are calculated in metres rather than angular degrees resulting in figure 6.

Geocoded Cluster 6 company locations by sector

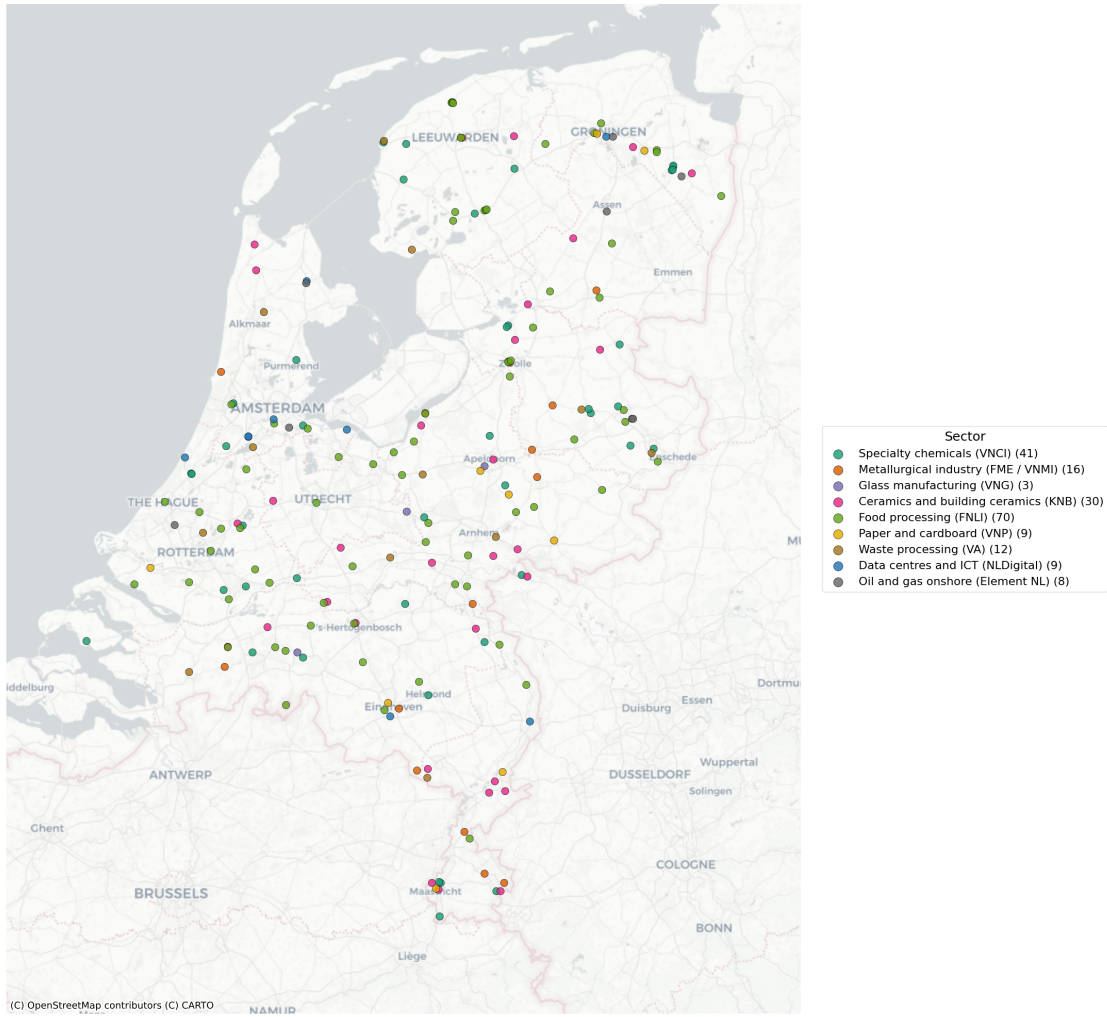


Figure 6: Geocoded Cluster 6 company locations

5.3 Sector Classification

Each company is assigned to one of the Cluster 6 sector categories. The sector classification follows the sector associations used in the LCES. These associations represent the main industrial branches within Cluster 6 and are used as the link between company-level locations and sector-level demand assumptions.

Table 7 shows the sector categories used in the dataset.

Table 7: Sector categories used for Cluster 6 company classification

Association	Sector	Role in dataset
VNCI	Specialty chemicals	Anchor demand
FME / VNMI	Metallurgical industry	Anchor demand
VNG	Glass manufacturing	Anchor demand
KNB	Ceramics and building ceramics	Anchor demand
VNP	Paper and cardboard	Marginal demand
FNLI	Food processing	Marginal demand
VA	Waste processing	Marginal demand
NLDigital	Data centres and ICT	Marginal demand
Element NL	Oil and gas onshore	Marginal demand

5.4 Hydrogen Demand Estimation

Company-level hydrogen demand is not directly available for all Cluster 6 sites. Therefore, potential hydrogen demand is estimated using sector-level natural gas consumption data from CBS StatLine (Statistics Netherlands (CBS), 2026) and allocated equally across the identified companies in each sector.

The demand estimation follows three steps. First, sector-level natural gas consumption is collected. Second, this demand is allocated to the companies belonging to each sector. Third, a sector-specific hydrogen substitution rate is applied to estimate the share of natural gas demand that could realistically be replaced by hydrogen.

Company-level hydrogen demand is estimated by applying a sector-specific hydrogen substitution rate to the allocated natural gas demand. For each company $i \in \mathcal{N}$, let s_i denote the sector to which the company belongs. The estimated hydrogen demand is calculated as:

$$D_i = P_{s_i} \cdot G_i, \quad \forall i \in \mathcal{N}, \quad (4)$$

where D_i is the estimated hydrogen demand of company i , G_i is the natural gas demand allocated to company i , and P_{s_i} is the hydrogen substitution rate associated with sector s_i , as presented in Table 5.

The baseline substitution rates are based on the technological relevance of hydrogen per sector and the actor analysis. Sectors with high-temperature processes or feedstock applications receive higher substitution rates, while sectors with stronger electrification alternatives receive lower values.

5.5 Actor-informed Demand Labels

The actor analysis shows that Cluster 6 companies should not be treated as a homogeneous group. Some sectors have a clear technical need for hydrogen and stronger incentives to participate in collective infrastructure development. Other sectors have lower hydrogen relevance or more attractive alternatives, such as electrification. To reflect this difference, each company node is assigned an actor-informed demand label.

Two categories are used. The first category is anchor demand. This includes sectors with relatively high hydrogen relevance, stronger coordination potential and greater likelihood of participating in a collective hydrogen connection. The anchor sectors are specialty chemicals, metallurgy, glass and, conditionally, ceramics.

The second category is marginal demand. This includes sectors where hydrogen demand is more uncertain because of competing decarbonisation pathways, lower process-temperature requirements or weaker coordination incentives. The marginal sectors are food processing, paper and cardboard, waste processing, data centres and onshore oil and gas.

This actor-informed classification is used as an interpretive layer after the spatial clustering and network optimisation. All estimated demand is included in the main network model. The anchor and marginal labels are then used to assess whether the resulting clusters are supported by firms with a more credible role for hydrogen. In this way, the actor-informed classification does not form a separate network configuration, but helps to evaluate the implementation credibility of the clustering outcomes.



Figure 7: Actor-informed classification of Cluster 6 demand nodes

5.6 Hydrogen Backbone Data

In addition to demand nodes, the model requires information on the national hydrogen backbone. The backbone represents the main supply infrastructure to which Cluster 6 clusters may connect. Because the exact final routing and connection availability of the backbone remain uncertain, the backbone is represented in a simplified spatial form based on publicly available information from Hynetwork and Gasunie (Hynetwork Services, 2025).

The backbone data is represented as a line dataset. In the model, clusters are not restricted to a fixed set of predefined connection points. Instead, each cluster is allowed to connect to the nearest point on the backbone line. This means that the connection location is determined by calculating the clusters centre and then the closest point to the hydrogen backbone.

Rather than assuming that only specific stations or terminals can be used, the model estimates the shortest

spatial connection between clustered demand and the backbone. The resulting connection point is therefore a modelled point on the backbone line, not a fixed existing location.

The distance between clustered demand and the nearest point on the backbone line is an important variable in the network model. It affects both the internal network layout and the cost of connecting clustered demand to the national hydrogen system. Figure 8 shows the simplified backbone representation used in the model.



Figure 8: Simplified representation of the national hydrogen backbone and Cluster 6 company locations

5.7 Cost Conversion

The ONLT produces relative cost values rather than costs in euros. To support interpretation and comparison, an indicative conversion factor was derived using the pipeline cost formulation of Oostdijk (Oostdijk, 2025).

The conversion was calibrated using the ($k = 70$) configuration, which has an ONLT cost of (743,681) and a total network length of (1,292.59) km. Applying the cost formulation with a fixed pipeline diameter of (0.508) m resulted in an indicative pipeline cost of (70.89) million euros. The conversion factor is therefore:

$$\gamma = \frac{70,892,239}{743,681} \approx 95.33 \text{ euro per ONLT cost unit.} \quad (5)$$

For each configuration, the indicative euro-equivalent cost is calculated as:

$$C_{\text{euro}}(k) = \gamma \cdot C_{\text{ONLT}}(k). \quad (6)$$

The conversion is used solely to place the relative ONLT outcomes on an approximate euro scale. The resulting values should therefore be interpreted as indicative estimates for comparing configurations, rather than as precise investment costs.

5.8 Backbone Connection Cost

In addition to the internal network cost, each cluster may require a connection to the hydrogen backbone. This is represented through a T-junction cost, denoted by C_T . The T-junction cost captures the additional infrastructure cost associated with creating a collective connection point to the backbone.

For each clustering configuration, the total cost including T-junction costs is calculated as:

$$C_{\text{total}}(k) = C_{\text{euro}}(k) + k \cdot C_T. \quad (7)$$

Here, k represents the number of clusters and therefore the maximum number of collective connection points. This formulation assumes that each cluster requires one connection to the backbone. The T-junction cost is varied across configurations to test how sensitive the preferred clustering configuration is to different connection cost assumptions.

It is important to note that the cost analysis takes a network-planning perspective. The purpose of this analysis is to compare clustering configurations rather than estimate exact project costs, only the order of magnitude of the infrastructure costs is required. This is why only a single ONLT-to-euro conversion factor is therefore applied to all configurations. This places the results on an indicative euro scale without changing their relative ranking.

The calculated cost is not the total cost of the hydrogen system. Hydrogen production, the hydrogen price, storage, financing and company-specific conversion costs are not included. The results are therefore only meant to show which network configurations are relatively cost-efficient, not whether they are commercially attractive for every individual firm.

6 Model Setup and Configurations

This chapter translates the methodology from Chapter 4 and the input data from Chapter 5 into a set of model configurations. The chapter has two purposes. First, it defines how clustered industrial demand is represented as input for the Optimal Network Layout Tool (ONLT). Second, it explains how different clustering configurations are compared. The cost parameters used in this comparison are described in Section 5.7 and Section 5.8.

The model should be interpreted as a comparative planning tool rather than as an exact infrastructure design. Its purpose is not to determine the final layout of the hydrogen infrastructure for Cluster 6, but to make the trade-off between different clustering configurations visible.

6.1 Clustering Configurations

The main design variable in this study is the number of spatial clusters, denoted by k . A low value of k creates a small number of large clusters. This reduces the number of collective backbone connection points, but increases the internal distances over which demand must be aggregated. A high value of k creates more, smaller clusters. This reduces internal transport distances, but increases the number of possible connection points to the backbone.

The model is therefore run for multiple values of k . Comparing these runs makes it possible to assess whether additional clusters continue to reduce internal network cost, or whether a point of diminishing returns is reached.

6.2 ONLT Implementation

The Optimal Network Layout Tool is used to calculate a cost-efficient network layout for each clustering configuration. The ONLT compares possible network structures and searches for a layout that balances distance, capacity and economies of scale. In this study, the ONLT output is used comparatively. The goal is not to produce a final engineering design, but to evaluate how different levels of spatial clustering affect infrastructure requirements.

For every value of k , the model records the total ONLT cost, C_{ONLT} , and the total length of the resulting network, L_{network} . These values are compared across clustering configurations. The ONLT cost is treated as a relative cost indicator. The cost conversion in Section 5.7 is then used to translate this indicator into an indicative euro value.

6.3 Cluster-to-Backbone Connection

Each cluster is connected to the point on the backbone that is closest to its centroid. The centroid is used because it represents the demand-weighted centre calculated by K-means. This keeps the connection rule consistent for all values of k .

The selected connection point is added as a node in the ONLT run. As a result, the pipeline from the cluster to the backbone is optimised together with the internal cluster network and is included in C_{ONLT} . This is important because the distance to the backbone changes when the number of clusters changes, just as the internal pipeline distances do. Including this distance in the optimisation ensures that its effect is captured in the comparison.

This approach is a modelling choice rather than a fixed design rule. Other options are possible. For example, the connection point could be placed at the company closest to the backbone, several clusters could share one connection point, or elongated clusters could use multiple connection points. Comparing these alternatives would help determine how sensitive the results are to the chosen connection rule and would be a useful topic for further research.

6.4 Configuration Design

The model is evaluated under multiple configuration settings. These configurations are used to test whether the preferred clustering configuration is robust under different assumptions. The first configuration dimension is the clustering level k . This directly tests the trade-off between fewer large clusters and more small clusters. The second dimension is the T-junction cost. This tests how strongly the preferred clustering level depends on the cost of connecting clusters to the backbone. The third dimension concerns actor-informed demand interpretation. This is used to assess whether the clustered demand is supported by sectors with a credible role for hydrogen. The fourth dimension is the demand threshold, which tests whether clusters contain enough demand to justify a collective connection.

Table 8: configuration dimensions used in the model setup.

configuration dimension	Tested values	Purpose
Clustering level k	$k = 20$ to $k = 140$	Tests the effect of spatial aggregation on network length and cost.
Selected configurations	$k = 34$, $k = 70$	Used for detailed comparison of cost, actor composition and threshold sensitivity.
T-junction cost C_T	€150.000 to €1,000,000 per connection	Tests how sensitive the preferred clustering level is to connection cost assumptions.
Actor-informed configurations	Anchor and marginal demand labels	Tests whether clusters are supported by credible hydrogen demand.
Demand threshold	Low, medium and high threshold	Tests whether clusters contain enough demand to justify a collective connection.

6.5 Output Indicators

The results are evaluated using a small set of indicators. These indicators are selected because they directly relate to the research question, but also because they correspond to the decision criteria of the actors identified in Chapter 3. Different actors evaluate the same clustering configuration on different grounds. A configuration that is attractive from a network-planning perspective is not necessarily attractive for an individual firm, and a configuration with credible aggregated demand is not necessarily the cheapest one. Making this explicit is important, because the model does not produce a single optimum but a trade-off between criteria that are weighted differently by different actors.

The first indicator is the total ONLT cost, C_{ONLT} , which shows how the internal network cost changes across clustering configurations. The second indicator is the euro-equivalent network cost, C_{euro} , calculated using Equation 6. The third indicator is the total network length, L_{network} , which shows how much pipeline infrastructure is required. The fourth indicator is the number of clusters, which represents the maximum number of collective connection points. The fifth indicator is the total cost including T-junction costs, C_{total} , calculated using Equation 7.

In addition to these cost and network indicators, the results are interpreted using actor-informed indicators. These include the number of clusters that contain at least one anchor-demand firm and the number of clusters that meet minimum demand thresholds. Together, these indicators make it possible to assess the central trade-off of this thesis: whether dispersed industrial demand is better represented as a small number of large collective clusters or as a larger number of smaller regional clusters.

Table 9 links each indicator to the actors for which it is most relevant, based on the resources and interests.

Table 9: Link between output indicators and actor decision criteria.

Indicator	Primary actor	Why it matters for this actor
Total network cost C_{euro}	Hynetwork, Ministry of KGG	Determines whether extending infrastructure towards dispersed demand is defensible as a public investment.
Number of clusters	Hynetwork	Each cluster is a separate connection request, study trajectory and operational connection point.
Total cost including T-junctions C_{total}	Hynetwork, ACM	Relevant for total system cost and for future tariff and cost-allocation decisions.
Network length L_{network}	Provinces / IPO	Longer networks require more spatial planning, permits and routing coordination.
Demand per cluster D_c	Hydrogen producers, Hynetwork	Determines whether a connection point represents a volume worth supplying and contracting.
Anchor presence A_c	Stichting Cluster 6, Cluster 6 firms	Indicates whether a cluster has a firm with a technical reason to lead and co-fund a collective connection.
Cluster stability	Stichting Cluster 6	A stable cluster represents a durable regional grouping that can be used as a basis for coordination.
Feasible clusters under subsidy	RVO, Ministry of KGG	Shows how many clusters become viable under different levels of public support.

The table shows that the indicators are not equally important to all actors. Hynetwork is mainly sensitive to the number of connection points and total cost, which favours configurations with fewer clusters. Individual Cluster 6 firms are mainly sensitive to internal distance and to whether their cluster contains a credible anchor firm, which favours more detailed configurations. The Ministry and RVO are mainly interested in how many clusters become feasible under realistic support levels. This means that the trade-off analysed in this thesis is not only a technical one, but reflects a genuine difference in criteria between actors.

7 Results and Analysis

This chapter presents the results of the clustering and network optimisation analysis. It is important to recognize that the aim is not to determine an exact hydrogen infrastructure layout for Cluster 6, but to compare how different levels of spatial aggregation affect network structure, cost and demand credibility. The analysis first evaluates the baseline network outcomes without additional backbone connection costs. It then adds T-junction costs to examine how the cost trade-off changes when each cluster is assumed to require a separate cost to connect to the hydrogen backbone. After this cost-based comparison, the selected clustering configurations are further assessed using actor-informed indicators and demand-threshold sensitivity. Together, these results show how the preferred clustering level depends not only on network cost, but also on connection assumptions and the credibility of the underlying hydrogen demand.

7.1 Baseline Network Costs

The model is run for clustering configurations ranging from $k = 20$ to $k = 140$. For each value of k , the ONLT calculates the total network cost and the corresponding network layout. The total ONLT cost is first evaluated without additional T-junction costs. This shows the internal network cost of connecting the clustered demand nodes.

Figure 9 shows the total ONLT cost for each clustering configuration. The curve follows a clear U-shaped pattern. At low values of k , the total cost is relatively high. This is because a small number of clusters creates large aggregated demand areas, which increases internal transport distances. As k increases, the network can represent the dispersed demand structure more accurately, and the total ONLT cost decreases.

The lowest ONLT cost is observed at $k = 70$, with a cost of 743,681 and a total network length of approximately 1,292.59 km. This does not mean that $k = 70$ is a clear or unique optimum. As shown in Figure 9, the cost curve is relatively flat between $k = 40$ and $k = 90$. Within this range, several clustering configurations result in similar internal network costs.

The value at $k = 70$ should therefore be interpreted as the lowest point found in the model runs, rather than as a strong technical optimum. In practical terms, the model suggests that a broader range of medium-to-high clustering configurations performs similarly when only internal network cost is considered.

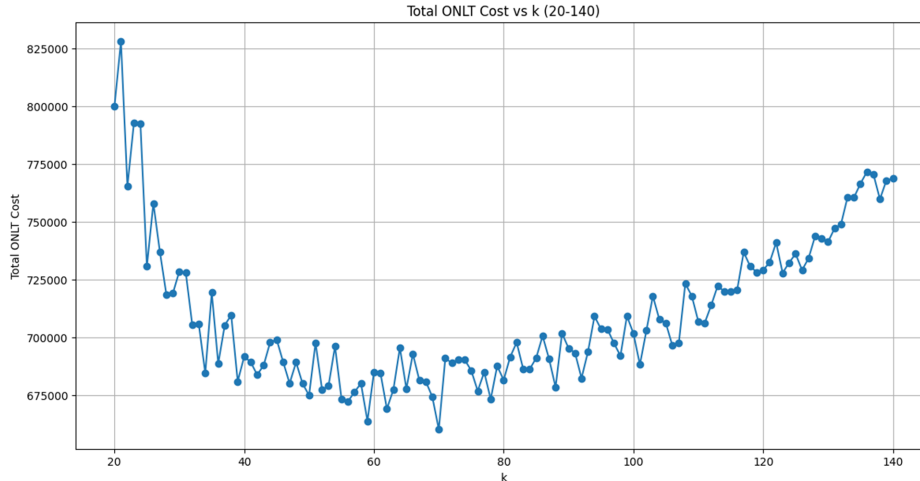


Figure 9: Total ONLT cost for clustering configurations between $k = 20$ and $k = 140$, excluding T-junction costs.

This finding is important for the interpretation of the later results. Internal network cost alone does not clearly distinguish between all clustering configurations. Although $k = 70$ gives the lowest observed ONLT

cost, configurations with fewer clusters can achieve comparable costs. This becomes especially relevant once T-junction costs and actor-informed feasibility are included in the coming chapters.

The network layout for $k = 70$ is visualised in Figure 10.



Figure 10: $k=70$ network configuration

7.2 Effect of Adding T-junction Costs

The baseline results in the previous section only include the network cost calculated by the ONLT. However, these results do not account for the additional cost of creating a collective connection point to the hydrogen backbone. In Hynetwork's terminology, such a connection is called a T-junction. It can be developed through a Valve Study Agreement. During the study phase, Hynetwork requires a guarantee of €150,000. The final connection cost is not known in advance. Instead, it is calculated separately for each project, based on the study results and the actual work needed to realise the connection (Hynetwork Services, 2026a).

This means that T-junction costs should not be treated as a single known unit cost. Instead, they are project-specific and depend on the location, capacity and technical configuration of the connection point. For this reason, this thesis treats the T-junction cost as a sensitivity parameter rather than as a fixed engineering estimate. This makes it possible to test how strongly the preferred clustering configuration depends

on different assumptions about the cost of connecting clusters to the hydrogen backbone.

To represent this effect, an additional cost term is added for each cluster. The total cost for each clustering configuration is calculated as:

$$C_{\text{total}}(k) = C_{\text{euro}}(k) + k \cdot C_T, \quad (8)$$

where $C_{\text{euro}}(k)$ is the euro-equivalent ONLT cost for clustering configuration k , and C_T is the assumed cost per T-junction. This means that higher values of k are penalised because they require more collective connection points.

This formulation is based on a simplifying assumption. The T-junction cost is treated as a fixed cost per cluster, meaning that each cluster receives the same connection penalty regardless of its demand size. In reality, connection costs are likely to depend on many other factors as well. The T-junction configurations should therefore be interpreted as an order-of-magnitude sensitivity analysis rather than precise connection-cost estimates.

Figure 11 shows the total cost for T-junction costs between 150,000 and 1,000,000, tested in steps of 50,000. For each value of C_T , the total cost is calculated by adding a penalty of $k \cdot C_T$ to the euro-equivalent ONLT cost. As a result, higher values of k are penalised more strongly when the assumed T-junction cost increases.

The figure shows that the effect of the T-junction penalty becomes stronger as k increases. This is because the penalty term is proportional to the number of clusters: configurations with more clusters receive a larger additional cost when C_T rises. As a result, the right-hand side of the curve becomes increasingly expensive for higher T-junction costs.

This changes the preferred clustering configuration. For tested T-junction costs between €150,000 and €500,000, the minimum remains at $k = 34$. At a T-junction cost of €600,000 and above, the penalty on configurations with more clusters becomes large enough to shift the optimum to a lower value of k . In this case, the cost-minimising configuration moves from $k = 34$ to $k = 25$. This shows that increasing connection costs favour fewer, larger clusters over more spatially detailed clustering configurations. This shows that increasing connection costs favour fewer, larger clusters over more spatially detailed clustering configurations.

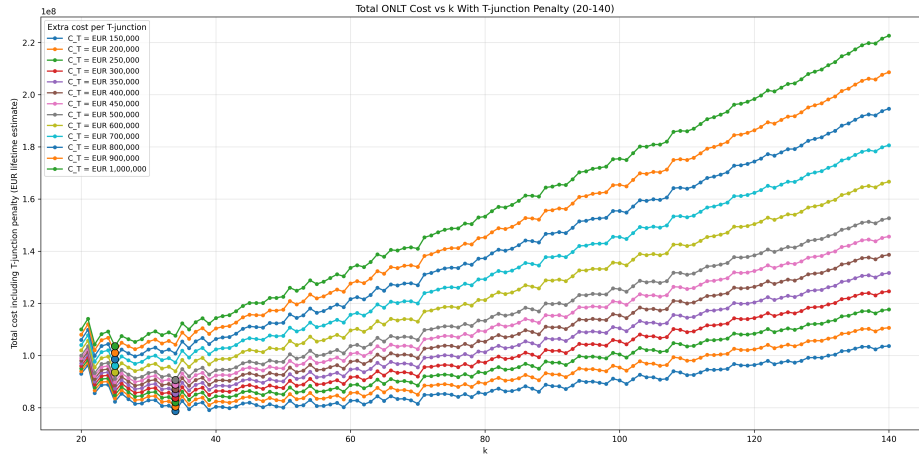


Figure 11: Total cost including T-junction penalties for C_T values between 150,000 and 1,000,000.

Figure 12 shows four representative cases in more detail. This makes the shift in the cost-minimising clustering configuration easier to interpret. Without T-junction costs, the minimum is found at $k = 70$. When a T-junction cost of 150,000 is added, the minimum shifts to $k = 34$. The same minimum remains

visible for $C_T = 500,000$. When the T-junction cost increases to 1,000,000, the minimum shifts further downward to $k = 25$.

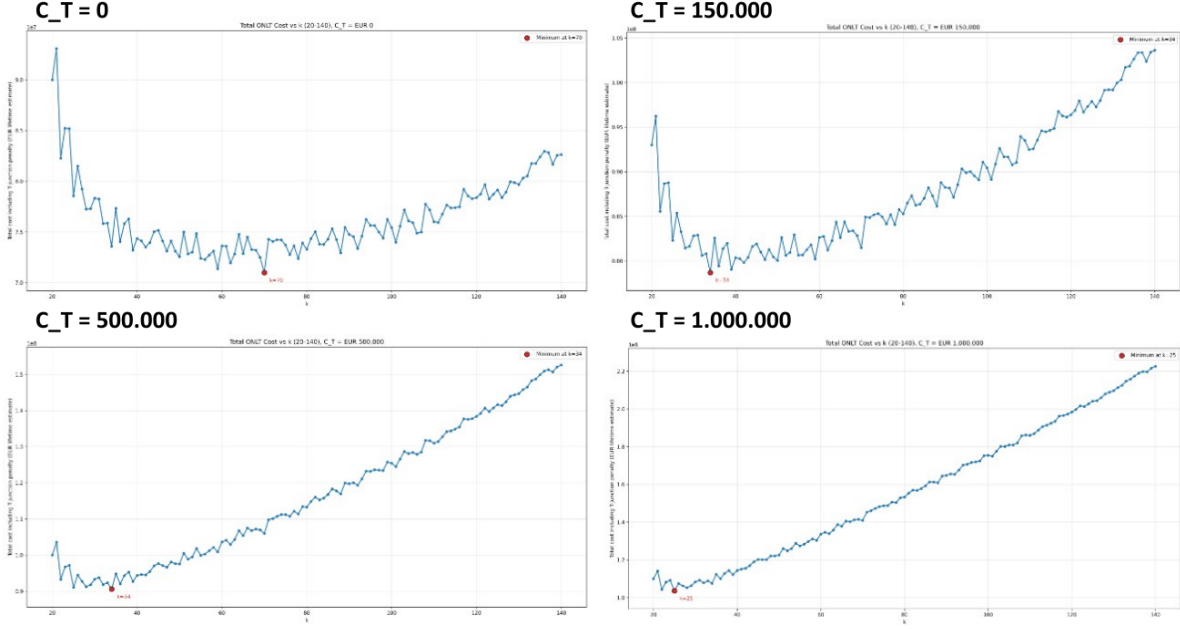


Figure 12: Total cost curves for four representative T-junction cost assumptions: $C_T = 0$, $C_T = 150,000$, $C_T = 500,000$ and $C_T = 1,000,000$.

Figure 13 summarises how the cost-minimising value of k changes when the T-junction penalty increases. For all tested values from 150,000 to 500,000, the optimal clustering configuration remains $k = 34$. When the penalty increases to 600,000 and above, the optimum shifts downward to $k = 25$. This shows that higher connection costs make smaller numbers of larger clusters more attractive.

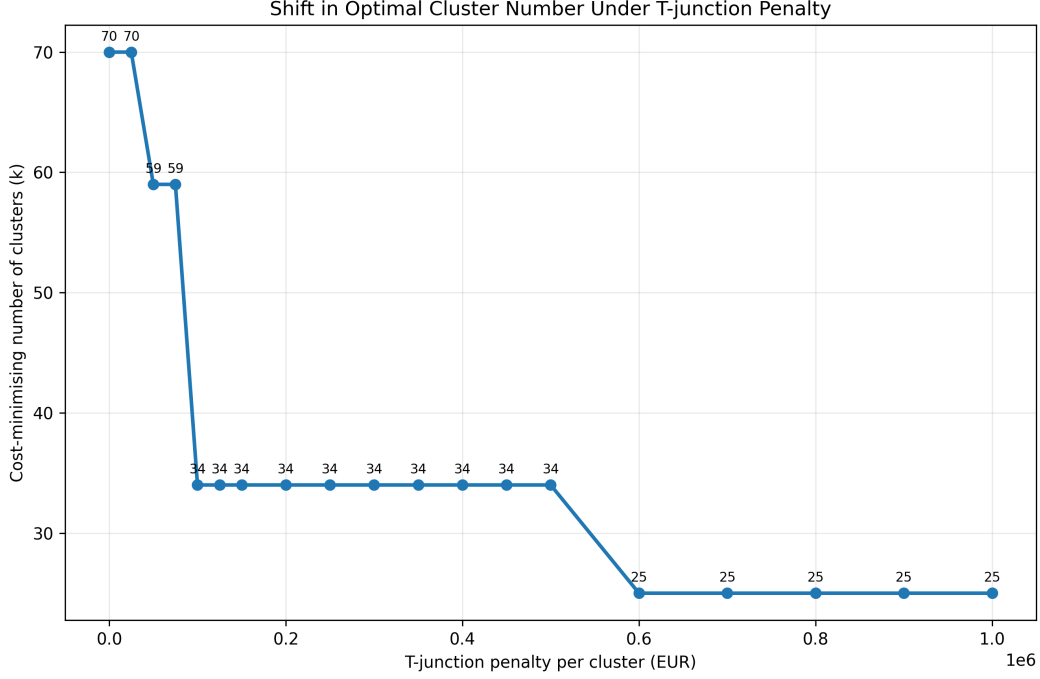


Figure 13: Shift in the cost-minimising number of clusters under increasing T-junction penalties.

These results show that the preferred clustering configuration depends strongly on whether connection costs are included. Without T-junction costs, the lowest internal ONLT cost is found at $k = 70$. After adding T-junction costs, the optimum shifts to lower values of k . In the tested range with the more realistic costs around €500,000, $k = 34$ is the most stable practical optimum, while the higher T-junction costs further reduce the preferred number of clusters to $k = 25$.

The important result is that the cost-minimising clustering configuration remains $k = 34$ for tested T-junction costs between €150,000 and €500,000. At €600,000 and above, the additional penalty on configurations with many clusters becomes large enough to change the optimum. For the tested values from €600,000 to €1,000,000, the cost-minimising configuration is $k = 25$.

This result changes the interpretation of the optimal clustering level. Without T-junction costs, the model favours $k = 70$. Once realistic connection costs are included, the optimum shifts to $k = 34$. This shows that the preferred clustering configuration is not only determined by internal network length and layout, but also by the number of required backbone connection points.

The network with $k=34$ is visualised in figure 14



Figure 14: $k=34$ network configuration

7.3 Interim Finding

The first results show a clear distinction between the technical network optimum and the practical cost optimum. When T-junction costs are excluded, the lowest internal ONLT cost occurs at $k = 70$. This means that, from a purely internal network perspective, a relatively detailed clustering configuration performs best.

When T-junction costs are included, the preferred clustering configuration shifts downward. For tested T-junction costs between €150,000 and €500,000, the cost-minimising solution is consistently found at $k = 34$. For tested T-junction costs between €600,000 and €1,000,000, the optimum shifts further downward to $k = 25$. This shows that higher connection costs increasingly favour fewer and larger clusters.

This finding is central to the research question. Spatial clustering affects infrastructure cost in two opposite ways. More clusters reduce internal transport distances, but also increase the number of required connection points. Fewer clusters reduce the number of T-junctions, but increase internal network distances. The preferred configuration is therefore found in the balance between these two effects.

7.4 Anchor-site Composition of Selected Clustering Configurations

The previous sections identified two relevant clustering configurations. The first is $k = 70$, which gives the lowest internal ONLT cost when T-junction costs are excluded. The second is $k = 34$, which becomes the most relevant practical configuration once T-junction costs around €500,000 are included. However, these cost results alone do not show whether the resulting clusters are institutionally credible.

For this reason, both selected clustering configurations are also evaluated using the actor-informed distinction between anchor and marginal demand introduced in Section 3.4. Anchor sites are firms from sectors with relatively high hydrogen relevance and stronger coordination potential. Marginal sites are firms from sectors with lower hydrogen relevance, stronger electrification alternatives or more uncertain participation.

This analysis is applied to both $k = 34$ and $k = 70$. The $k = 34$ configuration represents a more aggregated clustering outcome with fewer collective connection points, while $k = 70$ represents a more spatially detailed outcome with lower internal ONLT cost. Comparing both configurations helps to assess how the level of spatial aggregation affects the presence of anchor-sector firms within clusters.

Figure 15 shows the anchor-site composition for the $k = 34$ clustering configuration. The figure shows for each cluster the number of anchor sites relative to the total number of sites in that cluster. Clusters with at least one anchor site are more likely to represent credible collective hydrogen connection opportunities, because they contain demand from sectors that are more likely to require hydrogen and to participate in collective infrastructure planning.

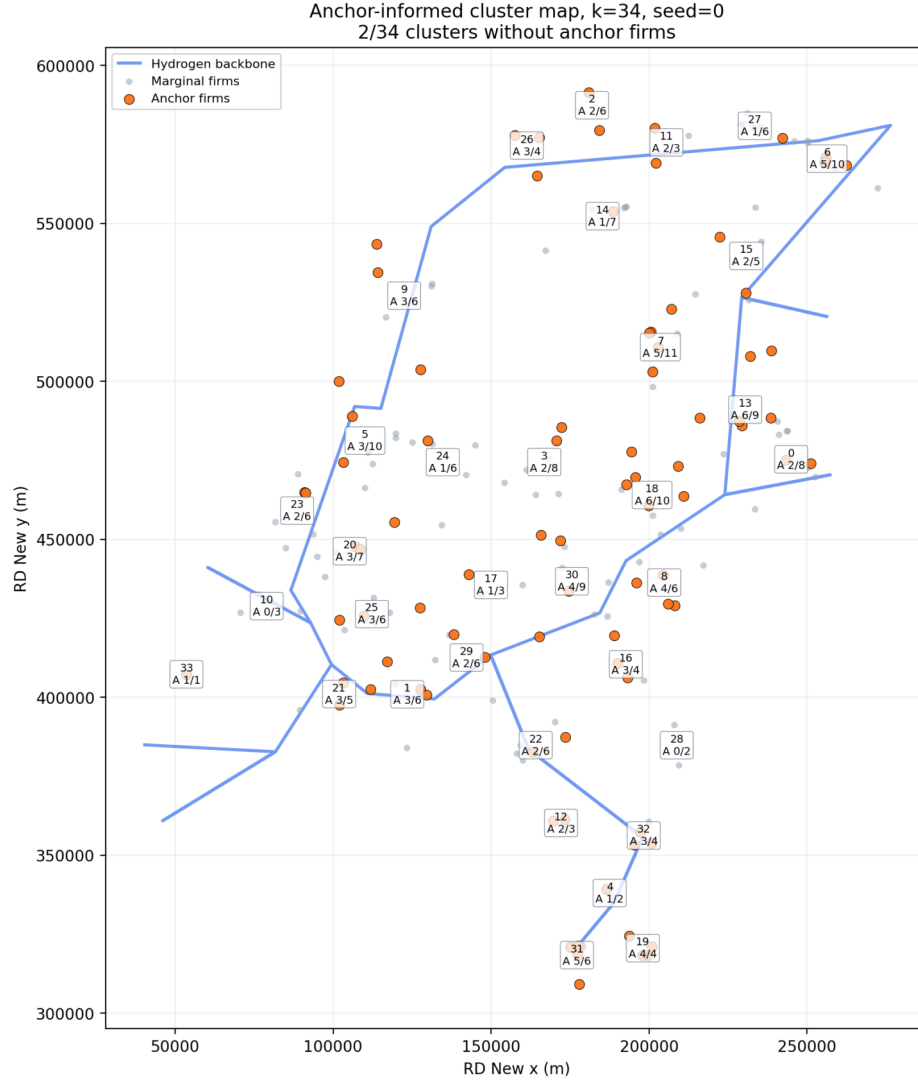


Figure 15: Anchor and marginal sites within the $k = 34$ clustering configuration.

For $k = 34$, in the reference clustering shown in Figure 15 using K-means seed 0, 2 out of 34 clusters contain no anchor-sector company. This means that 32 out of 34 clusters include at least one firm from a sector with relatively high hydrogen relevance and stronger coordination potential.

Figure 16 shows the same assessment for the $k = 70$ clustering configuration.

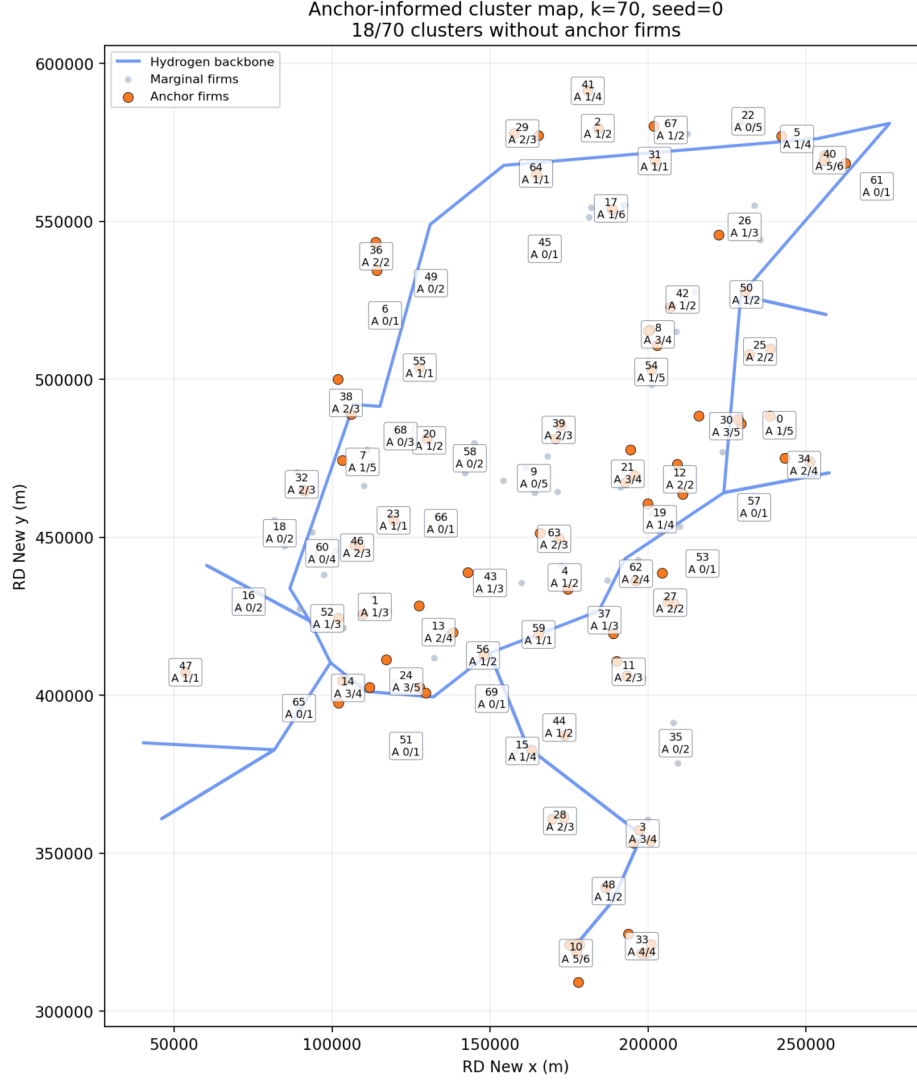


Figure 16: Anchor count per K-means cluster for the $k = 70$ clustering configuration.

For $k = 70$, in the reference clustering shown in Figure 16 using K-means seed 0, 18 out of 70 clusters contain no anchor-sector company. This does not make the configuration unsuitable, but these clusters are less likely to drive collective connection development on their own.

The comparison shows that $k = 34$ is stronger from an actor-informed perspective. In the reference seed used in Figures 18 and 19, namely K-means seed 0, the $k=34$ configuration has 2 clusters without an anchor-sector company, while the $k=70$ configuration has 18. This means that 32 out of 34 clusters are anchor-supported for $k=34$, compared with 52 out of 70 clusters for $k = 70$. The more detailed $k=70$ configuration therefore creates more clusters that may be weaker from an actor-informed implementation perspective.

7.5 Robustness of Selected Clustering Configurations

The anchor-site analysis compares the actor credibility of the $k = 34$ and $k = 70$ configurations. However, this still treats each configuration as a fixed clustering outcome. Since K-means can produce different cluster memberships depending on the initial centroid positions, both configurations are tested across multiple random seeds.

For both $k = 34$ and $k = 70$, the K-means algorithm was repeated for 50 random seeds. For each value of k , the clustering run with the lowest within-cluster sum of squares was used as the reference solution. The stability of each reference cluster was then assessed using the mean Jaccard similarity between that cluster and the most similar cluster found in the other seed runs. A Jaccard similarity close to 1 indicates that the same group of companies is repeatedly assigned to the same cluster. Lower values indicate that the cluster membership is more sensitive to the initial centroid placement.

Figure 17 shows the per-cluster stability for $k = 34$. Most clusters have relatively high Jaccard similarity scores. This means that the main regional demand groupings are reproduced across different K-means initialisations. Some clusters show lower stability, especially in areas where industrial sites are located close to several neighbouring clusters.

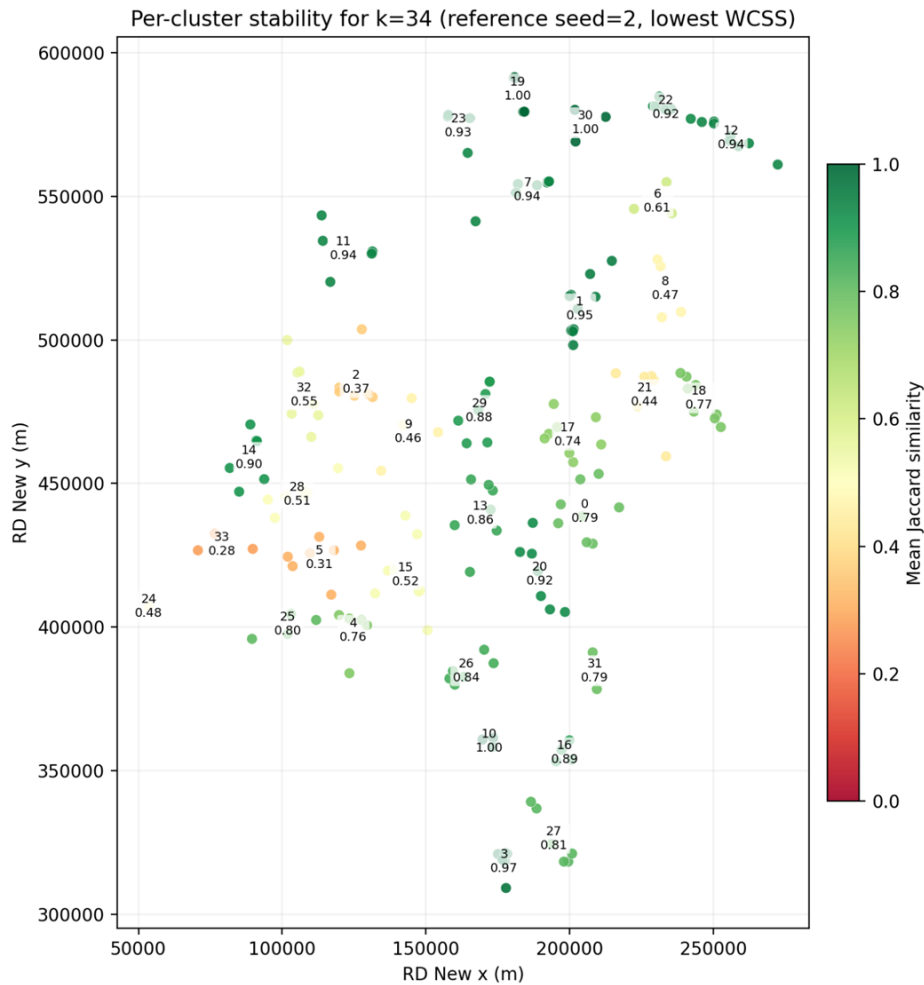


Figure 17: Per-cluster stability for $k = 34$, measured using mean Jaccard similarity across K-means seed runs.

Figure 18 shows the same stability analysis for $k = 70$. The results show that many clusters are highly stable, especially small and spatially isolated clusters. However, compared to $k = 34$, the configuration also contains more clusters with intermediate stability scores. This is expected because a higher number of clusters creates smaller groups and more cluster boundaries. As a result, small changes in centroid initialisation can more easily move one or two companies between neighbouring clusters.

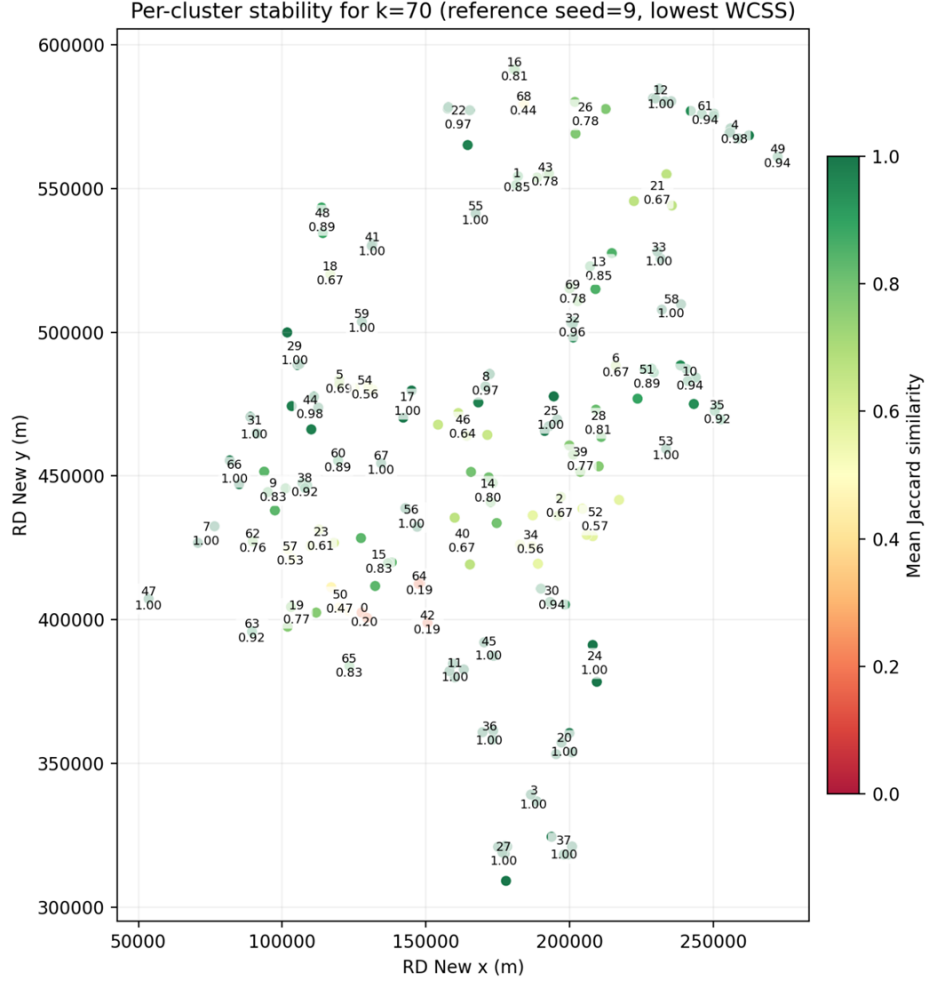


Figure 18: Per-cluster stability for $k = 70$, measured using mean Jaccard similarity across K-means seed runs.

The stability analysis shows that both $k = 34$ and $k = 70$ contain clusters that are stable across different K-means initialisations, while other clusters remain more sensitive to the chosen seed.

The stability analysis therefore opens up a further interpretation of the results. Instead of only selecting one optimal value of k , the model can also be used to identify promising alternative cluster configurations. Clusters with high stability can be treated as relatively robust demand groups, while lower-stability clusters may require further investigation.

7.6 Threshold Sensitivity of Selected Clustering Configurations

The previous sections show that the preferred clustering configuration depends on more than internal network cost alone. Without T-junction costs, the lowest ONLT cost is found for the more spatially detailed $k = 70$ configuration. Adding T-junction costs shifts the cost balance towards fewer clusters, making $k = 34$ more attractive under moderate connection-cost assumptions. The actor-informed analysis then examines whether these configurations are supported by credible demand, based on the presence of anchor-demand firms. This section adds a final feasibility layer by testing whether the selected clusters also contain enough hydrogen demand to recover a fixed connection cost.

The threshold analysis is not used to select a new optimal value of k . Instead, it is used to interpret

the selected configurations more carefully. Four configurations are compared: the lowest-cost $k = 34$ seed, the actor-informed $k = 34$ seed, the lowest-cost $k = 70$ seed and the most actor-favourable $k = 70$ seed.

A cluster is treated as a credible feasible connection candidate only when two conditions are met. First, the cluster must contain enough hydrogen demand to recover the fixed connection cost within the assumed payback period. Second, the cluster must contain at least one anchor-sector firm. This makes the anchor filter explicit in the feasibility calculation:

$$Feasible_c = \begin{cases} 1, & \text{if } H_c \geq H_{\min} \text{ and } A_c \geq 1, \\ 0, & \text{otherwise.} \end{cases} \quad (9)$$

where H_c is the annual hydrogen demand of cluster c , $H_{\min}$ is the minimum demand threshold, and A_c is the number of anchor-sector firms in the cluster. Clusters with $A_c = 0$ are excluded from the main feasible-cluster count, even if their estimated demand exceeds the threshold. This distinction is important because a cluster may be demand-feasible but still weak from an actor perspective if it contains no anchor firm. The main policy indicator in this section is therefore the number of credible feasible clusters, meaning clusters that pass both the demand threshold and the anchor condition.

The threshold analysis requires an assumption about how much money per kilogram of delivered hydrogen could be available to recover a fixed connection or T-junction cost. This value is not directly observable, because it depends on future tariffs, cost allocation rules, commercial agreements and possible policy support. Therefore, this study does not use a single assumed value, but applies a sensitivity range grounded in three reference points.

The assumed cost-recovery contributions are based on three reference points. First, Hynetwork’s 2026 hydrogen transport tariff is used as a conservative infrastructure-cost proxy. The combined entry and exit tariff is 48.71 euro per kW per year. Under full utilisation, this is equivalent to approximately 0.19 euro/kg H_2 . This supports the low contribution case of 0.20 euro/kg H_2 (Hynetwork Services, 2026b).

Second, the IEA Global Hydrogen Review 2025 assumptions annex is used as a commercial-margin reference. The IEA assumes a margin of 6% in its hydrogen price build-up. For delivered hydrogen prices of roughly 4 to 7 USD/kg H_2 , this corresponds to approximately 0.24 to 0.42 USD/kg H_2 . This supports the use of 0.50 euro/kg H_2 as a moderate cost-recovery case (International Energy Agency, 2025a).

Third, the Dutch OWE subsidy scheme shows that public support for renewable hydrogen can be much larger than normal commercial margins. Support can reach up to 9 euro/kg H_2 . This does not represent a connection tariff, but it shows that policy support can operate at a multi-euro-per-kilogram scale. The higher sensitivity cases of 1.00 and 2.00 euro/kg H_2 are therefore interpreted as high or policy-supported cost-recovery cases (Netherlands Enterprise Agency (RVO), 2024).

Based on these reference points, four contribution levels are used in the threshold analysis, as shown in Table 10.

Table 10: Contribution levels used in the threshold sensitivity analysis.

Contribution	Interpretation
0.20 euro/kg H ₂	Conservative infrastructure-tariff proxy, based on the converted Hynetwork transport tariff.
0.50 euro/kg H ₂	Moderate cost-recovery case, broadly aligned with a commercial margin reference.
1.00 euro/kg H ₂	Optimistic high cost-recovery case, above ordinary margin-based recovery.
2.00 euro/kg H ₂	Strong policy-supported case, representing a situation with substantial public support.

Hydrogen demand values reported in kilograms were converted to terajoules using the lower heating value of hydrogen:

$$E_{\text{H}_2} = 0.00012 \cdot m_{\text{H}_2}, \quad (10)$$

where E_{H_2} is the hydrogen demand in TJ and m_{H_2} is the hydrogen mass in kg. This conversion is based on 1 kg of hydrogen corresponding to approximately 0.00012 TJ.

The converted annual demand was then compared with the minimum demand threshold for each contribution and subsidy configuration. A cluster was counted as a credible feasible connection candidate only when it exceeded this threshold and contained at least one anchor-sector firm. This ensures that the feasibility count reflects both sufficient hydrogen demand and actor-informed credibility.

Figure 19 shows the number of credible feasible clusters for each selected configuration under the tested contribution and subsidy combinations.

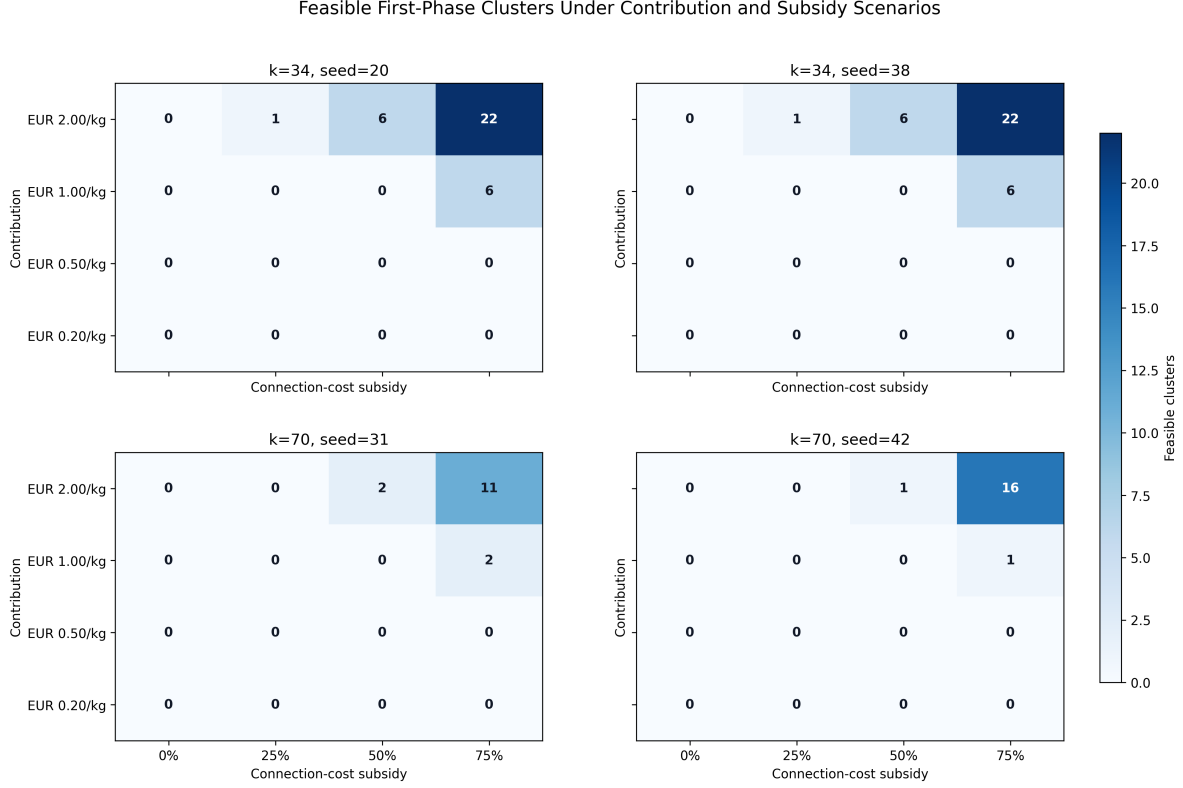


Figure 19: Number of credible feasible clusters under different contribution and subsidy assumptions.

The heatmap shows that credible feasible clusters only start to appear under relatively favourable cost-recovery and subsidy assumptions. At 1.00 euro/kg H_2 with a 75% subsidy, six clusters become feasible in both $k = 34$ configurations. At 2.00 euro/kg H_2 with a 50% subsidy, the same threshold is reached, again resulting in six credible feasible clusters for $k = 34$. When the contribution is 2.00 euro/kg H_2 and the subsidy is 75%, the threshold falls even further. In this case, 22 clusters become credible feasible candidates in both $k = 34$ configurations.

The comparison with $k = 70$ strengthens the earlier interpretation. Although $k = 70$ performs well in terms of internal ONLT cost, it splits demand into more and smaller clusters. These smaller clusters are less able to recover a fixed connection cost. At 1.00 euro/kg H_2 with a 75% subsidy, only one cluster is credible feasible in the lowest-cost $k = 70$ seed and two clusters in the actor-favourable $k = 70$ seed. Even in the strong policy-supported case of 2.00 euro/kg H_2 with a 75% subsidy, the number of credible feasible clusters reaches only 16 for the lowest-cost $k = 70$ seed and 11 for the actor-favourable $k = 70$ seed.

A key conclusion from this analysis is that the preferred clustering configuration depends on the balance between spatial efficiency and demand credibility. A more detailed clustering configuration can reduce internal network costs, but it also creates smaller clusters that may not contain enough demand to recover fixed connection costs. A more aggregated configuration reduces the number of connection points and creates larger demand groups, but may increase internal network distances.

8 Discussion

This chapter discusses the main findings of this thesis and reflects on what they mean for hydrogen infrastructure planning for Cluster 6. The results show that spatial clustering has a strong influence on how dispersed industrial hydrogen demand is represented and assessed. Clustering is therefore not only a data preparation step, but an important modelling choice. The way in which demand is grouped affects internal network length, the number of collective backbone connection points, the cost outcome and the credibility of the resulting clusters.

The results should not be interpreted as an exact infrastructure plan for Cluster 6. Instead, they show how different clustering choices create different trade-offs. A technically attractive clustering configuration may not automatically be the most practical one if it requires many backbone connections, lacks anchor-demand firms or does not contain enough demand to justify a collective connection.

8.1 Interpretation of the Main Results

The results show a clear trade-off between many small clusters and fewer large clusters. When only the internal ONLT network costs are considered, a more detailed clustering configuration performs better. In this case, the lowest internal network cost is found at 70 clusters. This result is logical, because smaller clusters reduce the distance between companies within each cluster. Shorter internal distances lead to shorter internal pipeline networks and therefore lower internal network costs.

However, this internal-cost optimum does not by itself determine the most useful clustering configuration. A configuration with many clusters also creates more collective connection points to the hydrogen backbone. Once T-junction costs are included, the cost advantage of highly detailed clustering decreases. Under the tested connection-cost assumptions, the preferred configuration shifts from 70 clusters towards 34 clusters. When T-junction costs become even higher, the preferred configuration shifts further towards 25 clusters.

The actor-informed and threshold analyses add another layer to this interpretation. A cluster is more credible when it contains anchor-sector firms, because these firms have a stronger expected role for hydrogen and may be more likely to support collective infrastructure development. At the same time, anchor presence alone is not sufficient. A cluster also needs enough demand scale to make a collective connection economically plausible.

This means that the preferred clustering configuration cannot be selected on the basis of network cost alone. Internal network cost, the number of backbone connection points, anchor-sector presence and demand scale all play an important role. For Cluster 6, the configuration with 34 clusters provides a relevant middle ground. It reduces the number of connection points compared to 70 clusters, while still keeping internal transport distances relatively limited. It also performs better from an actor-informed perspective because most clusters contain anchor-demand firms.

The result does not mean that 34 clusters is the exact final solution for future hydrogen infrastructure planning. Rather, it shows that a medium level of aggregation can provide a more useful planning basis than either highly detailed site-level clustering or very coarse regional aggregation. The final choice requires a trade-off between spatial efficiency, connection costs and the credibility of the underlying hydrogen demand.

8.2 The Role of T-junction Costs

The analysis also shows that T-junction costs are an important factor in hydrogen infrastructure planning. Without these costs, the model mainly favours configurations with short internal distances. Once connection costs are included, the number of clusters becomes much more important. This changes the interpretation of the results.

This is relevant because the connection between Cluster 6 clusters and the hydrogen backbone will probably

not be free or standardised. Connection costs may differ by location, required capacity, technical complexity and the distance to the backbone. The exact cost of a T-junction is uncertain, which means that the results should not be interpreted as exact investment estimates. Instead, the T-junction analysis should be seen as a sensitivity analysis. It shows how strongly the preferred clustering configuration depends on assumptions about connection costs.

The results indicate that infrastructure planning for dispersed demand should not only focus on the cost of internal pipelines. Even if a network layout is efficient within a cluster, it may still become less attractive when every cluster requires a separate connection to the backbone. This is especially important for Cluster 6, because the companies are spread across the country and do not naturally form a few large industrial demand centres.

8.3 Actor-informed Interpretation of the Results

A second important finding is that a low-cost cluster is not automatically a credible cluster. The actor analysis showed that Cluster 6 firms differ strongly in their expected role for hydrogen. Some sectors, such as specialty chemicals, metallurgy, glass and ceramics, have a stronger technical reason to use hydrogen. These sectors were labelled as anchor demand. Other sectors, such as food processing, paper and cardboard, waste processing, data centres and onshore oil and gas, have more uncertain hydrogen demand and were labelled as marginal demand.

This distinction matters when interpreting the clustering results. A cluster that contains at least one anchor-sector company is more likely to represent a credible collective hydrogen connection point. Such a company may provide stronger demand, more technical need and a clearer reason to participate in hydrogen infrastructure development. A cluster that only contains marginal-demand firms may still have estimated hydrogen demand, but this demand is more uncertain. These firms may choose electrification or other decarbonisation routes instead.

From this perspective, the 34-cluster configuration is stronger than the 70-cluster configuration. The 70-cluster configuration creates more detailed spatial clusters, but also includes more clusters without anchor-sector companies. These clusters may be harder to organise and may have less credible demand for a collective hydrogen connection. The 34-cluster configuration contains anchor-demand companies in almost all clusters. This makes it more credible from an implementation perspective.

The actor-informed results therefore support the idea that clustering should not only be assessed with technical and cost indicators. It should also be assessed by looking at the type of companies within each cluster. For implementation, a cluster needs more than a low network cost. It also needs credible demand, coordination potential and firms that have a realistic reason to use hydrogen.

8.4 Practical Implications

The results have practical implications for Hynetwork, Stichting Cluster 6, regional authorities and policymakers. First, they suggest that Cluster 6 should not be approached only as a collection of individual sites. Connecting companies one by one would probably lead to many separate connection points, which can become expensive and difficult to coordinate.

At the same time, the results also show that very strong aggregation is not always desirable. If too many companies are grouped into a small number of large clusters, internal transport distances increase. This can make the internal network more expensive and less realistic. A medium level of regional clustering appears to be more useful, because it combines demand bundling with manageable internal distances.

For Stichting Cluster 6 and regional authorities, the results can help identify where collective connection planning may be most promising. Clusters with sufficient demand, reasonable internal distances and at least one anchor-sector company are likely to be stronger starting points for further investigation. These clusters

could be prioritised in discussions with Hynetwork and policymakers.

For Hynetwork, the results show that dispersed industrial demand can become more actionable when it is presented as regional clustered demand instead of separate company-level demand points. This can help reduce uncertainty about where collective connection points may be located. However, the results also show that connection planning depends strongly on assumptions about T-junction costs and minimum demand thresholds. More transparency about these conditions would make it easier for Cluster 6 companies to assess whether collective hydrogen connections are realistic.

8.5 Validation of the Results

Given additional time and access to more detailed data, the results could be validated through a combination of empirical, expert-based and methodological validation.

First, the estimated hydrogen demand could be compared with company-specific data on natural gas consumption, production capacity and decarbonisation plans. Interviews or workshops with Cluster 6 companies and sector associations could be used to determine whether the estimated demand and the anchor and marginal classifications reflect their actual expectations regarding hydrogen use.

Second, the spatial clustering results could be reviewed with Stichting Cluster 6, Hynetwork and regional authorities. These actors could assess whether the identified clusters represent realistic regional collaborations and whether the proposed connection areas are compatible with existing infrastructure plans, industrial relationships and spatial constraints.

Third, the network and cost results could be compared with detailed engineering estimates for a selection of representative clusters. This would allow the assumed pipeline costs, T-junction costs and backbone connection distances to be tested against more location-specific information.

Finally, the robustness of the results could be assessed by applying alternative clustering methods and demand assumptions. Comparing k-means with hierarchical, demand-weighted or network-constrained clustering would show whether the main trade-off between internal network costs and backbone connection costs remains consistent across different modelling approaches.

Together, these validation steps would help determine whether the identified configurations are not only internally consistent within the model, but also technically, economically and institutionally plausible in practice.

8.6 Limitations

Several limitations should be considered when interpreting the results. First, hydrogen demand is estimated using available sector-level information on natural gas use and assumed hydrogen substitution rates. This provides a useful approximation of potential hydrogen demand, but it does not capture company-specific production processes, investment plans, operating conditions or decarbonisation strategies. Actual hydrogen demand may therefore differ from the estimated demand used in this study.

A related limitation concerns the resolution of company-level demand. Due to limited availability of site-specific energy consumption data, companies within the same sector are represented using the same estimated hydrogen demand. This means that differences in plant size, production capacity, operating hours and process heat requirements are not captured. For example, a large paper mill and a smaller paper producer may be represented as identical demand nodes if they belong to the same sector.

This simplification affects more than the demand estimate itself. It also propagates into the clustering results, the estimated network capacities and the threshold sensitivity analysis. Clusters with several smaller firms may therefore appear more credible than they would be in practice, while clusters containing one or more

large industrial sites may be underestimated. The results should therefore be interpreted as a sector-level approximation of potential hydrogen demand rather than as a company-specific demand forecast.

Second, the substitution rates are assumptions. They are based on sector-level reasoning about hydrogen relevance and alternative decarbonisation options. However, future hydrogen use depends on many uncertain factors, including hydrogen prices, electricity prices, policy support, technology development and company-specific investment decisions. The substitution rates should therefore not be interpreted as predictions, but as indicative assumptions used to compare clustering configurations under a consistent modelling approach.

Third, the clustering method simplifies reality. K-means clustering groups companies based on geographic proximity, but it does not account for all practical routing constraints. Real infrastructure development would also depend on land use, existing pipeline corridors, permitting, safety zones, environmental constraints and local spatial planning. These factors are not included in the clustering model.

In addition, other clustering methods could lead to different spatial groupings. Hierarchical clustering could show how individual sites merge into larger regional groups at different aggregation levels. DBSCAN could identify dense industrial regions without predefining the number of clusters, although it may classify remote sites as outliers. Demand-weighted or capacitated clustering could account more directly for hydrogen demand volumes and minimum demand thresholds. Network-constrained clustering could also include existing infrastructure corridors, routing constraints and distance to the hydrogen backbone. Future research could compare these methods to test how sensitive the results are to the chosen clustering algorithm.

Fourth, the ONLT provides tree-like network layouts without redundancy. This is useful for comparing cost-efficient network structures, but it does not include backup routes or security-of-supply requirements. In real infrastructure planning, redundancy and reliability may be important, especially for industrial users that require continuous hydrogen supply. The model therefore provides insight into cost-minimising network structures, but not into fully resilient infrastructure designs.

Fifth, T-junction costs are uncertain. The analysis uses a range of assumed connection costs to test the sensitivity of the results, but these values are not exact project estimates. Actual costs may differ strongly between locations depending on technical complexity, capacity requirements, pressure levels, available space, permitting conditions and the distance to suitable backbone connection points.

Sixth, the euro conversion relies on a fixed 20-inch pipeline diameter and a single ONLT-to-euro conversion factor. This makes the ONLT results easier to interpret in monetary terms, but it also simplifies the actual design problem. In practice, pipeline diameters would differ between clusters depending on hydrogen demand, required capacity, pressure levels, routing conditions and future expansion needs. The euro-equivalent costs should therefore be interpreted as indicative order-of-magnitude estimates rather than detailed engineering cost estimates.

Finally, the actor-informed classification simplifies real company behaviour. Labelling firms as anchor or marginal demand helps to interpret the clustering results, but firms within the same sector can still differ substantially. Some marginal-sector firms may become strong hydrogen users, while some anchor-sector firms may delay or avoid hydrogen investment. The classification should therefore be seen as an interpretation tool rather than as a fixed prediction of company behaviour.

9 Conclusions

This thesis investigated how spatial clustering of dispersed industrial hydrogen demand affects transport infrastructure design and cost. The focus was on Cluster 6 companies in the Netherlands and their possible connection to the national hydrogen backbone. These companies are spread across the country and are not automatically included in the current backbone planning. This creates a challenge, because individual sites may be too small or too remote to justify a separate connection, while their combined demand may still be relevant for collective hydrogen infrastructure.

To study this problem, company-level demand was estimated, Cluster 6 firms were spatially clustered using K-means, and the resulting clusters were analysed with the Optimal Network Layout Tool. Connection costs to the hydrogen backbone were added through T-junction cost configurations. Finally, the results were interpreted with actor-informed demand labels, distinguishing between anchor-demand and marginal-demand firms.

9.1 Answering the Research Questions

9.1.1 How can dispersed industrial demand within Cluster 6 be spatially clustered into representative regional groups?

Dispersed industrial demand within Cluster 6 can be clustered by using the geographic location of individual companies as input for K-means clustering. By varying the number of clusters, different levels of spatial aggregation can be created. Lower values of k result in fewer and larger regional clusters, while higher values of k result in more detailed clusters with shorter internal distances.

The clustering approach makes it possible to transform a dispersed set of company-level demand points into regional demand groups. This is useful because the national hydrogen backbone is unlikely to connect every individual Cluster 6 site separately. By grouping nearby companies, demand can be bundled into more visible and potentially more credible regional connection points.

However, the results also show that clustering should not be interpreted only as a mathematical grouping exercise. The actor-informed classification is needed to understand whether the clusters contain credible hydrogen demand. Clusters that include anchor-sector companies are more relevant for collective connection planning than clusters that mainly depend on marginal-demand firms.

9.1.2 How do different clustering configurations affect the spatial structure, connection requirements and feasibility of Cluster 6 hydrogen networks?

Different clustering configurations strongly affect the spatial structure of the resulting Cluster 6 hydrogen networks. Configurations with a higher number of clusters create more localised networks with shorter internal transport distances, but they also require more separate connection points to the national hydrogen backbone. In contrast, configurations with fewer clusters reduce the number of required backbone connections, but increase the internal distances over which demand must be bundled.

The results show that this creates a clear trade-off. When only internal network cost is considered, a more disaggregated configuration is preferred, because shorter internal pipeline distances reduce the ONLT network cost. However, when T-junction costs are included, configurations with many clusters become less attractive because each additional cluster may require an additional connection to the backbone. This shifts the preferred configuration towards fewer, larger clusters.

The analysis therefore shows that clustering affects not only the layout of the regional network, but also the number and credibility of backbone connection points. The results do not provide detailed hydraulic capacity sizing. Instead, connection requirements are interpreted through the number of clusters, the spatial distribution of connection points and the demand credibility of each cluster. From this perspective, the

medium-level clustering configuration provides the most useful balance between internal network length, connection cost and actor-informed feasibility.

9.1.3 What is the relationship between cluster configuration and total infrastructure cost, and at what point do additional clusters yield diminishing returns?

The relationship between cluster configuration and total infrastructure cost depends on whether only internal network costs are considered or whether backbone connection costs are also included. When only internal ONLT network costs are included, a more detailed clustering configuration performs best. In this case, the lowest internal network cost is found at 70 clusters. This is because smaller clusters reduce the internal distances between companies.

However, this changes when T-junction costs are added. Every additional cluster may require an additional connection to the hydrogen backbone. This means that configurations with many clusters become more expensive once connection costs are included. Under the tested T-junction cost assumptions, the preferred configuration shifts from 70 clusters towards 34 clusters. When T-junction costs become higher, the preferred configuration shifts further towards 25 clusters.

This shows that additional clusters only reduce costs up to a certain point. After that point, the reduction in internal network cost no longer compensates for the extra cost of additional backbone connections. For Cluster 6, the results indicate that a medium level of aggregation gives the best balance between internal network cost and connection cost.

9.1.4 How does spatial clustering of dispersed industrial hydrogen demand affect transport infrastructure design and cost?

The central research question can be answered as follows. Spatial clustering strongly affects both the design and cost of hydrogen infrastructure for dispersed industrial demand. It changes the internal network length, the number of required backbone connections and the total infrastructure cost.

A highly detailed clustering configuration is attractive from the perspective of internal network design, because it keeps companies close together within each cluster. However, it also creates many potential connection points to the hydrogen backbone. A strongly aggregated configuration reduces the number of backbone connections, but creates longer internal networks. The most useful solution is therefore found between these two extremes.

For Cluster 6, the results show that the configuration with 34 clusters is a relevant practical outcome. It reduces the number of connections compared to the 70-cluster configuration, while still keeping internal distances acceptable. It also performs better from an actor-informed perspective, because almost all clusters contain at least one anchor-demand firm. This makes the configuration more credible for collective hydrogen connection planning.

9.2 Final Conclusion

The main conclusion of this thesis is that the connection of Cluster 6 to the Dutch hydrogen backbone should not be approached as a purely technical routing problem. It is also a question of how dispersed demand is grouped and which companies are likely to support collective hydrogen infrastructure.

Spatial clustering has a clear influence on infrastructure design and cost. More detailed clustering reduces internal network distances, but increases the number of required backbone connections. Stronger aggregation reduces the number of connections, but increases internal transport distances. When T-junction costs and actor feasibility are included, a medium level of aggregation becomes more relevant than the purely internal-cost optimum.

For the case of Cluster 6, this means that regional clustering can help create credible collective demand

for hydrogen infrastructure. The most promising clusters are those with reasonable internal distances, sufficient bundled demand and anchor-sector companies. Spatial clustering is therefore not only a modelling tool, but also a way to support the integration of dispersed industrial demand into the future Dutch hydrogen system.

10 Recommendations

Based on the conclusions of this thesis, several recommendations can be made for future hydrogen infrastructure planning for Cluster 6 and for further research. The recommendations focus on improving the practical use of the clustering approach, reducing uncertainty in the model inputs and making the results more suitable for real infrastructure planning.

10.1 Recommendations for Practice

First, Cluster 6 demand should be assessed through regional clusters rather than through individual company connections. The results show that connecting companies one by one would lead to many separate backbone connections, which can become expensive and difficult to coordinate. At the same time, grouping all companies into only a few large regions increases internal transport distances. Therefore, a medium level of spatial aggregation should be used as a starting point for further planning.

Second, clusters should not only be selected based on cost. The actor-informed analysis shows that the type of companies within a cluster is also important. Clusters with at least one anchor-sector company are more credible, because these firms have a stronger expected role for hydrogen and are more likely to support collective infrastructure development. Stichting Cluster 6, Hynetwork and regional authorities should therefore prioritise clusters that combine sufficient demand, reasonable internal distances and the presence of anchor-sector firms.

Third, the configuration with 34 clusters should be used as a practical starting point for further investigation. This does not mean that 34 clusters is the final or exact infrastructure design. Instead, it provides a useful balance between internal network cost, number of backbone connection points and actor-informed feasibility. The 34-cluster configuration can therefore be used to identify promising regions for more detailed technical, economic and institutional analysis.

Fourth, more clarity is needed about the costs and requirements of connecting to the hydrogen backbone. The results show that T-junction costs have a large influence on the preferred clustering configuration. If connection costs are low, more detailed clustering becomes attractive. If connection costs are high, fewer and larger clusters become more favourable. Hynetwork and policymakers should therefore provide more transparent information about connection procedures, cost ranges and minimum demand requirements. This would help Cluster 6 companies assess whether collective connections are realistic.

Fifth, regional coordination should be strengthened. Even when a cluster appears technically and economically promising, a collective connection will only be possible if companies, regional authorities, Stichting Cluster 6 and Hynetwork coordinate their plans. Regional authorities can play an important role in bringing companies together, supporting spatial planning and linking hydrogen infrastructure to broader regional energy strategies.

10.2 Recommendations for Future Research

First, future research should use more detailed company-specific energy data. In this thesis, hydrogen demand is estimated using sector-level natural gas consumption and substitution rates. This provides a useful first approximation, but it does not capture the specific processes, energy use and investment plans of individual companies. More detailed data would improve the accuracy of the demand estimates and make the clustering results more reliable.

Second, future research should compare different clustering methods. This thesis uses K-means clustering because it provides a clear and consistent way to compare different values of k . However, K-means tends to create compact clusters around centroids and requires the number of clusters to be chosen in advance. Other methods could produce different spatial groupings. Hierarchical clustering could show how individual sites merge into larger regional groups at different aggregation levels. DBSCAN could identify dense indus-

trial regions without predefining the number of clusters, although it may classify remote sites as outliers. Demand-weighted or capacitated clustering could account more directly for hydrogen demand volumes and minimum demand thresholds. Network-constrained clustering could include existing infrastructure corridors, routing constraints and distance to the hydrogen backbone. Comparing these methods would show how sensitive the results are to the chosen clustering algorithm.

Third, future studies should include more detailed spatial routing. The current model uses spatial clustering and network optimisation, but does not fully include real-world routing constraints such as existing pipeline corridors, land use, waterways, railways, roads, safety zones, permitting issues and protected areas. Including these factors would make the resulting network layouts more realistic for implementation.

Fourth, future research should improve the pipeline cost estimation. In this thesis, the euro conversion is based on a fixed 20-inch pipeline diameter. This makes the ONLT results easier to interpret in monetary terms, but it simplifies the actual design problem. In practice, pipeline diameter would depend on demand volume, required capacity, pressure level, route characteristics and future expansion needs. Future research could use diameter-specific cost functions or hydraulic modelling to estimate pipeline sizes and costs more accurately for each cluster.

Fifth, T-junction costs should be studied in more detail. This thesis treats T-junction costs as a sensitivity parameter because exact connection costs are uncertain. More detailed cost information from Hynetwork or comparable infrastructure projects would make it possible to replace broad assumptions with more specific estimates. This would improve the reliability of the total cost comparison between clustering configurations.

Sixth, the actor-informed classification should be validated through interviews. The distinction between anchor-demand and marginal-demand firms is useful for interpreting the results, but it is based on sector-level reasoning. Interviews with Cluster 6 companies, branch associations, Hynetwork and regional authorities could provide more insight into actual willingness to participate in collective hydrogen connections.

Seventh, future research should compare pipeline connections with other hydrogen supply options. Some dispersed or low-demand clusters may not justify a direct pipeline connection to the backbone. For these clusters, truck delivery, local electrolysis or other hydrogen carriers may be more suitable. Comparing these options would make it possible to identify which clusters are suitable for pipeline infrastructure and which clusters require alternative supply solutions.

Finally, future research should study the development of Cluster 6 hydrogen infrastructure over time. Hydrogen demand, supply and infrastructure availability will not develop all at once. A staged development model could show which clusters should be connected first and how the network could expand when demand becomes more certain. This would make the results more useful for long-term planning and investment decisions.

A Anchor-informed Comparison of K-means Seed Runs

This appendix reports the anchor-informed comparison of the tested K-means seed runs for the two selected clustering configurations. For each seed, the table reports the total ONLT network length, total ONLT cost, the number of clusters without anchor-sector firms and the number of anchor-supported clusters.

Table 11: Anchor-informed seed comparison for $k = 34$.

Seed	Network length (km)	ONLT cost	Clusters without anchors	Anchor-supported clusters
38	1273.93	763,075.13	1	33
25	1295.82	770,368.93	1	33
20	1319.46	773,598.71	0	34
41	1294.22	774,162.89	2	32
33	1319.98	777,975.83	0	34
42	1297.96	780,882.82	2	32
10	1313.69	781,585.85	0	34
39	1300.76	781,805.11	1	33
12	1307.26	782,552.95	1	33
47	1310.32	782,974.83	1	33
4	1312.55	783,770.29	2	32
48	1315.82	785,034.36	2	32
11	1314.75	785,751.89	2	32
14	1331.77	785,905.96	2	32
46	1341.34	786,010.10	1	33
36	1331.76	786,249.22	0	34
1	1311.41	787,081.89	3	31
40	1319.71	787,203.05	2	32
24	1303.68	788,549.60	1	33
0	1308.68	789,300.40	2	32
13	1321.56	790,139.95	3	31
6	1322.78	790,248.83	1	33
15	1324.25	790,512.13	2	32
3	1341.09	791,121.93	0	34
49	1326.34	792,572.21	2	32
35	1323.05	793,006.82	2	32
21	1340.17	793,639.31	1	33
2	1331.41	793,667.90	2	32
17	1335.96	794,826.99	3	31
30	1321.55	795,183.83	1	33
29	1339.52	796,612.60	2	32
19	1345.61	796,627.55	0	34
32	1356.76	798,463.25	0	34
18	1351.14	798,652.82	1	33
16	1332.70	798,882.28	1	33
31	1344.28	799,322.36	1	33
23	1307.65	799,431.26	1	33
22	1349.21	800,789.49	2	32

Continued on next page

Seed	Network length (km)	ONLT cost	Clusters without anchors	Anchor-supported clusters
34	1338.06	802,231.93	0	34
37	1355.57	802,897.33	2	32
27	1338.83	804,332.36	1	33
44	1357.13	804,699.58	2	32
8	1361.47	805,101.44	1	33
5	1364.58	807,275.47	2	32
9	1342.29	809,019.72	2	32
28	1351.80	809,045.23	3	31
26	1355.96	809,155.15	3	31
7	1365.47	818,104.02	2	32
43	1390.41	818,287.78	2	32
45	1376.50	820,710.59	3	31

Table 12: Anchor-informed seed comparison for $k = 70$.

Seed	Network length (km)	ONLT cost	Clusters without anchors	Anchor-supported clusters
42	1292.59	743,681.22	19	51
6	1297.16	749,008.64	19	51
40	1315.67	749,625.42	19	51
15	1306.59	750,554.48	19	51
18	1324.11	751,510.19	21	49
19	1319.73	753,930.44	20	50
21	1336.55	757,868.89	20	50
3	1349.77	758,590.50	19	51
1	1337.70	759,361.29	20	50
0	1342.73	760,098.90	18	52
49	1336.57	761,270.49	18	52
26	1334.02	762,090.83	22	48
39	1332.25	762,194.55	18	52
14	1349.19	762,596.55	19	51
38	1354.46	763,839.60	19	51
41	1344.60	764,361.51	18	52
30	1360.02	764,435.69	20	50
2	1342.21	764,537.80	19	51
27	1353.33	765,420.39	19	51
34	1345.46	765,595.71	18	52
17	1350.01	765,780.13	19	51
22	1349.47	766,434.69	19	51
46	1353.17	766,578.06	18	52
7	1355.78	766,765.80	20	50
29	1356.40	767,011.10	18	52
13	1350.92	767,818.85	18	52
28	1366.31	767,830.46	20	50

Continued on next page

Seed	Network length (km)	ONLT cost	Clusters without anchors	Anchor-supported clusters
45	1368.23	768,742.87	18	52
16	1357.52	768,833.09	19	51
31	1375.39	769,274.76	17	53
10	1362.29	769,283.24	19	51
8	1365.79	769,628.38	20	50
23	1361.26	770,385.89	18	52
44	1375.10	770,403.60	19	51
20	1371.50	770,682.07	21	49
35	1369.26	771,027.53	21	49
43	1368.51	772,643.62	19	51
47	1380.45	773,421.01	20	50
9	1374.96	773,442.74	18	52
32	1375.57	774,046.57	17	53
12	1386.53	776,458.44	18	52
5	1385.77	777,666.24	20	50
48	1384.57	777,868.28	18	52
24	1383.17	777,900.35	19	51
4	1386.86	777,941.06	18	52
36	1382.86	778,559.68	18	52
11	1396.19	779,368.07	18	52
33	1388.76	779,680.05	19	51
37	1401.51	782,070.18	18	52
25	1427.44	787,194.45	18	52

References

- Beagle, E., Lewis, M., Pecora, B., Rhodes, J., Webber, M., & Hebner, R. (2023). Model to inform the expansion of hydrogen distribution infrastructure. <https://doi.org/10.1016/j.ijhydene.2023.07.017>
- Bique, A. O., & Zondervan, E. (2018). An outlook towards hydrogen supply chain networks in 2050: Design of novel fuel infrastructures in germany. <https://doi.org/10.1016/j.ched.2018.03.037>
- Busch, T., Groß, T., Linßen, J., & Stolten, D. (2023). The role of liquid hydrogen in integrated energy systems: A case study for germany. <https://doi.org/10.1016/j.ijhydene.2023.05.308>
- Cheng, W., & Cheng, Y. F. (2023). A techno-economic study of hydrogen transport by pipelines in canada. <https://doi.org/10.1016/j.jpse.2023.100112>
- Detz, R., Weeda, M., & van Stralen, J. (2020). *Hydrogen in the netherlands: A review of recent dutch scenario studies* (TNO 2020 M10971). TNO. <https://publications.tno.nl/publication/34637043/DB5hwV/TNO-2020-M10971.pdf>
- Enserink, B., Bots, P., van Daalen, E., Hermans, L., Kortmann, R., Koppenjan, J., Kwakkel, J., Ruijgh-van der Ploeg, T., Slinger, J., & Thissen, W. (2022). *Policy analysis of multi-actor systems* (2nd ed.). Eleven / TU Delft Open. <https://doi.org/10.5074/T.2022.004>
- Franco, A., & Giovannini, C. (2023). Routes for hydrogen introduction in industrial hard-to-abate sectors. <https://doi.org/10.3390/en16166098>
- Gasunie. (2024). Hydrogen network netherlands. <https://www.gasunie.nl/en/projects/hydrogen-network-netherlands>
- Gasunie. (n.d.). *Why the hydrogen network is starting with industrial clusters* [Accessed: 2026-04-15]. <https://www.gasunie.nl/en/expertise/hydrogen/hydrogen-and-industry/why-the-hydrogen-network-is-starting-with-industrial-clusters>
- Hammond, J., Rosenberg, M., & Brown, S. (2025). Understanding costs in hydrogen infrastructure networks. <https://doi.org/10.1016/j.ijhydene.2024.12.273>
- Heijnen, P. (2025). Optimal network layout tool (onlt). <https://gitlab.tudelft.nl/pheijnen/optimal-network-layout>
- Heijnen, P. (2026). *Minimum-cost network-layout tool with multi-sinks and multi-sources* [Version 4, 21/01/2026, TU Delft].
- Hynetwork / Gasunie. (2025). Hydrogen infrastructure on track to link up supply and demand [Status update on Rotterdam pipeline (operational H1 2026) and northern Netherlands connections]. <https://www.hynetwork.nl/en/knowledge-base/article/hydrogen-infrastructure-on-track-to-link-up-supply-and-demand>
- Hynetwork Services. (2025). Connection policy for the national hydrogen network [Accessed: 2026-05-03].
- Hynetwork Services. (2026a). Contracts – Connection Study Agreement, Investment Agreement, Valve Study Agreement and Valve Investment Agreement [Accessed: 2026-05-15].
- Hynetwork Services. (2026b). Transport tariff [Indexed 2026 tariff: EUR 24.36/kW/year entry and EUR 24.36/kW/year exit, ex VAT. Accessed 5 July 2026].
- International Energy Agency. (2025a). *Global hydrogen review 2025: Assumptions annex* (tech. rep.) (Licence: CC BY 4.0. Accessed 5 July 2026). IEA. Paris. <https://iea.blob.core.windows.net/assets/15673ab3-a86a-4434-bff4-490bb42d3563/GlobalHydrogenReview2025AssumptionsAnnex.pdf>
- International Energy Agency. (2025b, January). *The netherlands 2024: Energy policy review* (tech. rep.). International Energy Agency. Paris. Retrieved June 23, 2026, from <https://iea.blob.core.windows.net/assets/2b729152-456e-43ed-bd9b-ecff5ed86c13/TheNetherlands2024.pdf>
- Kim, C., Lee, Y., Lee, H., Lee, U., & Kim, K. (2023). Economic and environmental potential of green hydrogen carriers. <https://doi.org/10.1016/j.enconman.2023.117302>
- Kotek, P., Tóth, B. T., & Selei, A. (2023). Designing future-proof hydrogen infrastructure for europe. <https://doi.org/10.1016/j.enpol.2023.113641>
- Kountouris, I., et al. (2024). A unified european hydrogen infrastructure planning. *Nature Communications*, 15. <https://doi.org/10.1038/s41467-024-49867-w>
- Lafeber, M. (2024). *Hydrogen infrastructure planning under uncertainty in an industrial port cluster: A robust over time approach* (Master's thesis). Delft University of Technology.
- Lipiäinen, S., et al. (2023). Hydrogen transport options for industrial users. <https://doi.org/10.1016/j.spc.2023.11.021>

- Liu, H., et al. (2024). Coordinated planning model for hydrogen supply chains. <https://doi.org/10.1016/j.apenergy.2024.124456>
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations.
- Mendler, F., et al. (2025). Spatial modelling framework for hydrogen valleys. *Advances in Applied Energy*. <https://doi.org/10.1016/j.adapen.2025.100207>
- Namazifard, N., Vingerhoets, P., & Delarue, E. (2024). Cost optimization of hydrogen infrastructure. <https://doi.org/10.1016/j.ijhydene.2024.02.324>
- Netherlands Court of Audit (Algemene Rekenkamer). (2025). Hydrogen network under severe pressure – ineffective grant to offset startup losses [Accessed: 2026-05-15].
- Netherlands Enterprise Agency (RVO). (2024). Subsidy scheme for large-scale hydrogen production using an electrolyser (OWE) [Subsidieregeling grootschalige productie volledig hernieuwbare waterstof via elektrolyse. Accessed 5 July 2026].
- Neuwirth, M., Fleiter, T., Manz, P., & Hofmann, R. (2024). The future potential hydrogen demand in energy-intensive industries — a site-specific approach applied to germany. <https://doi.org/10.1016/j.enconman.2021.115052>
- Oostdijk, S. (2025, October). *Optimizing hydrogen supply transport for dutch regional industry by use of multiple transport options* [Master's thesis]. Delft University of Technology [Faculty of Technology, Policy and Management].
- Parolin, F., Colbertaldo, P., & Campanari, S. (2022). Multi-modality hydrogen delivery infrastructure. <https://doi.org/10.1016/j.enconman.2022.115650>
- Parolin, F., Colbertaldo, P., & Campanari, S. (2024). Optimal hydrogen delivery infrastructure design. <https://doi.org/10.1016/j.ijhydene.2024.06.049>
- Statistics Netherlands (CBS). (2026). Energy balance sheet; supply and consumption, sector [Accessed: 2026-05-03]. <https://www.cbs.nl/en-gb/figures/detail/83989ENG>
- Stichting Cluster 6. (2025). Landelijke cluster energie strategie (lces) – cluster 6 [Accessed: 2026-05-03].
- Timalsina, S., et al. (2025). Optimizing hydrogen transport infrastructure under uncertainty. <https://doi.org/10.1016/j.ijhydene.2025.02.336>
- van 't Veer, J., et al. (2025). Hydrogen distribution in the netherlands: Addressing ambiguities in the regulatory framework. *Energy Policy*, 206. <https://www.sciencedirect.com/science/article/pii/S0301421525003015>
- Vergara, D., et al. (2024). Hydrogen revolution in europe: Bibliometric review. *Energies*. <https://doi.org/10.3390/en17153658>
- Yang, M., et al. (2023). Review of hydrogen storage and transport technologies. *Clean Energy*. <https://doi.org/10.1093/ce/zkad021>