



Analysis and Optimization of the Charging Station Network in Germany

Scientific work for the attainment of the degree

Master of Science

at the TUM School of Management
of the Technical University of Munich

Examiner: Univ.-Prof. Dr. Dr. h. c. mult. Horst Wildemann
Research Institute
Corporate management, logistics and production

Supervisor: Sebastian Junker M.Sc.

Course of study: Master TUM-BWL

Submitted by: Giovanni Filomeno
Arcistraße 21
80333 Munich
Matricula 03669006

Submitted on: 29.11.2023.

Table of contents

Table of contents	I
List of Figures	III
List of Tables.....	IV
List of Symbols	5
List of Abbreviations	6
1. Introduction.....	7
1.1 Research Objectives.....	7
1.2 Thesis structure	8
2. Literature review	9
2.1 Electric Vehicle Market Penetration Main Factors	9
2.2 Selection of Charging Station Location.....	10
2.3 Optimization Models for EV Charging Station Distribution.....	11
2.4 Algorithm for Optimal EV Charging Station Configuration	13
2.5 Thesis novelty.....	14
3. Theoretical Background.....	16
3.1 Network Theory	16
3.1.1 Diameter	16
3.1.2 Average Distance	17
3.1.3 Average Clustering	17
3.1.4 Density.....	18
3.2 Genetic Algorithm	19
3.2.1 Algorithm components	19
3.2.2 Variants	21
3.2.3 Strengths and Weakness of Genetic Algorithms.....	21
3.3 Predictions Methods	22
3.3.1 Random Forest.....	22
3.3.2 K-Nearest Neighbors	23
4. Methodology and Results	25
4.1 Pre-Processing	26
4.2 Network construction	27
4.3 Optimization.....	30
4.3.1 Initial Population	30

4.3.2	Fitness Function	31
4.4	Workflow overview	32
4.5	Results and Test Cases.....	33
4.5.1	Network 2015: Spread Across the Territory.....	33
4.5.2	Network 2012: Connection of Separated Subnets	35
4.5.3	Results Summary and Key Insights	37
5.	Conclusion and Further Implementations	39
5.1	Business Extension and Cross-Industry Applications of the Work.....	39
5.1.1	Strategic Advantages for Automakers	40
6.	Appendix	42
6.1	Network evolution of Germany.....	42
6.2	Per-year optimization results	45
6.3	Data Analysis of EV Infrastructure in Germany	52
	Bibliography.....	IV

List of Figures

Figure 1: Initial population on Wave function	20
Figure 2: Selected individuals from the initial population (cf. Figure 1). The individuals closest to the peak are preserved since they have the best fitness.....	20
Figure 3: Example of two constructed network: 2013 (a) and 2015 (b).....	29
Figure 4: Method overview.....	33
Figure 5: Optimized network (a) and real network (b) for year 2015	34
Figure 6: Change of the fitness over the generations with the value of the real network fitness (2015) as reference.....	35
Figure 7: Starting network (2011)	36
Figure 8: Simulated and real network in 2012.....	36
Figure 9: Change of the fitness over the generations with the value of the real network fitness (2012) as reference.....	37
Figure 10: Evolution of the charging station network in Germany from 2009 to 2022 ..	44
Figure 11: Simulated and real network in 2011	45
Figure 12: Simulated and real network in 2012.....	46
Figure 13: Simulated and real network in 2013.....	47
Figure 14: Simulated and real network in 2014.....	48
Figure 15: Simulated and real network in 2015.....	49
Figure 16: Simulated and real network in 2016.....	50
Figure 17: Simulated and real network in 2017.....	51
Figure 18: Cumulative histogram per year	53
Figure 19: Cumulative histogram per year divided into Normal and Fast charging	53
Figure 20: Power clustering histogram per year	53
Figure 21: Charging Type clustering per year.....	54

List of Tables

Table 1: Characteristics of the different network created	29
Table 2: Comparison between the obtained network and the real network in 2015	34
Table 3: Parameters comparison between the obtained network and the real network in 2011	45
Table 4: Parameters comparison between the obtained network and the real network in 2012	46
Table 5: Parameters comparison between the obtained network and the real network in 2013	47
Table 6: Parameters comparison between the obtained network and the real network in 2014	48
Table 7: Parameters comparison between the obtained network and the real network in 2015	49
Table 8: Parameters comparison between the obtained network and the real network in 2016	50
Table 9: Parameters comparison between the obtained network and the real network in 2017	51

List of Symbols

C_{avg}	—	Average clustering coefficient
C_v	—	Clustering coefficient of node v
E_v	—	Number of edges between the neighbors of node v
K_v	—	Number of vertices that are neighbors of node v
e_i	—	Edge with index i
v_i	—	Vertex with index i
$D(G)$	m	Diameter of graph G
G	—	Graph
$L(G)$	m	Average distance of graph G
M	—	Number of edges
N	—	Number of vertices
$dist(u, v)$	m	Distance between node u and v
E	—	Set of edges
V	—	Set of vertices
$\delta(G)$	—	Density of graph G

List of Abbreviations

EV	Electric Vehicles
EVCS	Electric Vehicle Charging Stations
FRLP	Flow Refueling Location Problem
GA	Genetic Algorithm
HGA	Hybrid Genetic Algorithm
KNN	K-Nearest Neighbors
MAE	Mean Absolute Error
MINLP	Mixed-Integer Non-Linear Programming
ML	Machine Learning
MSE	Mean Squared Error
PGA	Parallel Genetic Algorithm
RF	RandomForest
RVGA	Real-valued Genetic Algorithm

1. Introduction

In light of the escalating global emphasis on energy transition, the automotive industry has substantially reallocated its resources to adapt its production line. This realignment aims to cater to an increasingly informed and environmentally-conscious consumer base, seeking vehicles that meet personal preferences, such as range autonomy, and adhere to the rigorous emission standards imposed by regulatory bodies. Consequently, the electric vehicle (EV) domain has witnessed heightened attention from established European automotive manufacturers, complemented by new entrants from China.

Nonetheless, it became evident that the market penetration of electric and hybrid vehicles is intrinsically tied to consumers' accessibility to public charging infrastructure and, for homeowners, the feasibility of home-based charging installations. Recognizing this, national electricity providers embarked on expansive projects to increase charging stations. Yet, the absence of a coordinated strategy among these providers precipitated several challenges: disparate pricing models, barriers in accessing diverse suppliers, and a geographically suboptimal charging network.

Moreover, the inertia exhibited by service providers, coupled with bureaucratic encumbrances, often compelled automakers to engage in the charging infrastructure landscape directly. An illustrative example may be seen in Tesla's construction of the Supercharger network, which served as an effort to address the lack of charging infrastructure and as a deliberate tactic to enhance consumer fidelity.

Combining these facts shows that the transition towards sustainable mobility requires a comprehensive infrastructure architecture that effectively handles complex technological, economic, and environmental concerns.

1.1 Research Objectives

This study aims to provide a comprehensive and efficient optimization network genetic algorithm to locate and distribute electric vehicle charging stations across Germany.

The main goals can be summarized as follows:

- Modeling and building the network
- Analysis of the current EVCS network analysis, including its evolution, distribution of charging stations, and user accessibility
- Strengths and weakness of the existing infrastructure and highlight areas for improvement

- Implement the GA to optimize the location of new EVCS
- Evaluate the impact of various factors, such as technical specifications and spatial considerations
- Validate the optimization network by comparing the proposed EVCS network with the existing infrastructure

Through these objectives, this research aims to provide valuable insights and practical solutions to address the challenges faced by the EVCS network in Germany. This work will contribute to the advancement of EV infrastructure and promote the transition towards a more sustainable and efficient transportation system.

1.2 Thesis structure

The thesis is organized into six chapters, which provide a detailed explanation of the theoretical backgrounds, a thorough analysis of the model and its rules, and an analysis of the algorithm's results.

Chapter 2 embarks on a literature review concerning the primary determinants of electric vehicle market penetration. It comprises three concise paragraphs illustrating the factors considered when choosing a suitable site for a new charging station, the key variables that impact this decision, and a summary of the approaches used in existing research articles to enhance these variables.

Chapter 3 is dedicated to the core theoretical networks undergirding this research. The chapter has three segments: a discourse on network theory entailing a mathematical delineation and explication of four parameters instrumental in network description, a walkthrough of the foundational steps constituting a genetic algorithm, and a segment explaining the prediction methods employed to extrapolate the initial population of new charging stations predicated on historical installations.

Chapter 4 exposes the algorithmic steps. This chapter includes the presuppositions adopted for data pre-processing alongside the construction of the graph. Further explanation is provided on the initial population's generation and the optimization function's formulation.

Chapter 5 contains the conclusions drawn and suggests future implementations. In contrast, Chapter 6 (Appendices) comprises tables and graphical representations of all the simulations undertaken, which were not directly incorporated into Chapter 4 for reading fluency.

2. Literature review

2.1 Electric Vehicle Market Penetration Main Factors

The transformation of the global automobile industry toward electric mobility is underway. The following factors influence the EV market penetration:

Anxiety of range: The phenomenon termed “range anxiety” is common among potential EV owners. Range anxiety refers to the worry that an EV has insufficient reach to get to its destination and would leave the driver stranded. It is one of the key barriers to EV adoption [1]. This fear can be mitigated by the availability of a comprehensive, convenient, and reliable charging network, thus directly linking EV market penetration to charging infrastructure [2].

Infrastructure: A critical factor influencing the EV market penetration trend is the availability and accessibility of Electric Vehicle (EV) charging infrastructure. It has been widely studied and proven that the prevalence of charging stations strongly influences the acceptance and adoption of EVs [3] [4] [5] [6]. In the U.S., a positive correlation between public infrastructure and EV uptake has been proven [6]. When charging infrastructure was more available and accessible, the likelihood of consumers purchasing an EV increased. Moreover, a study examined the influence of various factors on international electric vehicle policies and EV sales in 30 countries [7]. They found that charging infrastructure was among the most substantial determinants of national electric vehicle market share [8].

Policies and regulations: Government policies also play a crucial role in the proliferation of charging stations and, consequently, the promotion of EV adoption. Incentives like grants, subsidies, and tax reductions have been applied in various countries to accelerate the deployment of EV charging stations [9]. Such policy measures have substantially impacted both the expansion of charging infrastructure and the adoption of electric vehicles.

EV prices and battery capacity: While the charging infrastructure is a significant factor, other elements like EV prices, government subsidies, and battery capabilities are also crucial contributors to EV market penetration [9]. Reductions in battery prices and improvements in battery technology have increased the affordability and driving range of EVs, thereby supporting the growth of the EV market. Nevertheless, widespread EV adoption remains challenging without an extensive and reliable charging network.

2.2 Selection of Charging Station Location

The decision regarding the ideal placement of Electric Vehicle Charging Stations (EVCS) is a complex process with many considerations. Several interrelated quantitative and qualitative factors influence this decision, including but not limited to operator economics, driver satisfaction, vehicle power loss, traffic congestion, and power grid safety [10].

Falvo et al.'s work (2011) illustrates the role of reducing energy consumption by exploiting the capabilities of existing power plants. They draw attention to the interconnectivity of different transportation systems (e.g., EVs and subways), highlighting the potential for symbiotic relationships to optimize power usage. Their research highlights the strategic importance of aligning EVCS locations with the current power grid for energy efficiency, operational economics, and grid safety. It serves as a reminder that the placement of charging stations should be an integral part of broader urban energy planning [11]. This work's assumptions are also used to develop this thesis. However, instead of using pre-existing transport networks, already-installed charging stations are used as input for the calculation.

Guo et al. (2015) present an alternative approach to the problem, employing a fuzzy TOPSIS method to assess potential locations. Their approach considers practical or economic factors and various environmental, economic, and social benchmarks. This underscores the importance of a holistic, multi-faceted evaluation process for locating EVCS. Beyond the fundamental requirements of power supply and accessibility, Guo et al. emphasize the need to assess potential locations' broader societal impact, environmental implications, and economic viability. These findings underscore that the decision-making process should not be limited to infrastructure and logistics alone but should strive to align with broader sustainable development goals [12]. This approach was also used in the proposed model. In fact, the model tries to go against the trend and expand the network even in rural areas. Furthermore, the initial placement method of the charging stations also follows an economic trend, given that it is based on the historical series.

Similarly, Asamer et al. (2016) propose a comprehensive, integrative approach to the placement of EVCS. They contend that several variables must be factored into the decision-making process, ranging from environmental conditions to power availability and legislative considerations. Significantly, they also highlight the importance of empirical data, employing taxi data as a proxy to assess charging demand. The utilization of real-world data, they suggest, can provide invaluable insights into patterns of use and potential

demand hotspots, thus allowing for more targeted and effective placement of charging stations [13].

Building upon this foundation, Zhu et al. (2018) introduce an economic perspective into the analysis, evaluating how costs - both to the user and those associated with establishing and operating the charging stations - impact the final number and location of EVCS. This underscores the need for a detailed cost-benefit analysis in the decision-making process. It also raises an important question of user satisfaction and accessibility, emphasizing that the locations must be convenient for the end users to encourage uptake and continued use of EVs [14]. Despite the interesting approach, it was not possible to implement these considerations in the algorithm as it was not possible to find data on the demand or cost of electricity or construction. However, the flexibility of the chosen model perfectly allows the integration of these variables if necessary for other applications.

Complementing these perspectives, Sun et al. (2020) propose an innovative, user-centric approach. They consider residents' travel patterns, categorizing them as either short-distance or long-distance travelers. This differentiation aids in determining not only the optimal location for charging stations but also the appropriate number of stations needed. It serves as a reminder that the deployment of EVCS should not be a one-size-fits-all solution. Instead, it should be tailored to meet residents' needs and ensure maximum usability and efficacy [15]. This aspect is also considered in this work. The first objective in adding new stations is to connect existing sub-networks. In this way, the journey from Munich to Berlin is ensured even when the departure networks do not allow it.

In summary, the complex interplay of factors affecting the location of EVCS necessitates a multi-faceted and integrative approach. The studies above underline the need for strategies that balance technical requirements, economic feasibility, societal impact, and end-user needs. This careful balancing act determines the optimal location for EVCS, thereby promoting widespread EV adoption and the resultant environmental benefits. The findings from these studies collectively demonstrate that the placement of EVCS is an intricate process, interweaving numerous factors and requiring comprehensive, multidimensional planning and assessment [10] - [15].

2.3 Optimization Models for EV Charging Station Distribution

This section proposes a literature review of the various optimization models for the distribution of EVCS. These models consider a broad spectrum of factors to help enhance the adoption of EVs and user satisfaction.

Frade et al. (2011) used a maximal covering model to identify potential demand areas and possible EVCS locations in Lisbon, aiming to maximize covered demands [16]. He et al. (2015) proposed a double-layer mathematical model considering vehicle driving distances and charging needs, underlining the importance of daily mobility patterns of EV users [17].

Shahraki et al. (2015) presented an optimization model maximizing vehicle mileage based on driving patterns, emphasizing the role of real-world data in location decisions [18]. Wu et al. (2017) designed a stochastic flow-capturing location model reflecting the randomness in EV users' traveling behavior [19].

Models by Tu et al. (2016) and Luo et al. (2018) included temporal and spatial constraints, making these models more realistic by considering variable parking availability, congestion levels, and EV owners' home and work locations [20] [21].

Battery characteristics have been factored into models by Liu et al. (2018) and Mehrjerdi et al. (2019), highlighting the need for different strategies for different charging applications and the importance of power and capacity of charging facilities [22] [23].

He et al. (2019) and Davidov et al. (2019) incorporated economic aspects in their models, considering costs such as battery, charging station, and energy storage system expenses [24] [25].

Hosseini et al. (2019) integrated quantitative and qualitative aspects into their models, underlining the importance of subjective factors and user experience [26].

Zeng et al. (2021) integrated human behavior into their station-level optimization framework, pointing out that station networks must accommodate user preferences [27].

Hodgson et al. (1990) further refined the models by considering EV charging during long trips and the limited range of EVs, which resulted in the Flow Refueling Location Problem (FRLP) [28].

The FRLP model has been extended to consider limited charging station capacity, alternative paths, different types of stations and vehicles, and congestion at stations [29] [30] [31] [32] [33] [34] [35]. Multi-period deterministic extensions of FRLP have been suggested, allowing for a dynamic opening of new stations and considering limited station capacity [36] [37].

Few studies have addressed uncertainties in EV charging infrastructure planning, such as unpredictable driving range or variability in recharging demand. Some recent works have

introduced the concept of portable charging stations and advocated for robust optimization approaches [38] [39] [40] [41] [42].

These diverse models demonstrate the multi-faceted nature of EVCS placement and the need for comprehensive, flexible approaches incorporating demand characteristics, technical specifications, cost factors, and human behaviors to promote EV adoption and environmental benefits [16] - [27].

2.4 Algorithm for Optimal EV Charging Station Configuration

The available literature proposes various optimization techniques to determine optimal configurations of EVCS. The methods include optimization methods, genetic algorithms, decomposition algorithms, clustering methods, and combinations of these techniques.

Sadeghi-Barzani et al. (2014) used a mixed-integer non-linear programming (MINLP) optimization method and a genetic algorithm to find the optimal EVCS location and scale [43]. Arslan et al. utilized the Benders decomposition algorithm to optimize EVCS location, aiming to maximize mileage and minimize transportation costs [44].

Zhang et al. (2014) introduced a decentralized valley-filling charging strategy that used a cost-minimization pricing scheme, emphasizing the importance of pricing mechanisms [45].

Zhu et al. (2016) and Akbari et al. (2018) used genetic algorithms for optimizing EVCS locations, acknowledging the interplay between technical specifications, user charging demand, and spatial factors [46] [47]. Awasthi et al. (2017) combined a genetic algorithm with particle swarm optimization to determine optimal EVCS location and size [48].

Particle swarm optimization algorithms have been used by Li et al. (2016) and Chen et al. (2018) to determine EVCS deployment strategy and charging facility distribution [32] [49]. Clustering methods like k-means cluster analysis were employed by Zhang et al. (2019) and Straka et al. (2019) to understand dynamic charging demand trends and user behavior [50] [51].

Wu et al. (2019) used approximate dynamic programming and an evolutionary algorithm to determine optimal charging start times for EVs, demonstrating an interest in intelligent charging strategies [52].

Despite these advancements, the article identifies gaps, including the lack of comprehensive analyses considering total social cost, underutilization of genetic

algorithms, and limited case studies of specific charging station scenarios, like those in Ireland.

To address these gaps, the article proposes an optimal distribution model of EVCS based on total social cost, utilizing a genetic algorithm for iterative simulation. This model deviates from traditional reliance on Euclidean distance and includes a road bending coefficient. The article also incorporates a case study of Ireland, reflecting actual EV charging demand in five major cities. This approach provides a promising direction for future EVCS optimization research [43] - [52].

Based on previous works, the most used model appears to be the genetic algorithm. This type of optimization also allows flexibility in choosing the parameters to optimize and creating multi-objective functions. Considering the literature, the parameters to consider will certainly be the viability (e.g., connection between points) and the reduction of costs or diffusion of the network.

2.5 Thesis novelty

This thesis introduces an innovative approach to optimizing the location of EVCS by leveraging graph theory in conjunction with genetic algorithms. This methodology represents a significant departure from traditional models that often prioritize demand-population metrics, primarily focusing on urban agglomerations. In contrast, the proposed model integrates a national graph structure where each charging station is represented as a node, and connections are based on actual road distances, including one-way streets and dead ends. This approach enables a more holistic and realistic representation of the EVCS network, emphasizing the importance of connectivity in rural areas as well as urban centers.

The novelty of this research lies in its dual focus on optimizing both the geographical spread and the functional efficiency of the charging network. By assigning weights to the edges of the graph based on the distance between points, the model goes beyond mere population-centric approaches. It ensures that the network is not just concentrated in major cities like Berlin but also extends to peripheral areas, thereby fostering a more inclusive EV infrastructure. This strategy is crucial for addressing range anxiety and promoting EV adoption across a broader spectrum of society, including those in less densely populated areas.

Furthermore, the use of a genetic algorithm for network optimization is another key aspect of this thesis. This algorithm's flexibility allows for the incorporation of multiple parameters,

such as viability and cost reduction, which are essential for a practical and sustainable EVCS network. The inclusion of predictive methods within the algorithm to identify optimal positions for the initial population is particularly noteworthy. This approach mimics the varying economic and political landscapes across different regions, ensuring that the network's expansion is not only technically sound but also contextually relevant.

Additionally, this thesis incorporates the relationship between the area covered by the EVCS network and the territorial extent of Germany. This consideration is vital for understanding the true impact of the network's reach, particularly in ensuring that long-distance travel, such as from Munich to Berlin, is feasible with the available charging infrastructure. By doing so, the research addresses a critical gap in current EVCS planning, which often overlooks the necessity of connecting urban and rural regions effectively.

In summary, this thesis presents a groundbreaking approach to EVCS optimization. It blends advanced computational techniques with a deep understanding of geographical and societal needs, setting a new standard for future research and development in this field. The methodology's emphasis on extending the network beyond urban centers to include rural areas is a significant contribution to the field, promoting more equitable and widespread adoption of electric vehicles. This approach not only enhances the current understanding of EVCS distribution but also paves the way for more sustainable and efficient transportation systems in the future.

3. Theoretical Background

This chapter contains all the theorems, algorithms, and network parameters used to develop the method (cf. Section 1.2). Section 3.1 mathematically describes the three primary parameters used to establish the quality of a network. Section 3.2 contains a general explanation of the genetic algorithm used to optimize the graph. Finally, Section 3.3 presents the two methods that were compared for the general initial population.

3.1 Network Theory

3.1.1 Diameter

The concept of graph diameter provides valuable insights into the general structural characteristics of a graph. The formal definition of the diameter graph refers to the maximum length of the shortest route between any two vertices in the graph. It represents the longest of all the shortest paths that can be navigated from one vertex to any other vertex within the graph [53]. To formally represent this concept, a graph as $G = (V, E)$ is defined, where $V = \{v_1, v_2, v_3, \dots, v_N\}$ is the set of N vertices and $E = \{e_1, e_2, e_3, \dots, e_M\}$ is the set of M edges that exist inside the graph. The shortest route distance, denoted as $dist(u, v)$, between any pair of vertices, u and v , in the set of vertices V , is stated in Equation 3-1.

$$D(G) = \max_{u, v \in V} dist(u, v) \quad 3-1$$

Here, the diameter D is the maximum over all shortest path distances $dist(u, v)$ for every pair of vertices $u \in V$ and $v \in V$ in the graph [54]. In the case of a disconnected graph, where there exist vertices that do not have a path between them, the diameter is often defined as infinity. For a graph containing only a single vertex, the diameter is defined as 0. The concept of the diameter of a graph is applied in numerous applications, such as in network design and analysis, devising algorithms for effective information routing in distributed systems, and discerning the structure of social networks [55]. It's important to note that calculating the diameter of a graph can be computationally expensive, particularly for large graphs. Several algorithms have been developed to compute the diameter of a graph, including but not limited to the Floyd-Warshall algorithm and Johnson's algorithm [54].

3.1.2 Average Distance

The average distance in a graph, also known as the average path length or the characteristic path length, is another crucial concept in graph theory. This metric indicates the overall navigability and connectivity of the graph. Formally, the average distance $L(G)$ is the mean shortest path length between all pairs of vertices in the graph. It represents the expected distance between two vertices chosen uniformly at random [55]. The average distance $L(G)$ can be given by the formula stated in Equation 3-2.

$$L(G) = \frac{1}{N \cdot (N - 1)} \sum_{u,v \in V} dist(u, v) \quad 3-2$$

The summation is performed over all distances between distinct vertices. The calculation is composed of the sum of the shortest route lengths between every possible pair of vertices divided by the total amount of vertices N [54]. The average distance is a critical metric in understanding the properties of real-world networks, which often exhibit small-world properties. The small-world property describes networks where the average path length is relatively small, meaning that one can get from any given node to any other node in the network through a few steps [55]. It's important to note that the average distance only makes sense for connected graphs, where a path exists between every pair of vertices. This metric can't be defined without modification if a graph is disconnected.

3.1.3 Average Clustering

The average clustering coefficient is a commonly used metric for assessing nodes' propensity to form clusters [55]. The metric offers valuable insights into the global clustering patterns shown by the network as well as the interconnections between its individual nodes. The quantification of the clustering coefficient for an individual node in a graph measures the degree of proximity between its neighboring nodes in terms of forming a full network. A full graph is defined as a graph in which each vertex is linked to every other vertex by a distinct edge. The average clustering coefficient is calculated as the arithmetic mean of the clustering coefficient of all nodes inside the network [55]. Considering the previous nomenclature for graph G , K_v is defined as the count of vertices that are next of node $v \in V$, and E_v as the count of edges connecting the adjacent vertices of node $v \in V$. The clustering coefficient C_v for the node $v \in V$ can be calculated using the Equation 3-3.

$$C_v = \frac{2 \cdot E_v}{K_v \cdot (K_v - 1)} \quad 3-3$$

The average clustering coefficient C_{avg} of the graph is then calculated as stated in Equation 3-4, where sum is over all vertices in the graph, and N is the total number of vertices [54].

$$C_{avg} = \sum_v \frac{C_v}{N} \quad 3-4$$

When computing the average clustering coefficient, it is essential to underline that this metric generally applies only to nodes with a minimum of two neighbors. This condition arises from the fact that the denominator of the formula used to calculate the clustering coefficient, denoted as C_v , would provide a zero value for nodes with less than two neighbors. Calculating the average clustering coefficient may pose significant processing challenges, mainly when dealing with more extensive networks. Various algorithms have been developed to estimate it more efficiently, but the computational complexity is generally high due to the need to examine the local neighborhood of each node [54].

3.1.4 Density

Network density is a metric that serves as a barometer for understanding the proportion of potential connections within a network that have been realized.

Formally speaking, the density of a network is computed as the ratio of the number of actual edges present over the maximum possible number of edges that the network could have given its vertices [55].

Considering the previous nomenclature for graph G , the maximum possible number of edges can be expressed as $N \cdot (N - 1)/2$. Thus, the density $\delta(G)$ can be formalized as in Equation 3-5 [56].

$$\delta(G) = \frac{2 \cdot M}{N \cdot (N - 1)} \quad 3-5$$

For a fully connected graph, the density will be 1, whereas a graph without edges will have a density of 0 [57].

The density of a network provides intuitive insights into its interconnectedness. A denser network signifies a more robust connection and possibly greater redundancy. Conversely, a sparser network might indicate modules or clusters and potential bottlenecks.

As for the other metrics, the computational effort for large networks can be a challenge. This has spurred a suite of algorithms and approximations [58].

3.2 Genetic Algorithm¹

Genetic Algorithms (GAs) are a family of search and optimization techniques inspired by the concept of natural selection [59]. Throughout time, GAs have been effectively used to resolve a vast variety of engineering challenges, covering function optimization, machine learning, and scheduling, as cited in reference [60]. This section provides a comprehensive overview of Genetic Algorithms, addressing their theoretical basis, elements, and variants. It also points out the positive and negative aspects of GAs.

The theoretical inspiration of GAs is based on the fundamental principles of biological evolution. Darwin's theory of evolution posits that individuals with superior physical fitness are selected for reproductive purposes, transmitting their advantageous traits to subsequent generations. Over a temporal duration, this phenomenon leads to the adaptation of species to their environment, thereby increasing their probability of survival.

In GAs, a population consists of a set of potential solutions, each referred to as individuals. Genetic operators such as selection, crossover, and mutation are applied to this population. These operators simulate an evolutionary process that occurs over multiple generations [59]. The iterative nature of GAs enables efficient search space exploration and convergence to optimal or near-optimal solutions [61].

3.2.1 Algorithm components

Representation is an essential component of Genetic Algorithms, as it defines how potential solutions are encoded for the manipulation within the algorithm. The most used format is the binary string, where each possible answer is represented as a series of binary digits (0s and 1s) [59]. According to the nature of the issue, different representations, such as real-valued vectors, integers, and permutations, may also be utilized [60].

An **initial population** of potential solutions is created during the startup phase (cf. Figure 1). Both random and domain-specific information may be used to create this population [62]. The choice of population size is based on the complexity of the issue at hand and the preferred ratio of exploitation to exploration [63].

¹ This section is a slightly modified version of the Section 4.2. "Genetic algorithm" of the thesis "Automation of Design Synthesis for Electric Vehicle Transmission" published by the same author of this thesis and has been reproduced here with the permission of the copyright holder [80]

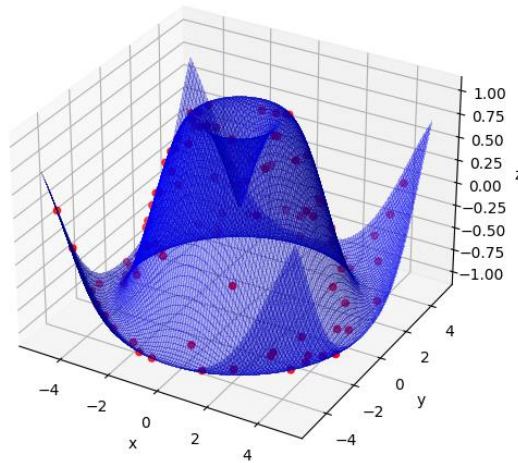


Figure 1: Initial population on Wave function

The **fitness function** is a way to measure the “quality” of each individual in the population. It acts as a gauge of how effectively the solution fulfills both the objectives and the constraints of the challenge [59]. The fitness function must represent the intended result and direct the search for the best solutions [64].

Selection is the process of choosing individuals from the present population to engage in reproduction to have the population aggregate around an ideal solution (cf. Figure 2). This mechanism prefers those with greater fitness levels, which mimes the concept of “survival of the fittest” [59]. Numerous selection methods have been proposed in the literature, including proportional selection [60], tournament selection [65], and ranking-based selection [66].

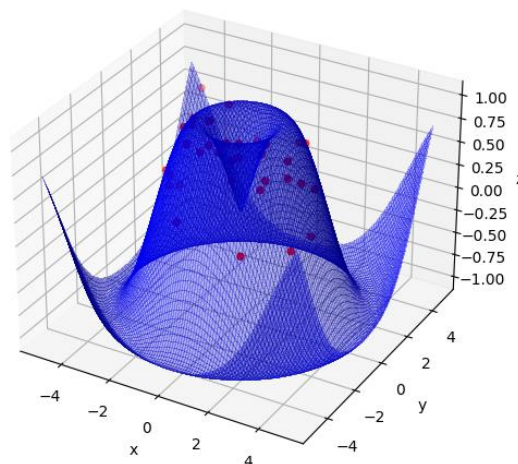


Figure 2: Selected individuals from the initial population (cf. Figure 1). The individuals closest to the peak are preserved since they have the best fitness

Integrating the genetic material from two parent solutions to produce one or more child solutions is known as a **crossover**, often referred to as recombination. Crossover aims to

explore the search area by developing new solutions that may have higher fitness values [59]. Common crossover operators include single-point, multi-point, and uniform crossover [60].

Elitism is an optional element of genetic algorithms that guarantees the preservation of the best person(s), called the incumbent, in each generation [61]. Elitism works to stop the loss of the best solutions brought on by genetic operators (e.g., crossover) and to direct search efforts in the direction of the ideal solution by keeping a copy of the fittest individual(s) [60].

3.2.2 Variants

In order to improve their efficiency and customize them for specific problem domains, many variations of genetic algorithms have been presented. Below is a discussion of a few of the more common variations:

- **Real-valued Genetic Algorithms (RVGAs):** employ real-valued representations rather than binary strings, enabling a more accurate and effective description of certain problem domains [67]. Arithmetic crossover [68] and non-uniform mutation [69] are two examples of the specific crossover and mutation operators needed for RVGAs.
- **Constraint-Handling Techniques:** GAs must consider restrictions to solve optimization issues in the real world effectively. Numerous methods for addressing constraints have been put forward, such as penalty functions [70], repair algorithms [71], and multi-objective approaches [64].
- **Hybrid Genetic Algorithms (HGAs):** it combines GAs with additional optimization methods to enhance convergence and search effectiveness [72]. Hill climbing or simulated annealing are two examples of local search techniques that are often used in HGAs to improve the solutions produced by GAs [73].
- **Parallel Genetic Algorithms (PGAs):** it uses of GAs' built-in parallelism to improve performance [74]. PGAs may be implemented using various parallelization techniques, including island models, fine-grained parallelization, and global parallelization.

3.2.3 Strengths and Weakness of Genetic Algorithms

Strengths:

- **Global search capability:** given their lower propensity to get stuck in local optima than gradient-based approaches, GAs are well-suited for identifying global optima in complicated search domains [60].
- **Applicability:** GAs can be applied to a wide range of optimization problems, as they only require the definition of a fitness function and an appropriate representation [59].
- **Robustness:** GAs are resilient and simple to apply in real-world problems since they are comparatively insensitive to parameter selection [61].
- **Parallelism:** GAs may significantly speed up by leveraging parallel computer resources since they are inherently parallelizable [74].

Weakness:

- **Slow convergence:** GAs may take a while to converge, particularly in cases where the search spaces are big and complex or the population is vast [60].
- **Premature convergence:** GAs can sometimes converge prematurely to suboptimal solutions, mainly if the selection pressure is too high or the mutation rate is too low [61].
- **Problem-dependent performance:** depending on the problem domain and the selected representation, crossover, and mutation operators, the performance of GAs might vary dramatically [60].

Parameter tuning: although GAs are generally reliable, determining the ideal set of parameters for a given issue may be difficult and time-consuming [63].

3.3 Predictions Methods

3.3.1 Random Forest

Random Forest (RF) is a versatile and popular learning technique that combines multiple decision trees to produce a more accurate and generalizable prediction. It was first introduced by Breiman in 2001 [75] and has since become a staple in the machine learning toolkit.

The primary intuition behind Random Forest is to create multiple decision trees during training and then aggregate their results (either by averaging for regression problems or voting for classification) to predict the final outcome. Each decision tree is built on a subset of the data, using a subset of the features. This kind of bootstrapped aggregation is known as “bagging”, which is instrumental in reducing variance and overfitting [76].

The RF has different advantages and peculiarities:

- **Robustness to Overfitting:** Because each tree in the forest is trained on a random subset of the data and features, the model is less likely to overfit the training data. This inherent diversity leads to a more generalizable model.
- **Feature Importance:** One of the unique aspects of Random Forest is its ability to gauge the importance of each feature. The importance is calculated based on how often a feature is used to split data across all trees and how much it improves the prediction.
- **Flexibility:** Random Forest can be used for classification and regression tasks, making it a versatile algorithm.
- **Handling Missing Data:** Random Forest can handle missing data more gracefully than many other algorithms. During training, the algorithm learns the best imputation strategy based on the existing data, and during prediction, it can use similar strategies to deal with missing features [77].

Although Random Forest has some benefits, it is important to acknowledge its limits. For instance, the computing demands might be substantial, particularly when dealing with a huge quantity of trees or a dataset with a high number of dimensions. Furthermore, while the model provides valuable information on the significance of features, it might be considered a black-box since it lacks the ability to deliver straightforward, rule-based logic similar to that of a single decision tree.

3.3.2 K-Nearest Neighbors

K-Nearest Neighbors is a type of instance-based learning where the algorithm does not explicitly learn a model. Instead, it memorizes the training dataset and makes predictions based on the proximity of new data points to the known data points. Introduced as early as the 1950s [78], KNN has remained a simple and powerful algorithm for various classification and regression tasks.

Given a new data point, the KNN algorithm searches the training dataset for the “K” training examples closest to the point. The prediction is then made based on the output values of these K-nearest neighbors.

The KNN has different advantages and peculiarities:

- **Simplicity:** The core concept behind KNN is straightforward, making it easy to implement and understand.

- **Versatility:** Like Random Forest, KNN can be used for classification and regression tasks.
- **No Training Phase:** Since KNN doesn't explicitly build a model, there's no training phase. This can be an advantage in settings where real-time decisions are required. However, it also means that the prediction phase can be computationally intensive. [79]

Despite its advantages, KNN has its limitations:

- **Computationally Intensive During Prediction:** Since KNN requires a distance calculation with all points in the training dataset for each prediction, it can be slow, especially with large datasets.
- **Sensitive to Irrelevant Features:** Since the algorithm relies on calculating distances, it's susceptible to irrelevant or redundant features, which can skew the distances and hence the predictions. Feature scaling and selection are crucial when working with KNN [79].
- **Choice of Distance Metric:** The choice of distance metric (e.g., Euclidean, Manhattan, Minkowski, etc.) can significantly impact the performance of KNN. The best metric often depends on the nature of the data.

4. Methodology and Results

The analytical foundation of this thesis is rooted in the dataset procured from the Federal Network Agency for Electricity, Gas, Telecommunications, Post, and Railways (Bundesnetzagentur für Elektrizität, Gas, Telekommunikation, Post und Eisenbahnen). This dataset undergoes annual updates, with the current version being refreshed in November 2022. It encapsulates comprehensive details pertaining to charging stations, encompassing the operator's identity, geographical specifics (including address, city, federal state, latitude, and longitude), and intricate technical parameters such as the type and quantity of connectors and the power delivered.

Upon rigorous data pre-processing (cf. Section 4.1), it becomes feasible to architect the charging network by interlinking each node with all proximate nodes within a 100 km road radius. This specified distance, determined by user discretion, offers an optimal compromise for Plug-in vehicles, especially when juxtaposed with the prevalent ranges of pure Electric Vehicles (EVs), which exceed 300km. Utilizing these extended ranges would have culminated in an overwhelmingly dense graph, rendering the graphical distinctions relatively indistinct (cf. Section 4.2).

Following the network's formulation, the subsequent phase involves the optimization of prospective charging stations' positioning. It is pivotal to note that the algorithm retains the spatial orientation of the extant charging stations, given the impracticality of alterations. Instead, it assimilates a predetermined number of new stations and refines their placement through a genetic algorithm (cf. Section 4.3).

Historical data served as the benchmark for algorithmic outcomes. An illustrative example is the installation of 532 charging stations between 2014 and 2015. The algorithmic model, initialized with the 2014 network and augmented with 532 stations, was juxtaposed against the 2015 network. Both networks maintain consistent pre-2014 station locations, with discrepancies manifesting solely in the orientation of the 532 newly incorporated stations.

To bolster the model's authenticity, the initial population derivation steered clear of randomness and instead hinged on a predictive algorithm contingent on location and federal state. This assumption predicates on the premise that federal states exhibit distinct adoption rates. The aforementioned predictive algorithm employs retrospective training levels. For instance, in the 2014 scenario, the initial population's prediction leveraged data spanning 2009 to 2014, eschewing the incongruity of simulating the 2015 station locations based on post-2019 occurrences (cf. Section 4.3.1).

Ensuring the model's realism extended to managing mutations and crossovers such that new points were seamlessly integrated into the road network. This precautionary measure mitigates the risk of placing a station in non-practical locations, such as a park or forest. In such instances, relocation to the nearest accessible road point was mandated, ensuring that the genetic algorithm retained its discrete nature.

The final layer of refinement encompassed the implementation of a multi-objective fitness function. This function, tailored for selecting the most viable individuals, amalgamated geographical considerations and graph-based metrics (cf. Section 4.3.2).

4.1 Pre-Processing

The raw dataset, procured from the Federal Network Agency, underwent meticulous pre-processing to render it suitable for subsequent analysis. This pre-processing has involved several steps to enhance the data's quality and readability, thereby facilitating more accurate and effective analysis. One key aspect of the pre-processing has been the elimination of duplicate entries from the dataset. It is common for datasets, particularly those aggregated from multiple sources or entered manually, to contain duplicate entries. If not removed, duplicate entries can lead to redundancy in the data and distort the subsequent analysis. Thus, duplicate entries in the dataset have been identified and carefully pruned to create a streamlined and precise representation of the unique electric vehicle charging stations distributed across the regions. Following this, the data has undergone a process of transformation aimed at standardizing its structure and making it more conducive to analysis. A prime example of this is the transformation of the *'type_of_charger'* variable. The original values, *'Schnellladeeinrichtung'* and *'Normalladeeinrichtung'*, have been remapped to 'fast' and 'normal', respectively. This alteration simplified the task of classifying charging stations based on their charging speed, thereby paving the way for more straightforward subsequent analysis.

In the original dataset, geographical coordinates have been presented in a format unsuitable for computational processing, with commas employed as decimal points. This posed significant challenges for any computational or spatial analyses that relied on these coordinates. The commas have been replaced with decimal points to address this, and these columns converted to the float data type. Consequently, calculations involving distances have become significantly more straightforward, and the feasibility of creating detailed geospatial visualizations was substantially improved. Another vital transformation step involved the *'commissioning_date'* column, which has stored the commissioning date

of each charging station as a string in the day-month-year format. It was necessary to convert this data into a datetime format for a more detailed and useful chronological analysis. This modification facilitated an overview of the temporal progression of charging station installations, offering valuable insights into their spread over time.

Additionally, leading and trailing spaces found in several dataset columns were removed to maintain uniformity in the data. Ensuring this consistency was especially important for object columns, such as city names, where minor differences in formatting could lead to inaccurate grouping or comparison of values. The city names present in the dataset were also standardized for improved accuracy and simplicity. Distinct neighborhoods or districts within a city initially represented separately, were consolidated under the name of the main city. This step significantly reduced geographical complexity during the analysis, fostering a more coherent data representation. In a further move to streamline the dataset, certain columns containing public keys were dropped. Although these keys are commonly used for cryptographic operations, they were deemed irrelevant to the intended study and were therefore excluded from the dataset, leading to a more simplified data structure. In the final pre-processing stage, the refined dataset was saved into a new CSV file termed '*ChargingStationCleaned.csv*'. This file stands as a meticulously cleaned and pre-processed version of the original dataset, primed for a detailed examination of the distribution, characteristics, and evolution of electric vehicle charging stations across Germany.

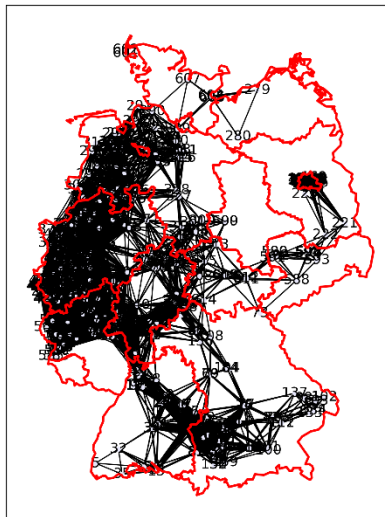
4.2 Network construction

To further analyze the distribution and accessibility of electric vehicle charging stations across various geographical regions, a process for network creation was deployed. The outcome was a graphical representation, visualizing connections between the charging stations for specified years and within user-determined distance parameters. Each with a unique role, several programming libraries were brought into play in this complex process. The `sqlite3` library provided the means for interaction with a SQLite database. The `requests` library was employed to handle HTTP requests required for distance calculation, whereas the `math` library was used to compute intricate mathematical operations, including trigonometric functions necessary for distance calculations. The `urllib3` library, an HTTP client for Python, managed and streamlined HTTP protocol details. The `Transformer` module from the `pyproj` library was crucial for geographic coordinate conversions.

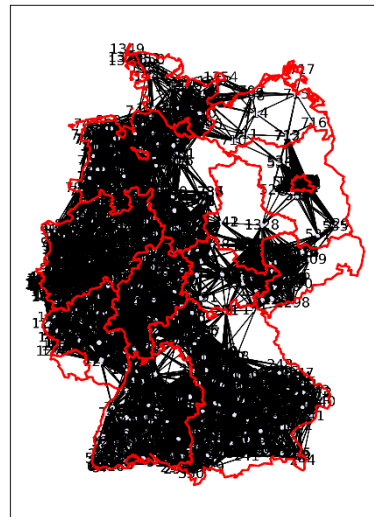
An essential part of this procedure was calculating the Haversine distance, a method utilized to determine the ‘great-circle’ distance between two points on the Earth’s surface, given their longitudes and latitudes. Here, the Earth was assumed to be a perfect sphere, and its radius was taken as 6371 kilometers. The function `get_osrm_distance` was developed for more accurate distance measurements. This function initially made a quick estimation based on the Haversine formula. If this approximated distance was found to be less than or equal to 100 kilometers, the Open-Source Routing Machine (OSRM) service was employed. OSRM is an open-source service that uses data from the OpenStreetMap project to find the shortest routes between coordinates. Efficiency and speed were prioritized by creating a SQLite database named ‘distances.db’, intended to save previously calculated distances. This efficient system considerably reduced the need for repetitive API calls to the OSRM service. If a previously calculated distance value was found in the database, it was retrieved and returned. If not, the function executes an API call to the OSRM service, acquires the precise distance, and subsequently stores this new distance value in the database for future use. The subsequent phase of the process was encapsulated in the `create_network` function. This function utilized the previously described distance calculations to construct an undirected graph representing the network of charging stations for a particular year and within a user-specified distance threshold. Each node in the graph corresponded to an electric vehicle charging station, assigned with specific attributes like latitude, longitude, and federal state. Edges between these nodes were added based on their distance, calculated by the `get_osrm_distance` function. After the graph’s construction, it was visualized on a map. For geographical reference, the boundaries of Germany were used as a backdrop. These boundaries were acquired from a GeoJSON file. The resulting visual output was a graph overlaid on Germany’s geographic boundaries, portraying the spread of charging stations and their connectivity. The final phase of this process was storing the constructed graph for future use. The graph was saved in the GraphML format, a comprehensive and easy-to-use file format for graphs, enabling potential future network analysis. All the networks constructed are shown in Section 6.1, while Table 1 contains the calculated parameters (cf. Section 3.1).

Year	Diameter	Average Distance	Average Clustering	Density	Total nodes	Subnetwork size
2009	1.00	15.10	0.93	1.00	60	[2, 2, 23, 4, 26]
2010	4.20	36.71	0.90	0.64	130	[14, 77, 38]
2011	8.27	41.38	0.86	0.42	320	[256, 61]
2012	11.00	43.65	0.85	0.18	494	[494]
2013	11.96	47.52	0.83	0.17	615	[613, 2]
2014	12.00	51.92	0.81	0.15	861	[861]
2015	11.00	50.75	0.80	0.11	1393	[1393]
2016	11.00	51.70	0.78	0.10	2503	[2503]
2017	10.00	53.26	0.77	0.09	4350	[4350]
2018	09.00	56.99	0.75	0.09	8135	[8135]
2019	09.00	57.93	0.76	0.09	13012	[13012]
2020	09.00	58.24	0.74	0.09	18103	[18103]
2021	09.00	58.43	0.77	0.10	25040	[25040]
2022	09.00	58.36	0.74	0.10	29481	[29481]

Table 1: Characteristics of the different network created



(a)



(b)

Figure 3: Example of two constructed network: 2013 (a) and 2015 (b)

4.3 Optimization

4.3.1 Initial Population

The first operation to perform the genetic algorithm is to generate the initial population. Although random positioning could be used, basing the positioning of new (future) charging stations, taking into account the distribution of past ones is a more effective method to simulate the real performance of the network. For this reason, two methods were tested: Random Forest (cf. Section 3.3.1) and K-Nearest Neighbors (cf. Section 3.3.2)

Initially, both models were trained and tested using the dataset columns containing information about existing charging stations, including location coordinates and the year of installation. The predictive performance of each model was evaluated based on three metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), and R2 score. The MAE measures the average magnitude of the errors in a set of predictions without considering their direction. MSE, similar to MAE, punishes more significant errors, which tends to be helpful in the real world as it punishes large errors. The R2 score measures the percentage of the variability in the dependent variable that can be explained by the independent variable(s) [80]. After the evaluation, it was seen that the Random Forest model exhibited superior performance compared to the K-Nearest Neighbors (KNN) model across all measures. This implies that the Random Forest model is more suitable for the specific prediction job at hand. The Random Forest (RF) model strongly predicted latitude and longitude, as seen by R2 scores of 0.921 and 0.867, respectively. These scores indicate that the model accounted for about 92.1% and 86.7% of the variability observed in the test dataset for latitude and longitude, respectively. Following the evaluation, the chosen RF model was trained on the entire dataset, divided by year. This process allows for the capture of annual trends in the establishment of charging stations, thereby providing a robust model that accounts for temporal variations. The trained models were then deployed in the genetic algorithm to generate an initial population, which suggests promising sites for new charging stations. Each “individual” in this application represents a potential layout of new charging stations. The predictive models were used to estimate the location of new stations in each federal state based on the state’s proportion of total stations. This approach ensures that the stations are distributed to reflect the historical patterns of station placement while also considering the need for more stations in areas with higher demand. The integration of predictive models into the genetic algorithm

allowed for a data-driven approach to generate initial populations, potentially improving the efficiency and effectiveness of the optimization process.

4.3.2 Fitness Function

In the GA for the charging station network optimization, the fitness function serves as a sort parameter of the individual. It balances the need for efficient charging station placements with the goal of wide-reaching accessibility across urban and non-urban areas. The function is designed as a multi-objective function that captures specific attributes of the desired optimal network.

The objectives considered in the fitness function are:

- **Network-to-Country Area Ratio:** this metric shows how uniformly the charging station network is spread across Germany. By comparing the area covered by the network to Germany's total areas, it gauges the spatial distribution of the stations. The network area is determined by creating a convex hull around all charging stations. The process leads to a polygon that envelops all nodes, representing the spatial footprint of the network. The ratio serves as an indicator of the special decentralization. A higher value means better spread, avoiding the risk of the network being too concentrated in particular regions.
- **Average Distance:** this metric reflects the average pairwise shortest path between stations (cf. Section 3.1.2). It measures the network's navigability and indicates how conveniently one can transition from one station to another. Decreasing the distance would lead to closer vertices, but it also increases the risk of clusters and big concentrations in already existing urban areas.
- **Diameter:** this metric captures its "largest smallest" path. It represents the longest distance one must travel between any two stations when taking the shortest possible route (cf. Section 3.1.1). This metric is an indicator of the network's maximum reach. Decreasing this metric means ensuring reducing the stops to reach two far vertices (e.g., decreasing the number of charging sessions between Munich and Hamburg)
- **Average Clustering:** this metric is a representation of the interconnectedness of stations (cf. Section 3.1.3). High clustering would mean stations are part of well-knit local groups, ensuring redundancy and better accessibility within federal states but not across them.

- **Density:** this metric measures the actual connections versus the potential connections in the network (cf. Section 3.1.4). A denser network would mean more stations are directly accessible from any given station, reducing transit times. However, a denser network also tends to add stations close to other clusters, increasing the disparity between urban and rural zones.

All the above metrics operate on different scales (e.g., average clustering is between 0 and 1, while average distance can be bigger than 50). For a balanced assessment, all the metrics were normalized to fall between 0 and 1, ensuring no single metric took absolute advantage of the fitness score. Additionally, weights are assigned to the metrics to give more importance to a specific characteristic.

4.4 Workflow overview

Figure 4 shows the schematic representation of the methodology employed in this research. The framework is underpinned by a series of Python scripts that manage the data processing pipeline. It is important to clarify that this illustration abstracts from certain components, such as the API calls for road distances and federal boundaries, to maintain clarity and focus on the core methodology.

The workflow is bifurcated into two distinct phases: data/model generation and optimization.

Data/Model Generation Phase: this initial phase is dedicated to preparing and transforming the input data. Commencing with the pre-processing step, the raw data furnished by the ministry in CSV format is subjected to a cleaning process, the outcome of which is also in CSV format (cf. Section 4.1). After that, the predictive model is calibrated using the cleaned data, yielding the trained model's weights corresponding to each year; for instance, the file labeled '2012.rf' encapsulates the model trained up to 2012 (cf. Section 3.3). Following this, network structures are constructed and archived, exemplified by the '2010.graphml' file, which encapsulates the network as it stood in 2010. This phase culminates with storing all the generated models and datasets (cf. Section 4.2).

Optimization Phase: this phase is managed within a singular script and is contingent upon user-defined parameters, such as the year of interest and the proposition of new charging stations. When the specified year is in the past, the default setting for new charging stations reverts to the actual number of stations installed that year.

For illustrative purposes, should the user parameters indicate the year 2015, the ensuing operation would be the generation of a network optimized for the year 2016 (denoted as ‘year+1’). The script uses the ‘2015.rf’ file to forecast the initial population, predicated on data up to 2015 (cf. Section 4.3.1) and utilizes the ‘2015.graphml’ network file. The resultant output is the ‘2016.graphml’ file, which is then segregated into a distinct directory from the actualized network files.

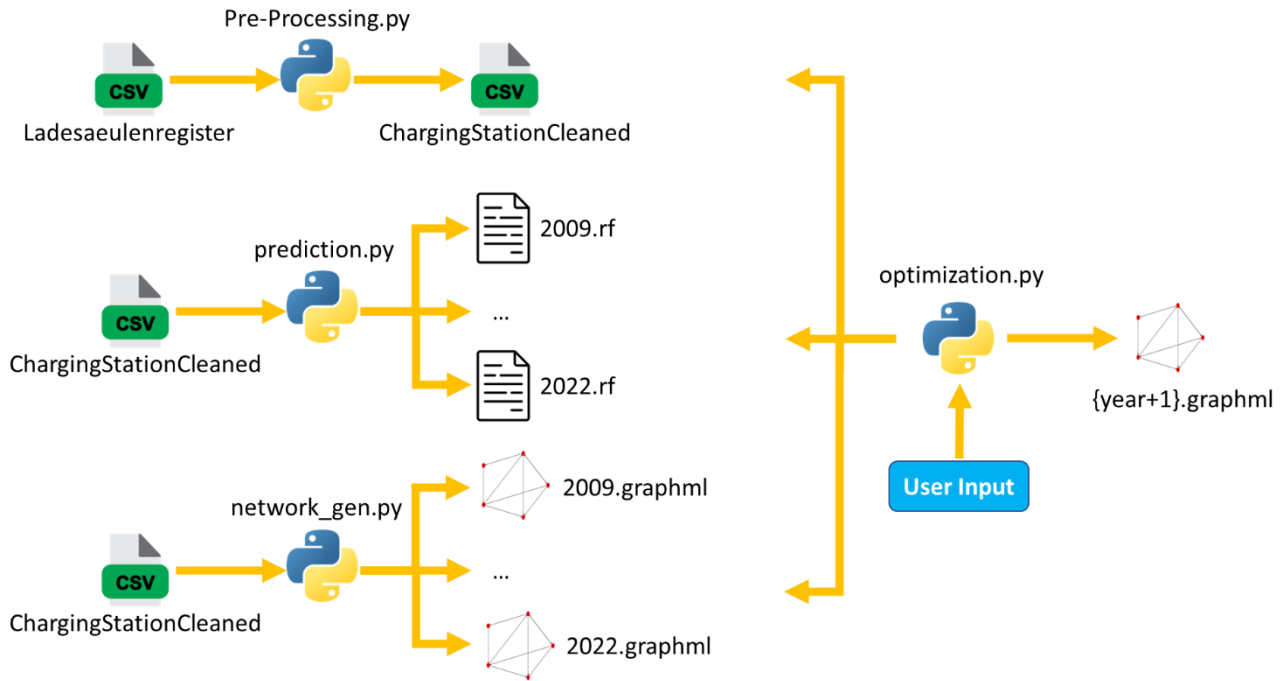


Figure 4: Method overview

4.5 Results and Test Cases

4.5.1 Network 2015: Spread Across the Territory

A clear illustration of the genetic algorithm’s efficacy in optimizing the distribution of the charging network across Germany can be discerned from the 2015 data. As previously delineated, the basis structure is the 2014 network (refer to Figure 14 (b)), augmented by the inclusion of 532 additional stations. The initial population, a confluence of the entrenched 2014 network and the novel stations, comprises 60 individuals, a user-defined parameter. The algorithm’s max generation is set at 30, with embedded stopping criteria triggered if the fitness remains static over six consecutive generations.

Figure 6 delineates the evolution of the fitness function for the incumbent, defined as the most optimal individual within the population. Notably, the algorithm’s performance surpasses the benchmark established by the authentic 2015 network’s fitness value. Table

2 compares the quantitative outcomes derived from the genetic algorithm against the real network. A discerning analysis reveals that, through the expense of the mean inter-nodal distance, the resultant network exhibits superior quality, manifesting a more equitable distribution across the German landscape.

This spatial disparity is visually presented in Figure 3 (a). Distinct regions, including Oberfranken (Bayern), Oberlausitz (Sachsen), Sachsen-Anhalt (entire Land), Lüneburg (Niedersachsen), Mecklenburg-Vorpommern (entire Land), and Brandenburg (entire Land), evince enhanced coverage. Remarkably, this expansive territorial coverage was achieved without augmenting the total number of stations, underscoring the algorithm's prowess in optimizing spatial distribution.

Parameters	Opt	Real
Density	0.1089	0.11
Average Distance	58.406	50.75
Diameter	10.0	11.0
Average Clustering	0.709	0.80
Num. of stations	1393	1393
Fitness	0.3981	0.4459
New installed stations	532	532

Table 2: Comparison between the obtained network and the real network in 2015

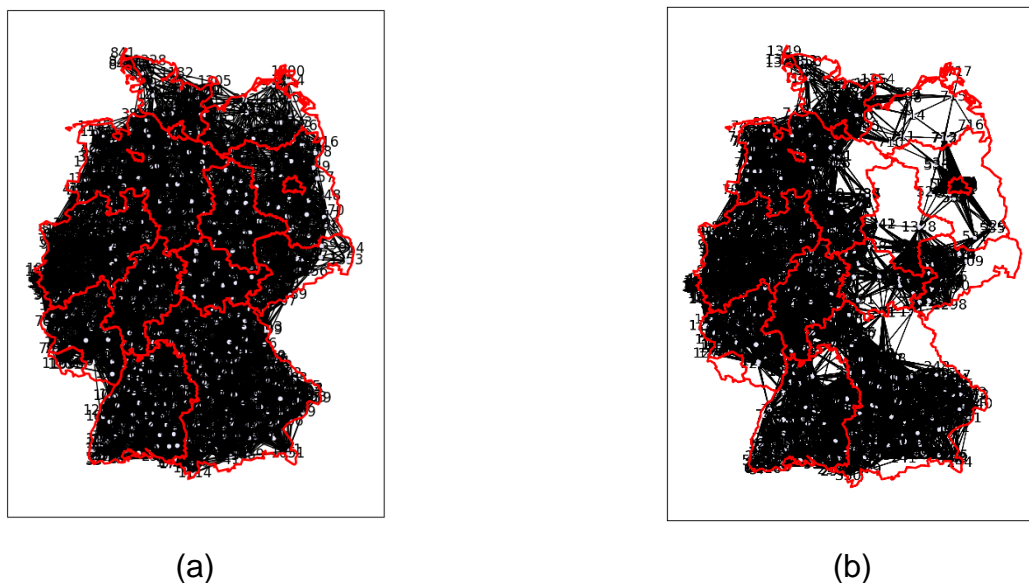


Figure 5: Optimized network (a) and real network (b) for year 2015

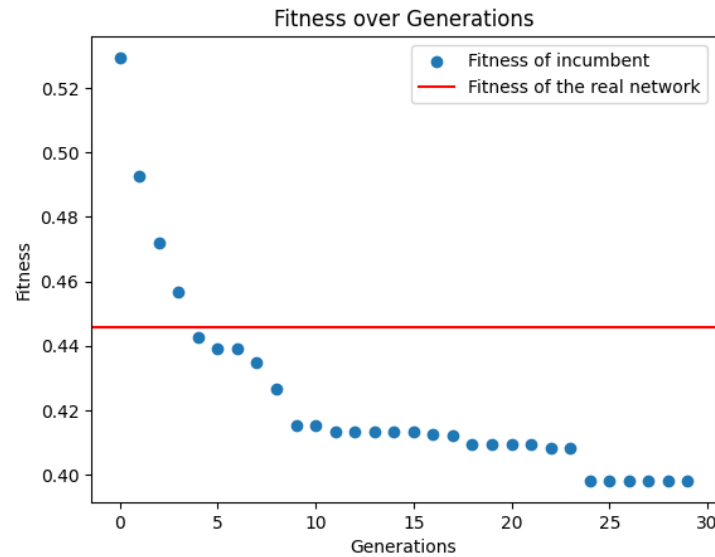


Figure 6: Change of the fitness over the generations with the value of the real network fitness (2015) as reference

4.5.2 Network 2012: Connection of Separated Subnets

The second test case presented within this thesis underscores a distinct feature inherent to the model in question: the amalgamation of pre-existing sub-networks.

The 2011 charging network is invoked as a reference point for illustrative purposes. As delineated in Figure 7, the German charging infrastructure in 2011 predominantly spanned the regions corresponding to the erstwhile West Germany, with notable proliferation in Nordrhein-Westfalen, Niedersachsen, the southern of Hessen, and the northern territories of Rheinland-Pfalz. Conversely, Berlin exhibited a concentrated cluster of charging stations, albeit in stark isolation, given the absence of stations in neighboring Sachsen-Anhalt and the minimal presence in Thüringen, Sachsen, Mecklenburg-Vorpommern, and Schleswig-Holstein. Consequently, inter-city commutes, such as those between Munich and Berlin or Desda and Dortmund, were infeasible. Therefore, the model's inherent objective gravitated towards synthesizing a unified network from these disparate sub-networks, ensuring connectivity across all federal states.

In fulfilling this objective, the model exhibits commendable efficiency, as displayed in Figure 8, adeptly interlinking all sub-networks into a cohesive federal network. This enhancement is particularly conspicuous in regions such as Sachsen-Anhalt.

Reiterating the analytical approach of the prior exemplification, Figure 9 charts the evolutionary trajectory of fitness across generational iterations. Notably, given the expanded operational territory in this scenario, the delta between the algorithmically

derived value and the empirical benchmark is accentuated. The result of the simulation can be found in Section 6.2.

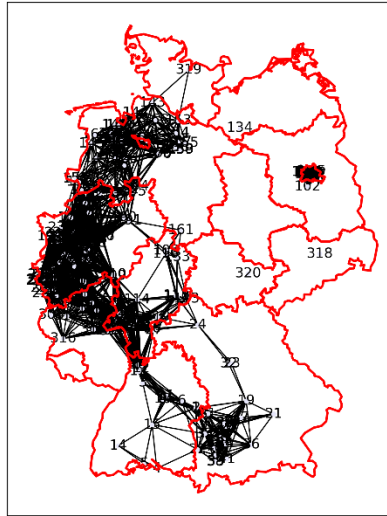
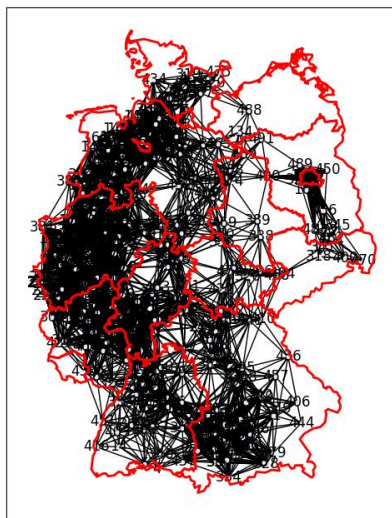
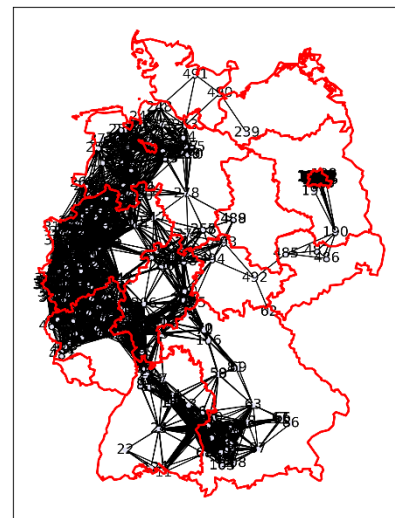


Figure 7: Starting network (2011)



(simulated)



(real)

Figure 8: Simulated and real network in 2012

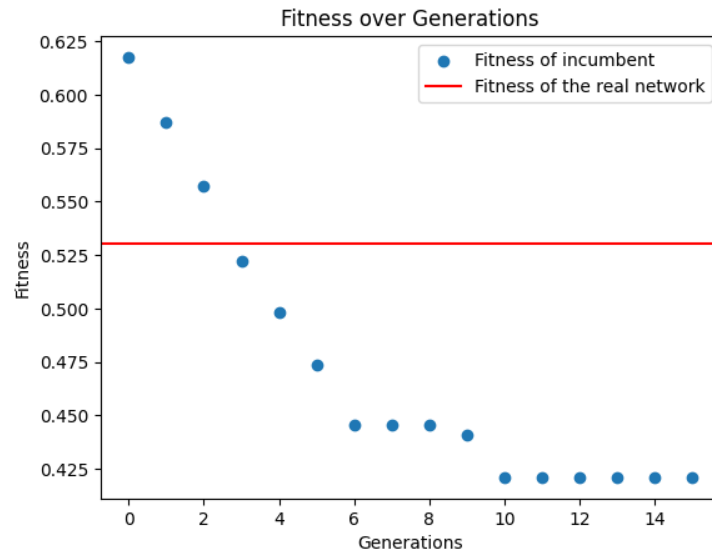


Figure 9: Change of the fitness over the generations with the value of the real network fitness (2012) as reference

4.5.3 Results Summary and Key Insights

This section presents a summary of the significant results obtained by using a GA to improve EV charging network in Germany over the years 2012 and 2015. The analysis focuses on the algorithm's proficiency in enhancing network distribution and connectivity.

Enhanced Distribution in 2015: The use of the GA enhanced the spatial arrangement of charging stations. The algorithm notably expanded the 2014 network by including 532 stations, which substantially improved coverage in areas such as Oberfranken, Oberlausitz, and Mecklenburg-Vorpommern. Although the optimized network kept the same number of stations, it exhibited a wide distribution and decreased average distances between nodes, suggesting improved network quality. Moreover, the algorithm's capacity to surpass the actual 2015 network in terms of fitness highlights its effectiveness in optimizing geographical distribution.

Connection of Subnets in 2012: The algorithm successfully resolved the issue of joining previously separate sub-networks in 2012, resulting in improved network cohesiveness across Germany. This was especially apparent in the consolidation of separate networks, which positively impacted areas like Sachsen-Anhalt and Berlin that had previously faced isolation. The algorithm improved the connection between places such as Munich and Berlin, making it easier to travel between them. This demonstrates the algorithm's effectiveness in creating a united network from smaller, separate networks.

Evolutionary Efficiency: Throughout both years, the genetic algorithm demonstrated remarkable efficacy. As shown in the corresponding pictures, the fitness function's

evolutionary trajectory presents the algorithm's capacity to exceed empirical standards and attain optimum configurations. This demonstrates the algorithm's resilience in adjusting to various operating situations and goals.

To summarize, the evolutionary algorithm has shown its effectiveness in improving the EV charging network in Germany. It has successfully achieved better geographical distribution and network unification, resulting in a more accessible and coherent charging infrastructure.

5. Conclusion and Further Implementations

This study examines the challenges associated with the development of the German electric vehicle charging network using graph theory as a foundational analytical tool. The objective was to develop an algorithmic approach to optimize the placement of new charging points, enhancing the connectivity and topological features of the resulting graph.

The proposed algorithm was rigorously tested on both continuous and discontinuous graph representations. In each instance, the algorithm demonstrated proficiency in bridging the connectivity gaps between major cities such as Berlin and Munich. Furthermore, the resulting graphs exhibited superior topological characteristics compared to their real-world counterparts, affirming the efficacy of the fitness function in capturing salient graph parameters.

To ensure applicability in real-world scenarios, the algorithm incorporates the generation of an initial population and a discrete genetic mechanism. This mechanism facilitates the repositioning of points on roads frequently used for vehicular transit, making the algorithm particularly pertinent during the implementation phase of charging infrastructure development.

Despite the promising outcomes, the model presents avenues for enhancement. Future iterations of the fitness function could benefit from the incorporation of additional parameters such as:

- Population density, which reflects potential user demand.
- Cost per kW to provide a cost-effectiveness analysis.
- Demand metrics, offering insights into charging point utilization.
- Implementation difficulty, offering a realistic feasibility measure.
- A differentiated weighting scheme that gives priority to fast transit routes over slower proximate roads, ensuring optimal positioning for high-speed routes.

5.1 Business Extension and Cross-Industry Applications of the Work

The optimization network that has been built for the EVCS has the potential to be adopted and used to address various placement and distribution challenges encountered across diverse industries. This work exhibits prospects for commercial expansions and cross-industry applications such as:

- **Transportation and Urban Planning:** the model adopted in this thesis may be utilized to improve the placement and dissemination of bike-sharing stations in

metropolitan regions to maximize accessibility and utilization. This has the potential to enhance the convenience of mobility and foster the use of environmentally-friendly modes of travel inside urban areas. Similarly, the approach may be used to ascertain the optimal placements of public transportation stations, minimizing the distances pedestrians must traverse and enhancing the efficacy of the public transit network.

- **Healthcare:** an optimization network in the healthcare sector can be used to identify optimal sites for hospitals, clinics, and other healthcare facilities. Furthermore, this methodology may be used to enhance the allocation of medical resources, including healthcare professionals among various healthcare establishments, ensuring the sufficient fulfillment of the population's medical requirements.
- **Retail and Services:** location optimization may be used in the retail sector to ascertain the most financially advantageous sites for retail outlets and service centers. This approach may be used to enhance the allocation of automated teller machines (ATMs) or bank branches to guarantee convenient accessibility of financial services for consumers.
- **Logistics and Supply Chain:** the method in the logistics and supply chain business may be utilized to optimize the placement of warehouses and distribution centers, aiming to reduce transportation costs and improve delivery times. Furthermore, it may be used to ascertain the optimal transportation vehicle routes, enhancing operational efficacy and reducing fuel usage.
- **Telecommunications:** the optimization method may be used within the telecommunications sector to ascertain the most advantageous positions for cell towers and other communication equipment to optimize coverage and signal quality. It can potentially enhance the efficiency of Wi-Fi hotspot deployment in public areas, guaranteeing optimal connection for individuals.

In conclusion, the methodologies and insights derived from this thesis have the potential to be applied across various industries to solve a wide range of location and distribution problems. The key is adapting the optimization network to each industry's specific needs and constraints, unlocking new opportunities and value in different business sectors.

5.1.1 Strategic Advantages for Automakers

This section explores the potential impact of the use of this work and its extension on automakers, particularly those producing electric vehicles (EVs) such as BMW, Mercedes, and others.

Targeting the Automaker Persona: Automakers are at the forefront of the EV revolution, constantly seeking ways to increase market penetration and customer satisfaction. The optimized EVCS network directly aligns with their strategic goals by addressing two critical areas: enhancing the attractiveness of EVs and reducing barriers to EV adoption.

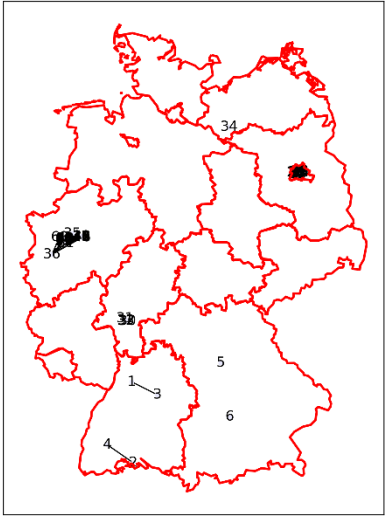
Financial Implications and Market Expansion: The initial investment in supporting the expansion of an optimized EVCS network can yield significant returns for automakers. By facilitating a more robust EV infrastructure, they contribute to increased EV sales and position themselves as leaders in sustainable transportation solutions. This approach can open new market opportunities and strengthen brand loyalty among environmentally conscious consumers.

The pains of limited charging options and long charging times can be mitigated through an optimized network, thus enhancing customer satisfaction and loyalty.

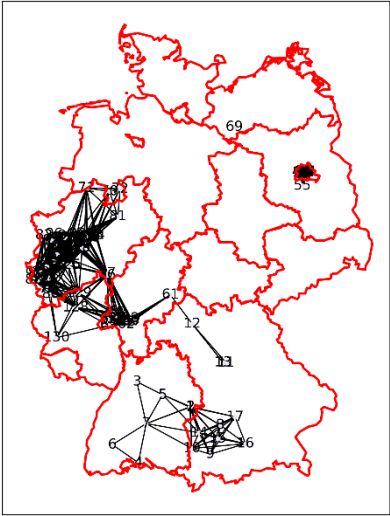
By addressing their customer base's specific needs and pains and actively participating in developing a more efficient charging infrastructure, they can drive greater EV adoption, open new markets, and establish a stronger competitive position in the evolving automotive industry.

6. Appendix

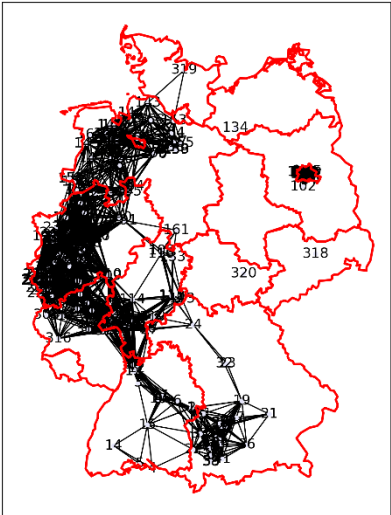
6.1 Network evolution of Germany



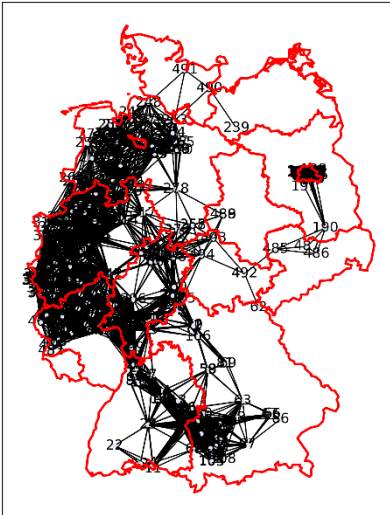
(2009)



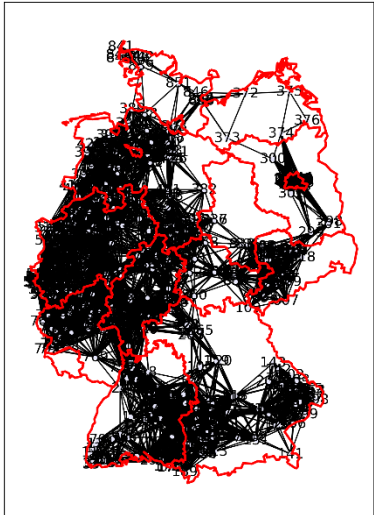
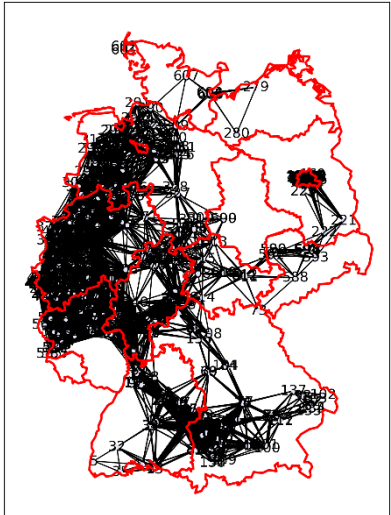
(2010)



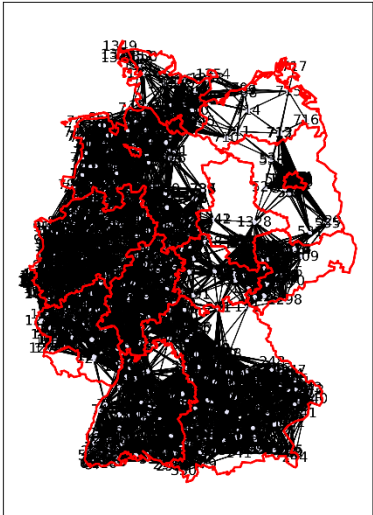
(2011)



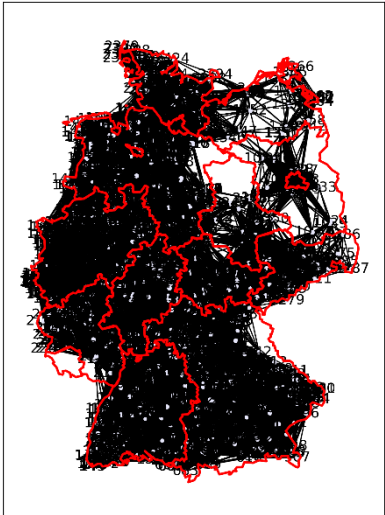
(2012)



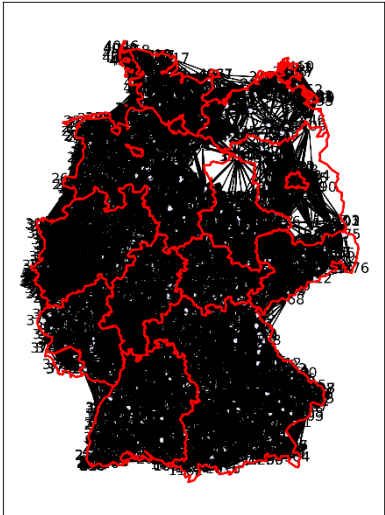
(2013)



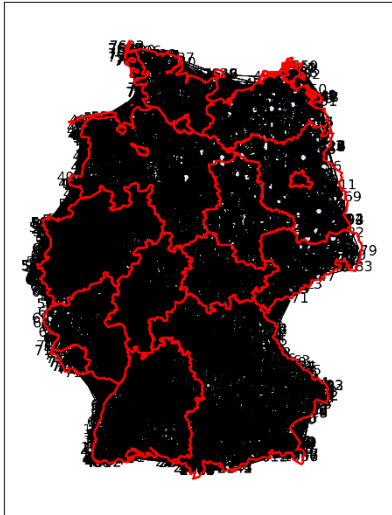
(2014)



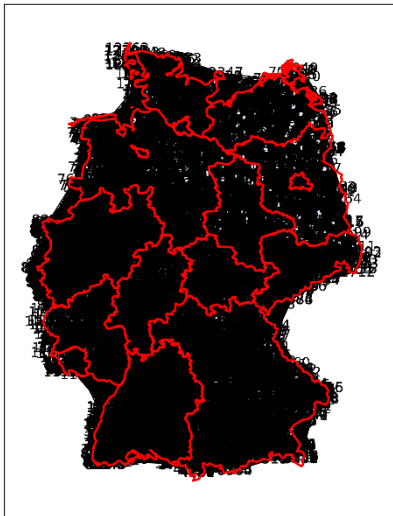
(2015)



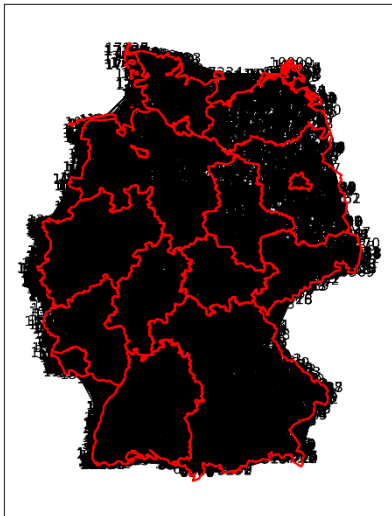
(2016)



(2017)



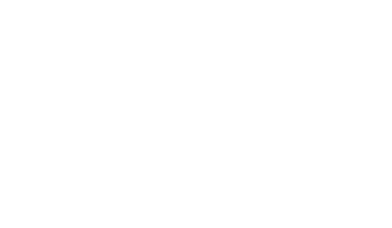
(2018)



(2019)



(2020)



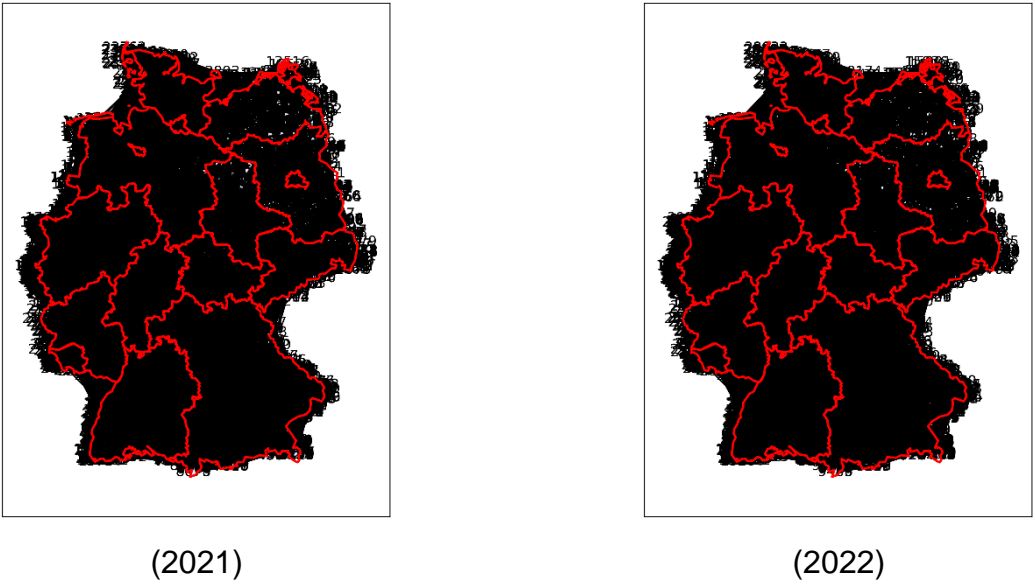


Figure 10: Evolution of the charging station network in Germany from 2009 to 2022

6.2 Per-year optimization results

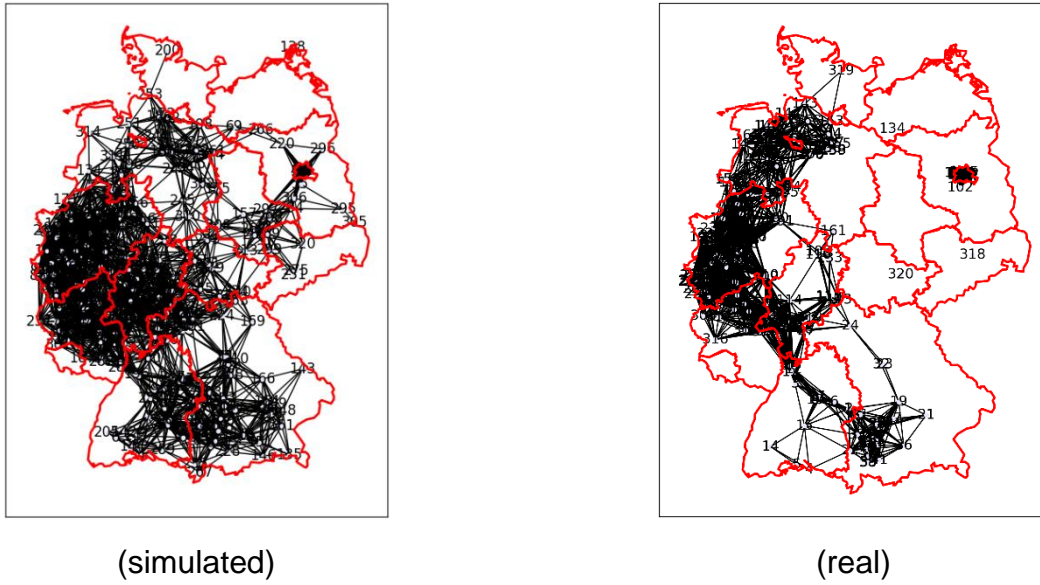


Figure 11: Simulated and real network in 2011

Parameters	Opt	Real
Density	0.125	0.42
Average Distance	53.786	41.38
Diameter	10.0	8.27
Average Clustering	0.712	0.86
Num. of stations	319	[256,61]
Fitness	0.4642	0.4912
New installed stations	188	188

Table 3: Parameters comparison between the obtained network and the real network in 2011

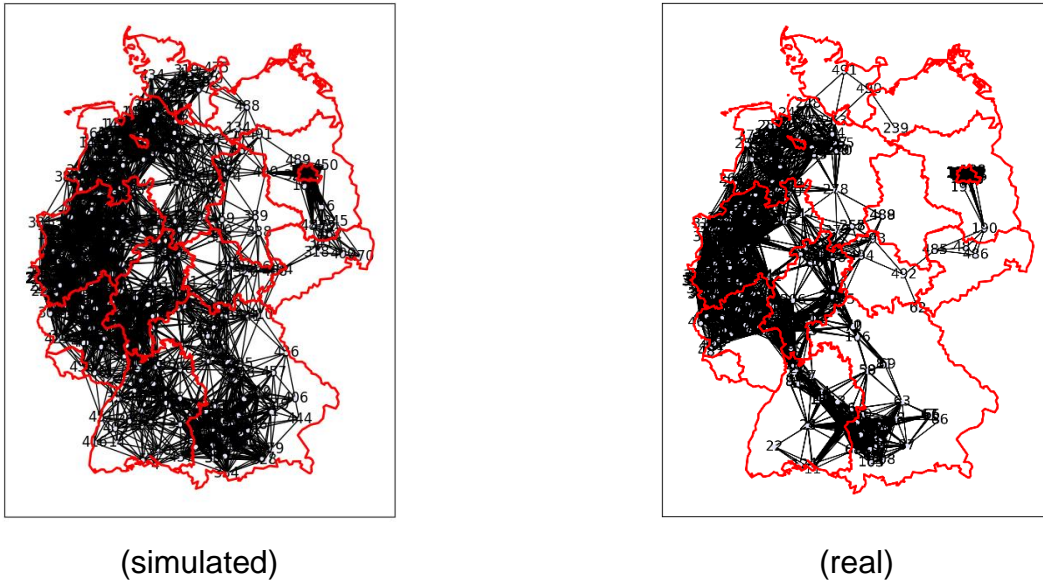


Figure 12: Simulated and real network in 2012

Parameters	Opt	Real
Density	0.14	0.18
Average Distance	50.001	43.65
Diameter	9.0	11
Average Clustering	0.768	0.85
Num. of stations	494	494
Fitness	0.421	0.5303
New installed stations	177	177

Table 4: Parameters comparison between the obtained network and the real network in 2012

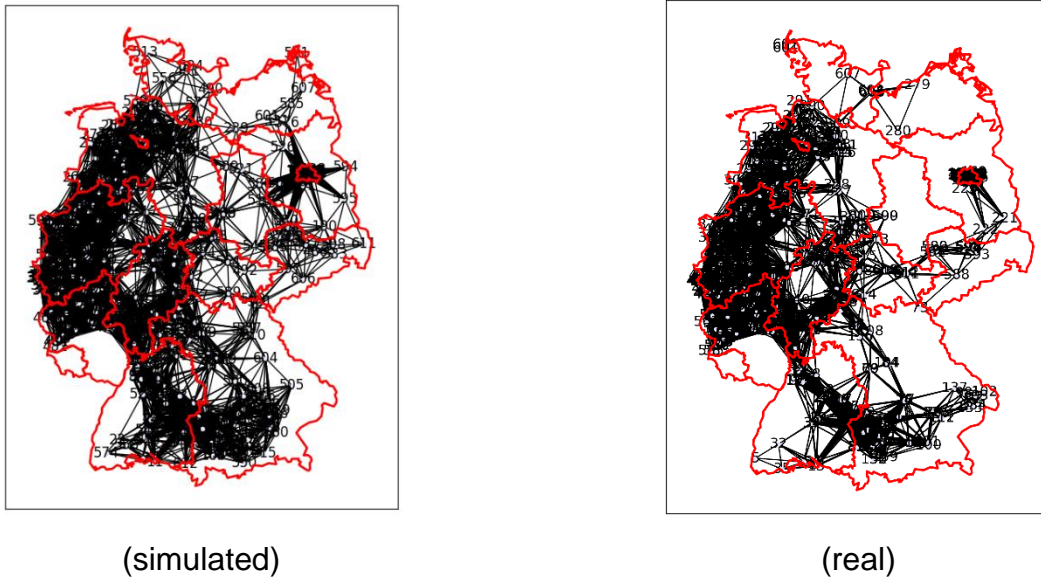


Figure 13: Simulated and real network in 2013

Parameters	Opt	Real
Density	0.144	0.17
Average Distance	48.08	47.52
Diameter	10.0	11.96
Average Clustering	0.797	0.83
Num. of stations	615	615
Fitness	0.4474	0.4861
New installed stations	121	121

Table 5: Parameters comparison between the obtained network and the real network in 2013

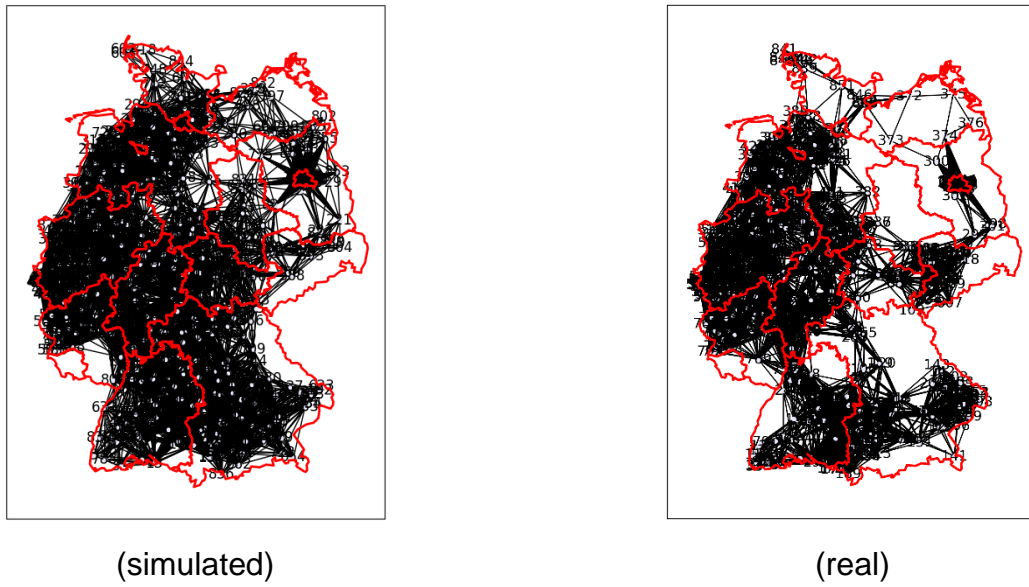


Figure 14: Simulated and real network in 2014

Parameters	Opt	Real
Density	0.126	0.15
Average Distance	53.343	51.92
Diameter	9.0	12
Average Clustering	0.76	0.81
Num. of stations	861	861
Fitness	0.4492	0.4599
New installed stations	246	246

Table 6: Parameters comparison between the obtained network and the real network in 2014

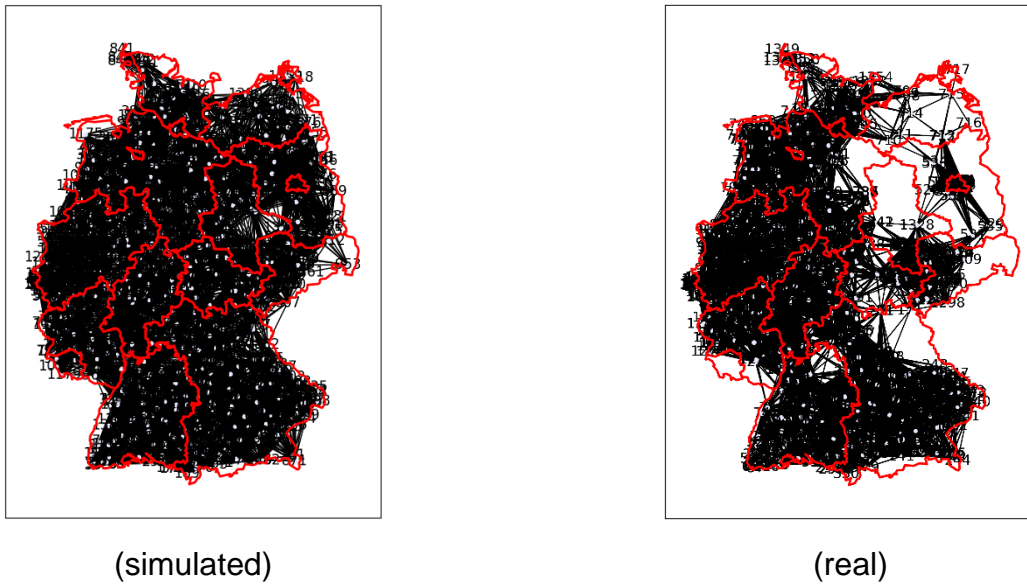


Figure 15: Simulated and real network in 2015

Parameters	Opt	Real
Density	0.1089	0.11
Average Distance	58.406	50.75
Diameter	10.0	11.0
Average Clustering	0.709	0.80
Num. of stations	1393	1393
Fitness	0.3981	0.4459
New installed stations	532	532

Table 7: Parameters comparison between the obtained network and the real network in 2015

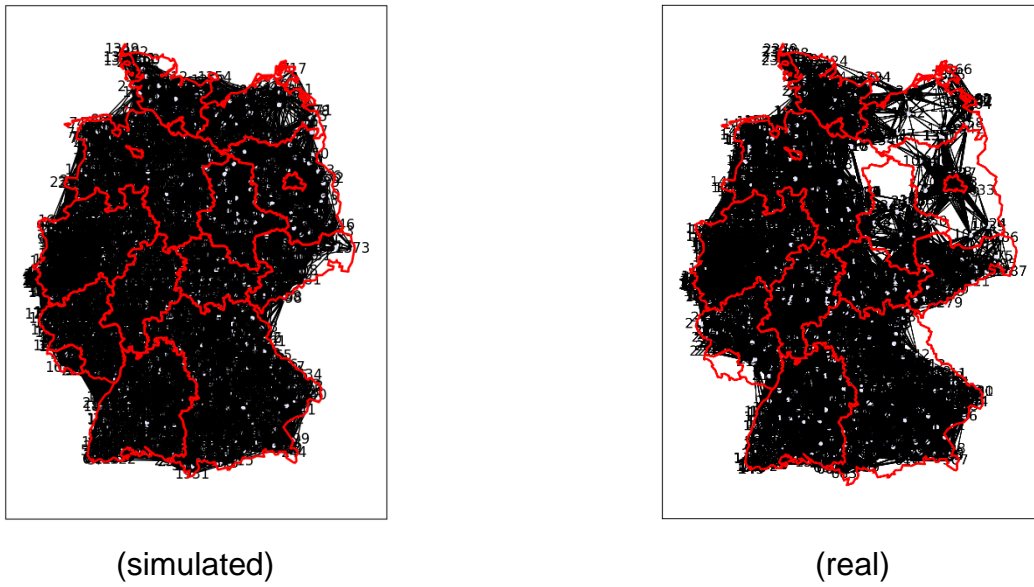


Figure 16: Simulated and real network in 2016

Parameters	Opt	Real
Density	0.094	0.10
Average Distance	59.563	51.70
Diameter	10	11
Average Clustering	0.709	0.78
Num. of stations	2503	2503
Fitness	0.4262	0.4336
New installed stations	1110	1110

Table 8: Parameters comparison between the obtained network and the real network in 2016

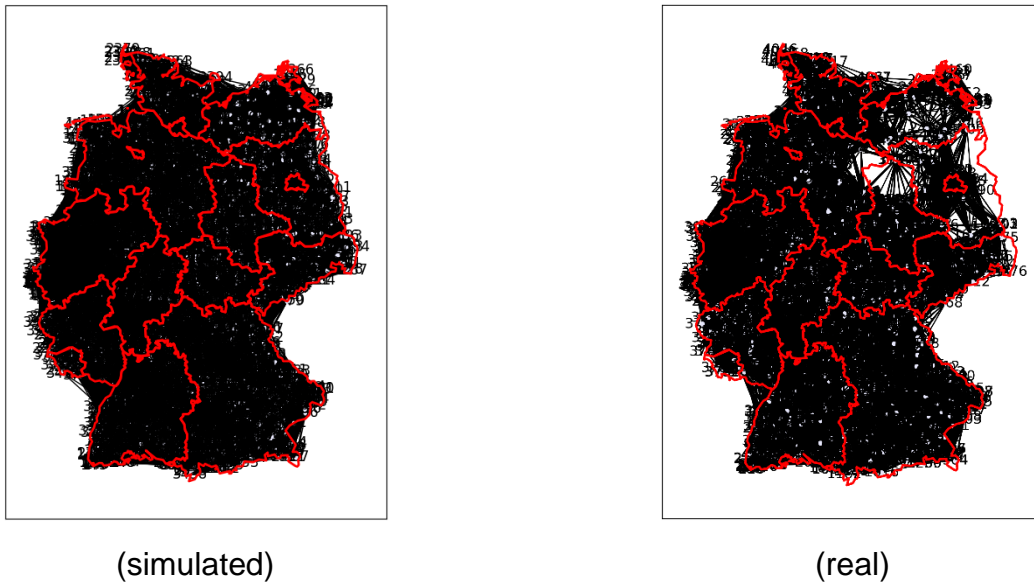


Figure 17: Simulated and real network in 2017

Parameters	Opt	Real
Density	0.089	0.9
Average Distance	60.256	53.26
Diameter	10	10
Average Clustering	0.719	0.77
Num. of stations	4350	4350
Fitness	0.4242	0.4395
New installed stations	1847	1847

Table 9: Parameters comparison between the obtained network and the real network in 2017

6.3 Data Analysis of EV Infrastructure in Germany

In examining the dataset, the overarching trajectory and evolution of Germany's charging network on an annual basis can be discerned. Specifically, there is an exponential growth in the charging infrastructure in Germany, as depicted in Figure 18. Between the years 2009 and 2016, a total of 2539 charging points were commissioned. However, in the subsequent period from 2016 to 2022, there was a tenfold increase, with the number surging to 29733. While this exponential growth is commendable, it also suggests that early adoption between 2009 and 2016 was relatively slow, possibly due to infrastructural or policy-related challenges.

Delving deeper into this growth trajectory, it becomes evident that the proliferation patterns of fast and normal charging stations diverge significantly, as illustrated in Figure 19. Fast charging points exhibit a more linear growth trajectory compared to their normal counterparts. This disparity can largely be attributed to the varying installation costs and distinct electrical grid requisites, which unfortunately serve as both technical and economic impediments to their widespread adoption. The evident divergence in growth trajectories between fast and normal charging stations suggests a potential market imbalance, wherein the uptake of fast charging points may be hindered by these challenges.

Another observation pertains to the power disparities within the charging stations. As represented in Figure 20, there are eight distinct power clusters. The 22-49 kW range emerges as the predominant and most rapidly expanding cluster, an outcome that can be linked to specific regulations and standards implemented over the years. However, the important increase in charging powers exceeding 49 kW have to be discussed. This trend underscores the growing market emphasis on fast charging, especially in high-speed transit corridors. Yet, the varying power clusters also indicate a lack of standardization, which could complicate the user experience.

Lastly, Figure 21 delineates the considerable variability concerning connector types within the charging infrastructure. Such diversity, while reflecting the evolving technological landscape, poses challenges for the proliferation of electric vehicles. Potential users may confront compatibility issues due to the absence of a suitable adapter. This underscores the pressing need for standardization in connector designs to foster seamless interoperability across diverse vehicle models and manufacturers. The lack of a unified connector standard can impede the broader acceptance and adoption of electric vehicles, potentially deterring prospective users.

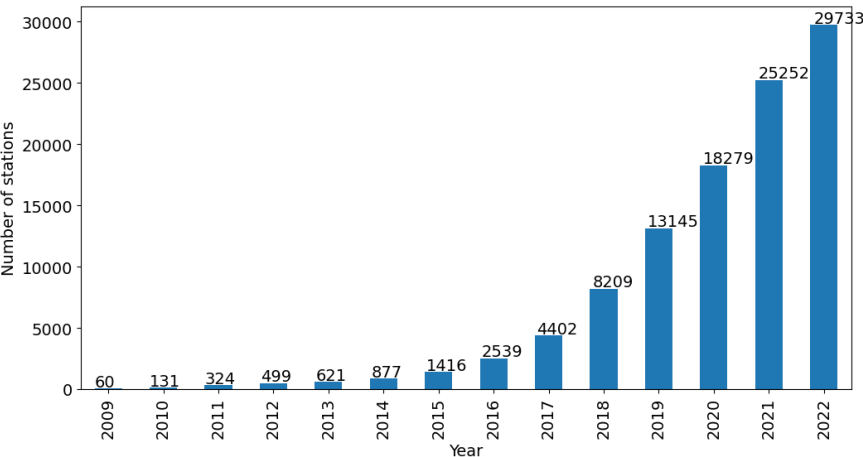


Figure 18: Cumulative histogram per year

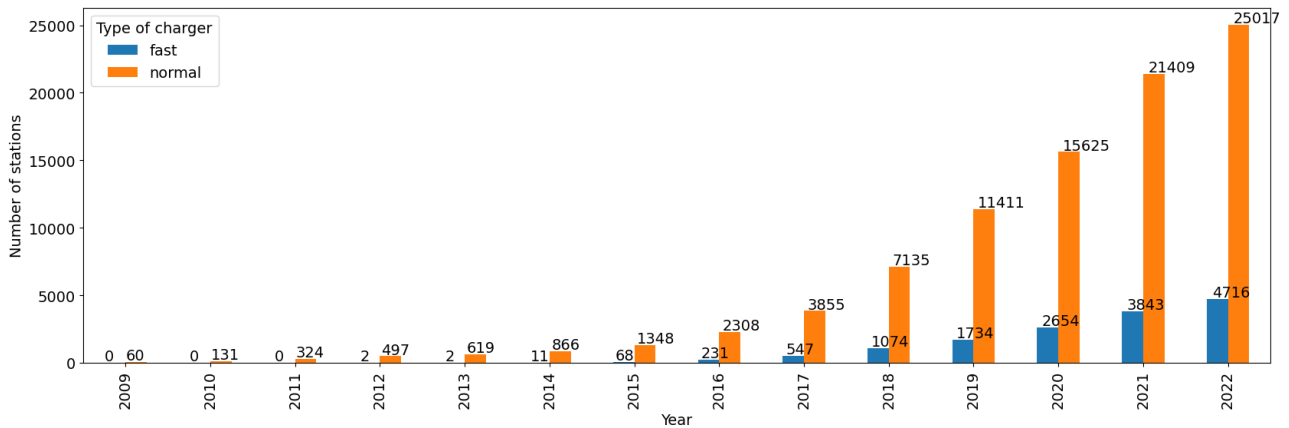


Figure 19: Cumulative histogram per year divided into Normal and Fast charging

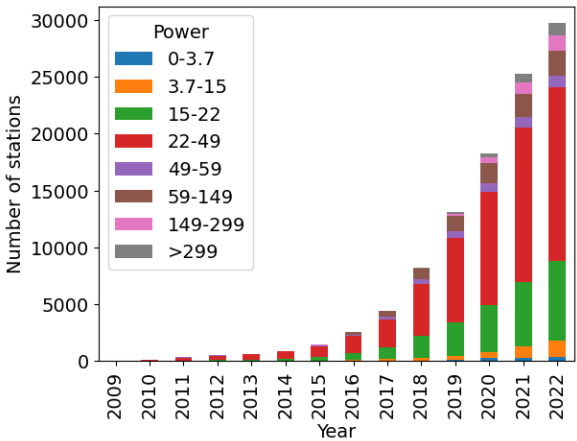


Figure 20: Power clustering histogram per year

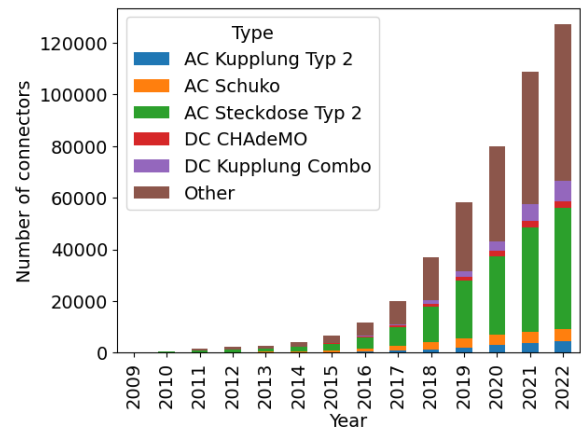


Figure 21: Charging Type clustering per year

Bibliography

- [1] T. Franke and J. F. Krems, "Understanding charging behaviour of electric vehicle users," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 21, pp. 75-89, 2013.
- [2] N. S. Pearre, W. Kempton, R. L. Guensler and V. V. Elango, "Electric vehicles: How much range is required for a day's driving?," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 6, pp. 1171-1184, 2011.
- [3] S. Jordan, D. Newport, S. Sandland and P. Vandergert, "Impact of Public Charging Infrastructure on the Adoption of Electric Vehicles," University of East London, 2020.
- [4] E. Kontou, C. Liu, F. Xie, X. Wu and Z. Lin, "Understanding the Linkage between Electric Vehicle Charging Network Coverage and Charging Opportunity Using GPS Travel Data," 2019.
- [5] B. Haidar and M. T. A. Rojas, "The relationship between public charging infrastructure deployment and other socio-economic factors and electric vehicle adoption in France," *Research in Transportation Economics*, DOI: 10.1016/j.retrec.2022.101208., vol. 95, 2022.
- [6] S. A. Funke, F. Sprei, T. Gnann and P. Plötz, "How much charging infrastructure do electric vehicles need? A review of the evidence and international comparison," *Transportation Research Part D: Transport and Environment*, vol. 77, pp. 224-242, 2019.
- [7] W. Sierzechula, S. Bakker, K. Maat and B. van Wee, "The influence of financial incentives and other socio-economic factors on electric vehicle adoption," *Energy Policy*, vol. 68, pp. 183-194, 2014.
- [8] O. Egbue and S. Long, "Barriers to widespread adoption of electric vehicles: An analysis of consumer attitudes and perceptions," *Energy Policy*, vol. 48, pp. 717-729, 2012.
- [9] Y. Zhang, Y. Yu and B. Zou, "Analyzing public awareness and acceptance of alternative fuel vehicles in China: The case of EV," *Energy Policy*, vol. 39, no. 11, pp. 7015-7024, 2011.

-
- [10] W. Kong, Y. Luo, G. Feng, K. Li and H. Peng, "Optimal location planning method of fast charging station for electric vehicles considering operators, drivers, vehicles, traffic flow and power grid," *Energy*, vol. 186, 2019.
- [11] M. C. Falvo, R. Lamedica, R. Bartoni and G. Maranzano, "Energy management in metro-transit systems: An innovative proposal toward an integrated and sustainable urban mobility system including plug-in electric vehicles," *Electric Power Systems Research*, vol. 81, no. 12, pp. 2127-2138, 2011.
- [12] S. Guo and H. Zhao, "Optimal site selection of electric vehicle charging station by using fuzzy TOPSIS based on sustainability perspective," *Applied Energy*, vol. 158, pp. 390-402, 2015.
- [13] J. Asamer, M. Reinthaler, M. Ruthmair, M. Straub and J. Puchinger, "Optimizing charging station locations for urban taxi providers," *Transportation Research Part A: Policy and Practice*, vol. 85, pp. 233-246, 2016.
- [14] Z. Zhu, Z. Gao, J. Zheng and H. Du, "Charging Station Planning for Plug-In Electric Vehicles," *Journal of Systems Science and Systems Engineering*, vol. 27, pp. 24-45, 2018.
- [15] Z. Sun, W. Gao, B. Li and L. Wang, "Locating charging stations for electric vehicles," *Transport Policy*, vol. 98, pp. 48-54, 2020.
- [16] I. Frade, A. Ribeiro and A. Antunes, "Optimal Location of Charging Stations for Electric Vehicles in a Neighborhood in Lisbon, Portugal," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2252, no. 1, 2011.
- [17] F. He, Y. Yin and J. Zhou, "Deploying public charging stations for electric vehicles on urban road networks," *Transportation Research Part C: Emerging Technologies*, vol. 60, pp. 227-240, 2015.
- [18] N. Shahraki, H. Cai, M. Turkay and M. Xu, "Optimal locations of electric public charging stations using real world vehicle travel patterns," *Transportation Research Part D: Transport and Environment*, vol. 41, pp. 165-176, 2015.
- [19] F. Wu and R. Sioshansi, "A stochastic flow-capturing model to optimize the location of fast-charging stations with uncertain electric vehicle flows," *Transportation Research Part D: Transport and Environment*, vol. 53, pp. 354-376, 2017.

-
- [20] W. Tu, Q. Li, Z. Fang, S. Shaw, B. Zhou and X. Chang, "Optimizing the locations of electric taxi charging stations: A spatial–temporal demand coverage approach," *Transportation Research Part C: Emerging Technologies*, vol. 65, pp. 172-189, 2016.
 - [21] L. Luo, W. Gu, S. Zhou, H. Huang, S. Gao, J. Han, Z. Wu and X. Dou, "Optimal planning of electric vehicle charging stations comprising multi-types of charging facilities," *Applied Energy*, vol. 226, pp. 1087-1099, 2018.
 - [22] K. Liu, L. Kang, H. Ma, J. Zhang and Q. Peng, "Multi-objective optimization of charging patterns for lithium-ion battery management," *Energy Conversion and Management*, vol. 159, pp. 151-162, 2018.
 - [23] H. Mehrjerdi and R. Hemmati, "Electric vehicle charging station with multilevel charging infrastructure and hybrid solar-battery-diesel generation incorporating comfort of drivers," *Journal of Energy Storage*, vol. 26, 2019.
 - [24] Y. He, Z. Song and Z. Liu, "Fast-charging station deployment for battery electric bus systems considering electricity demand charges," *Sustainable Cities and Society*, vol. 48, 2019.
 - [25] S. Davidov and M. Pantos, "Optimization model for charging infrastructure planning with electric power system reliability check," *Energy*, vol. 166, pp. 886-894, 2019.
 - [26] S. Hosseini and M. Sarder, "Development of a Bayesian network model for optimal site selection of electric vehicle charging station," *International Journal of Electrical Power & Energy Systems*, vol. 105, pp. 110-122, 2019.
 - [27] T. Zeng, S. Bae, B. Travacca and S. Moura, "Inducing Human Behavior to Maximize Operation Performance at PEV Charging Station," *IEEE Transactions on Smart Grid*, vol. 12, no. 4, pp. 3353-3363, 2021.
 - [28] M. J. Hodgson, "A Flow Capturing Location Allocation Model," *Geographical Analysis*, vol. 22, no. 3, pp. 270-279, 1990.
 - [29] C. Upchurch and M. Kuby, "A Model for Location of Capacitated Alternative-Fuel Stations," *Geographical Analysis*, vol. 41, no. 1, pp. 85-106, 2009.
 - [30] M. Hosseini and S. A. MirHassan, "A heuristic algorithm for optimal location of flow-refueling capacitated stations," *International Transactions in Operational Research*, vol. 24, pp. 1377-1403, 2017.

-
- [31] B. Yildiz, B. Arslan and O. E. Karaan, "A branch and price approach for routing and refueling station location model," *European Journal of Operational Research*, vol. 248, no. 3, pp. 815-826, 2016.
- [32] S. Li, Y. Huang and S. J. Mason, "A multi-period optimization model for the deployment of public electric vehicle charging stations on network," *Transportation Research Part C*, vol. 65, pp. 128-143, 2016.
- [33] Y. W. Wang and C. C. Lin, "Locating multiple types of recharging stations for battery-powered electric vehicle transport," *Transportation Research Part E*, vol. 58, pp. 76-87, 2013.
- [34] O. Arslan and O. E. Karasan, "A Benders decomposition approach for the charging station location problem with plug-in hybrid electric vehicles," *Transportation Research Part B*, vol. 93, pp. 670-695, 2016.
- [35] M. Ghamami, A. Zockaie and Y. M. Nie, "general corridor model for designing plug-in electric vehicle charging infrastructure," *Transportation Research Part C*, vol. 68, pp. 389-402, 2016.
- [36] S. H. Chung and C. Kwon, "Multi-period planning for electric car charging station locations: a case of Korean expressways," *European Journal of Operational Research*, vol. 242, no. 2, pp. 677-687, 2015.
- [37] A. Zhang, J. Kang and C. Kwon, "Incorporating demand dynamics in multi-period capacitated fast-charging location planning for electric vehicles," *Transportation Research Part B*, vol. 103, pp. 5-29, 2017.
- [38] H. De Vries and H. Duijzer, "Incorporating driving range variability in network design for refueling facilities," *Omega*, vol. 69, pp. 102-114, 2017.
- [39] C. Lee and J. Han, "Benders-and-price approach for electric vehicle charging station location problem under probabilistic travel range," *Transportation Research Part B*, vol. 106, pp. 130-152, 2017.
- [40] M. Hosseini and S. A. MirHassani, "Refueling-station location problem under uncertainty," *Transportation Research Part E*, vol. 84, pp. 101-116, 2015.
- [41] F. Wu and R. Sioshansi, "A stochastic flow-capturing model to optimize the location of fast-charging stations with uncertain," *Transportation Research Part D*, vol. 53, pp.

354-376, 2017.

- [42] M. Miralinaghi, Y. Lou, B. B. Keskin and A. Zarrinmehr, "Refueling station location problem with traffic deviation considering route choice and demand uncertainty," *International Journal of Hydrogen Energy*, vol. 42, no. 5, pp. 3335-3351, 2017.
- [43] P. Sadeghi-Barzani, A. Rajabi-Ghahnavieh and H. Kazemi-Karegar, "Optimal fast charging station placing and sizing," *Applied Energy*, vol. 125, pp. 289-299, 2014.
- [44] O. Arslan and O. Karasan, "A Benders decomposition approach for the charging station location problem with plug-in hybrid electric vehicles," *Transportation Research Part B: Methodological*, vol. 93, no. A, pp. 670-695, 2016.
- [45] K. Zhang, L. Xu, M. Ouyang, H. Wang, L. Lu, J. Li and Z. Li, "Optimal decentralized valley-filling charging strategy for electric vehicles," *Energy Conversion and Management*, vol. 78, pp. 537-550, 2014.
- [46] Z. Zhu, Z. Gao, J. Zheng and H. Du, "Charging station location problem of plug-in electric vehicles," *Journal of Transport Geography*, vol. 52, pp. 11-22, 2016.
- [47] M. Akbari, M. Brenna and M. Longo, "Optimal Locating of Electric Vehicle Charging Stations by Application of Genetic Algorithm," *Sustainability*, vol. 10, no. 4, 2018.
- [48] A. Awasthi, K. Venkitesamy, S. Padmanaban, R. Selvamuthukumaran, F. Blaabjerg and A. K. Singh, "Optimal planning of electric vehicle charging station at the distribution system using hybrid optimization algorithm," *Energy*, vol. 113, pp. 70-78, 2017.
- [49] Y. Chen, C. Cheng, S. Li and C. Yu, "Location optimization for multiple types of charging stations for electric scooters," *Applied Soft Computing*, vol. 67, pp. 519-528, 2018.
- [50] Y. Zhang, Q. Zhang, A. Farnoosh, S. Chen and Y. Li, "GIS-Based Multi-Objective Particle Swarm Optimization of charging stations for electric vehicles," *Energy*, vol. 169, pp. 844-853, 2019.
- [51] M. Straka and L. Buzna, "Clustering algorithms applied to usage related segments of electric vehicle charging stations," *Transportation Research Procedia*, vol. 40, pp. 1576-1582, 2019.

-
- [52] Y. Wu, A. Ravey, D. Chrenko and A. Miraoui, "Demand side energy management of EV charging stations by approximate dynamic programming," *Energy Conversion and Management*, vol. 196, pp. 878-890, 2019.
- [53] R. Diestel, *Graph Theory*, Springer, 2010.
- [54] T. H. Cormen, C. E. Leiserson, R. L. Rivest and C. Stein, *Introduction to algorithms*, MIT Press, 2009.
- [55] M. Newman, *Networks: An Introduction*, Oxford University Press, 2010.
- [56] W. L. Hamilton, *Graph Representation Learning*, Springer Link, 2020.
- [57] O. Sporns, "Graph theory methods: applications in brain networks," *Dialogues Clin Neurosci*, pp. 111-121, 2018.
- [58] A. S. d. Mata, "Complex Networks: a Mini-review," *Brazilian Journal of Physics*, vol. 50, p. 658–67, 2020.
- [59] H. Holland, *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*, MIT Press Direct, 1992.
- [60] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley Longman Publishing Co., Inc., 1989.
- [61] K. De Jong, "An Analysis of the Behavior of a Class of Genetic Adaptive Systems," Department of Computer and Communication Sciences, University of Michigan, 1975.
- [62] M. Mitchell, *An Introduction to Genetic Algorithms*, MIT Press, 1998.
- [63] A. E. Eiben and J. E. Smith, *Introduction to Evolutionary Computing*, Springer, 2003.
- [64] K. Deb, *Multiobjective Optimization Using Evolutionary Algorithms*, Jhon Wiley and Sons Ltd, 2001.
- [65] B. L. Miller and D. E. Goldberg, "Genetic Algorithms, Tournament Selection and the Effects of Noise," *Complex Symstems*, vol. 9, pp. 193-212, 1995.
- [66] J. E. Baker, "Adaptive Selection Methods for Genetic Algorithms," in *Proceedings of the 1st International Conference on Genetic Algorithms*, 1985.
- [67] A. H. Wright, "Genetic Algorithms for Real Parameter Optimization," *Foundations of*

Genetic Algorithms, vol. 1, pp. 205-218, 1991.

- [68] K. Deb and R. B. Agrawal, "Simulated Binary Crossover for Continuous Search Space," *Complex Systems*, vol. 9, no. 2, pp. 115-148, 1995.
- [69] Z. Michalewicz, *Genetic Algorithms + Data Structures = Evolution Programs*, Springer, 1996.
- [70] A. E. Smith and D. W. Coit, *Constraint-Handling Techniques - Penalty Functions*, Handbook of Evolutionary Computation, Institute of Physics Publishing and Oxford, 1997.
- [71] J. C. Bean, "Genetic Algorithms and Random Keys for Sequencing and Optimization," *ORSA Journal on Computing*, vol. 6, no. 2, 1994.
- [72] X. Yao, Y. Liu and G. Lin, "Evolutionary programming made faster," *IEEE Transactions on Evolutionary Computation*, DOI: 10.1109/4235.771163, vol. 3, no. 2, pp. 82-102, 1999.
- [73] D. E. Goldberg and S. Voessner, "Optimizing global-local search hybrids," in *GSECCO'99*, 1999.
- [74] E. Cantú-Paz, *Efficient and Accurate Parallel Genetic Algorithms*, Springer, 2001.
- [75] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, pp. 5-32, 2001.
- [76] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, pp. 123-140, 1996.
- [77] D. J. Stekhoven and P. Bühlmann, "MissForest—non-parametric missing value imputation for mixed-type data," *Bioinformatics*, vol. 28, no. 1, pp. 112-118, 2012.
- [78] E. Fix and J. L. Hodges, "Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties," *USAF School of Aviation Medicine*, 1951.
- [79] K. Q. Weinberger and L. K. Saul, "Distance Metric Learning for Large Margin Nearest Neighbor Classification," *Journal of Machine Learning Research*, vol. 10, pp. 207-244, 2009.
- [80] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *International Journal of Forecasting*, vol. 22, no. 4, pp. 679-688, 2006.
- [81] G. Filomeno, "Automation of Design Synthesis for Electric Vehicle Transmissions,"

Ruhr-Universität Bochum, Ph.D. Thesis, Bochum, 2023.

- [82] R. Chen, X. Qian, L. Miao and S. V. Ukkusuri, "Optimal charging facility location and capacity for electric vehicles considering route choice and charging time equilibrium," *Computer & Operations Research*, vol. 113, 2020.
- [83] M. Kuby and S. Lim, "The flow-refueling location problem for alternative-fuel vehicles," *Socio-Economic Planning Sciences*, vol. 39, no. 2, pp. 125-145, 2005.

Insurance

I certify that I have written the above thesis independently and have not used any outside help. All passages taken verbatim or in spirit from published or unpublished literature have been marked as such.

Giovanni Filomeno