Transmission Network Expansion Planning using the Railway Network

Master Thesis of

Chao Wang

At the Department of Informatics
Institute of Theoretical Computer Science

Reviewers:	Prof. Dr. Dorothea Wagner
	Prof. Dr. Peter Sanders
Advisors:	Franziska Wegner
	Matthias Wolf

Time Period: 15th November 2017 – 14th May 2018

Statement of Authorship

I hereby declare that this document has been composed by myself and describes my own work, unless otherwise acknowledged in the text.

Karlsruhe, 14th May 2018

Abstract

The main goal of Transmission Network Expansion Planning (TNEP) is to determine optimal investments on transmission line additions from the candidate network to satisfy reliability criteria under future load and generation scenarios. Up to now, most of the current researches focus on using different methods to find optimal solution of TNEP problem, however, how to generate the candidate network, where the new lines can be installed, has also drawn our attentions. Hence, this thesis explores a new design problem about using railway network to expand the power grid, which is defined as follows. Given the current power grid and the railway network, we build connections between both network and find, where and how many new lines along new connections and railway edges should be added with an object of minimizing the cost of lines construction under the electrical constraints. In order to build connections between the power grid and railway network, the algorithm KD-tree is proposed in our work, which can efficiently find k-nearest neighbors in railway network for each node in the power grid. For finding an optimal solution of TNEP problem, we have adopted mathematical models, such as mixed-integer non-linear model, binary linear model and mixed-integer linear model, which are been discussed in the most researches. Additionally, we have also proposed a heuristic algorithm and compared with the mathematical model. The heuristic algorithm performs more computational efficiency than the mathematical model.

In the end, a simple case study of 8-buses system is used to illustrate the non-linear and linear mathematical model. Furthermore, the realistic scenarios of German including the power grid and the railway network is implemented and tested by using the linear mathematical model and the heuristic algorithm. The non-linear mathematic model comparing to the linear model is easy to understand, but has lower computational efficiency and less optimal result. On the other hand, the heuristic model shows the highest efficiency but the result is way below the ones from linear mathematic model.

Deutsche Zusammenfassung

Das Hauptziel von TNEP (Transmission Network Expansion Planning) ist die Ermittlung optimaler Investitionen in Übertragungsleitungen aus dem Kandidatennetzwerk, um die Zuverlässigkeitskriterien unter zukünftigen Last- und Erzeugungsszenarien zu erfüllen. Bis jetzt konzentrieren sich die meisten aktuellen Forschungen auf die Verwendung verschiedener Methoden, um eine optimale Lösung des TNEP-Problems zu finden, jedoch hat es auch unsere Aufmerksamkeit auf sich gezogen, wie das Kandidatennetzwerk zu erzeugen ist, wo die neuen Linien installiert werden können. Die vorliegende Arbeit untersucht daher ein neues Designproblem zur Nutzung des deutschen Schienennetzes zum Aufbau deutschen Stromnetzes. Es ist definiert wie folgend. Angesichts des derzeitigen Stromnetzes und des Schienennetzes suchen wir mögliche Verbindungen zwischen beiden Netzen auf und entscheiden, wo und wie viele neue Leitungen entlang neuer Verbindungen und an der Kante hinzugefügt werden sollten, um die Kosten für den Bau von Leitungen unter den elektrischen Randbedingungen zu minimieren. Um Verbindungen zwischen dem Stromnetz und dem Eisenbahnnetz aufzubauen, wird in unserer Arbeit der Algorithmus KD-Baum verwendet, die nächste k Nachbarn im Eisenbahnnetz von jedem Knoten im Stromnetz mit hohe Effizient finden kann. Um eine optimale Lösung des TNEP-Problems zu finden, haben wir mathematische Modelle, wie gemischt-ganzzahliges nicht-lineares

Modell, binäres lineares Modell und gemischt-ganzzahliges lineares Modell, die in den meisten Untersuchungen diskutiert wurden, übernommen. Zusätzlich haben wir einen heuristischen Algorithmus vorgeschlagen und mit dem mathematischen Modell verglichen. Der heuristische Algorithmus führt eine höhere Recheneffizienz als das mathematische Modell aus.

Am Ende wird eine einfache Fallstudie mit 8-Busse Systems zur Veranschaulichung des nichtlinearen und linearen mathematischen Modells verwendet. Darüber hinaus werden die realistischen Szenarien von Deutsch einschließlich des Stromnetzes und des Eisenbahnnetzes implementiert und getestet, indem das lineare mathematische Modell und der heuristische Algorithmus verwendet werden.

Das nichtlineare mathematische Modell im Vergleich zum linearen Modell ist leicht zu verstehen, hat jedoch eine geringere Recheneffizienz und ein weniger optimales Ergebnis. Auf der anderen Seite zeigt das heuristische Modell die höchste Effizienz, aber das Ergebnis liegt weit unter denen aus dem linearen mathematischen Modell.

Contents

List of Figures

List of Tables

The next list describes several symbols that will be later used within the body of the document

List of Abbreviations

AC   alternating current

DBTP deutsche Bahn train paths

DC   direct current

LP   linear programming

MILP mixed-integer linear programming

MINLP mixed-integer non-linear programming

NLP   non-linear programming

List of Symbols

α weighting factor to make comparable investment and operation costs

b_{vw} the susceptance of the line along edge (v, w)

b_{vw}^k the susceptance of the kth parallel line along edge (v, w)

C_{vw} the cost of a line which is added from v to w

cap_{vw} the corresponding maximum power flow from vertex v to vertex w

cap_{vw}^k the corresponding maximum power flow of the kth parallel line along edge (v, w)

f_{vw} the power flow through the line from v to w

f_{vw}^k the power flow through the kth parallel line from v to w

g_v power generated at vertex v

$\bar{g_v}$ maximum Power generated at vertex v

$G = (V, E)$ the combined network of the power grid, railway network and the connection network, $(V_1 \cup V_2) = V$, $(E_1 \cup E_2 \cup E_3) \in E$

$G_1 = (V_1, E_1)$ the power graph

$G_2 = (V_2, E_2)$ the railway graph

$G_3 = (V, E_3)$ the connection graph between power grid and railway network

h the total investment

l_v power consumption at vertex v

lat the latitude

lon the longitude

n_{vw} integer variable, how many lines are installed along edge (v, w)

n_{vw}^k binary variable, whether the kth parallel line along edge (v, w) is built

n_{vw}^- the upper bound of the lines, which are installed along edge (v, w)

$N = (G, b, cap, l, \bar{g})$ the electrical network

P_v cost of the energy produced in bus v by the generating per MW

v arbitrary vertex

w arbitrary vertex

Z_{vw} decision variable, whether circuit is installed along edge (v, w)

ε the increased factor of the original power consumption

1. Introduction

In recent years due to the high power demand growth, optimal network expansion has become one of the hottest topics. Many studies have applied a famous optimization problem to expand the existing power system to meet the future energy delivery requirements, which is called Transmission Network Expansion Planning (TNEP). The TNEP is, how to minimize the investment cost while satisfying the constraints, such as the limits of the generator output power, the security limits and so on. TNEP faces many challenges of abstracting from the physical constraints, which is discussed in [HHK13a]. However, most TNEP studies consider only perfect the objective and constraints to improve the system's reliability and maximize the benefit. They approximate the transmission system on the existed power grid or abstract extended lines between the distributed generation or load buses. For these cases the environmental impact of adding new lines in power system is difficult to avoid and estimate. In this paper, we adopt the railway network when new lines are installed for expand the power grid. Because when the railway network were planned, the impact of disturbing the residents has been taken into consideration. Obviously, installing new lines in power system, which have the same paths as the railway network, makes more sense and reasonable. In other words the application of the railway networks in area power grid can be a attractive and similar topic on National economic infrastructure, which has the similar function as they both both spread all around the country and make the national power and economy healthy works, respectively.

In this thesis, the major work is adopt the widely-used DC power flow model to design the transmission system, which contains German railway network, German power network and connections between both networks, with mixed-integer linear programing(MILP). Our work starts with a mix-integer non-linear programming (MINLP) on the properly handled power system. Since the MINLP cannot provide an appealing tradeoff between solution quality and computational efficiency, linearizing the non-linear model to a linear model is another important part of this work, which can also support to build multiple lines to transfer power along railway tracks. We also develop the heuristic algorithm for installing new lines in the candidate networks to compare with the mathematical programming. In the end, test cases of 8-buses system and the German realistic networks are involved to evaluate the computational experiment.

In the first part of this work, the concepts of the involved networks are introduced. Besides concepts, symbols and definitions through all the work could also been found there. In chapter 4 we discuss how to extract the German railway network and German power grid from OpenStreetMap and build connections between both networks, where the designed

mathematical model (see in chapter 5) will be implemented and evaluated. In chapter 5 a mix-integer non-linear model is proposed in the DC system, furthermore, this MINLP model is linearized to a binary linear model, which cannot support to install multiple lines along one candidate path. In this way, the imperfection of this linear model is fixed in the end of this chapter by involving the multiple parallel line model. In chapter 6 the heuristic algorithm is proposed to selecting candidate edge to expand the power grid .In chapter 7 of 8-buses system and the German realistic networks are studied.

2. Preliminaries

In this chapter the basic concepts and definitions about our work are given. In Section 2.1 the basic mathematical formulas and definitions are provided, additionally the common scientific knowledge, which is adopted in our work, is described in details. In Section 2.2 we will introduce a number of basic graph theory terms and concepts. We talk about the fundamental knowledge of electrical definitions and properties in Section 2.3, then they are used to describe TNEP. Section 2.4 gives the common definitions of mathematical optimisation programming. In the end of this chapter we present the tools, that are used to analyze and devise solutions of our models.

2.1 Basic Concepts

In this section we give the definition and calculation formulas of the geographical distance between two points and the nearest neighbor problem in geographic coordinate system.

2.1.1 Geographical Distance

geographical distance is the distance measured along the surface of the earth. Given two vertices $v(lat_1, lon_1)$ and $w(lat_2, lon_2)$ in geographic coordinate system, we use the following specific formula to calculate the great-circle distance D_{vw} between the two points v and w, which is the shortest distance over the earth's surface. R is the radius of the earth, which is the approximate distance from earth's center to its surface, normally we set the it by $6373km$. The computation of the geographical distance D_{vw} is expressed as follow:

$$D_{vw} = R \cdot c \tag{2.1}$$

$$d_{lon} = lon_2 - lon_1 \tag{2.2}$$

$$d_{lat} = lat_2 - lat_1 \tag{2.3}$$

$$a = (sin(\frac{d_{lat}}{2}))^2 + cos(lat_1) \cdot cos(lat_2) \cdot (sin(\frac{d_{lon}}{2}))^2 \tag{2.4}$$

$$c = 2 \cdot atan2(\sqrt{a}, \sqrt{1-a}) \tag{2.5}$$

$$\tag{2.6}$$

2.1.2 The nearest Neighbour

Through this work, we will concern about the nearest neighbour problem in geographic coordinate system, in order to build the necessary data structure, which is described in Chapter 4. The Nearest Neighbor is: Given a collection of search points V and a query point w in geographic coordinate system, find the search point v with $v \in V$ that is closest to the query point, where $D_{vw} \leq D_{yw}, \forall y \in V$.

2.2 Graph Theory

A graph G can be defined as a pair (V, E) [Bal12], where V is a set of vertices, and E is a set of edges between the vertices $E = \{(u, v) | u, v \in V\}$. If vertex u has an adjacency relation with vertex v by one edge $e = (u, v)$, where $u, v \in V$ and $(u, v) \in E$, then vertex v is a neighbour of vertex u. The degree of a vertex u in a graph is the number of edges incident to the vertex u, we use $deg(u)$ to denote the degree of a vertex v. If graph G is directed, E is a set of ordinary pairs of vertices, called directed edges. For each directed edge $\vec{e} = (u, v)$ in graph G, the vertex u is called source of edge e, the vertex v is called target of edge e. In the case of the directed graph, for vertex u we distinguish between in-degree $deg^-(u)$, which is the number of head ends adjacent to vertex u, and out-degree $deg^+(u)$, which is the number of tail ends adjacent to vertex u. Either in undirected graphs or directed graphs, the number of vertices are denoted by $|V|$ and the number of edges are denoted by $|E|$. Note that, if we do not mentioned a graph is directed, then the graph must be undirected.
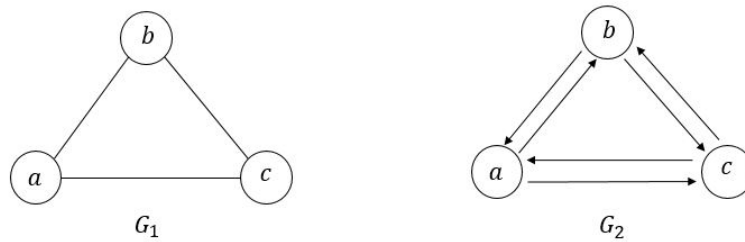
Figure 2.1: Example of a undirected graph G_1 and a directed graph G_2, for the undirected graph $G_1 = (V_1, E_1)$, where $V_1 = \{a, b, c\}$ and $E_1 = \{(a, b), (b, c), (a, c)\}$, $deg(a)$, $deg(b)$ and $deg(c)$ are 2, $|V_1| = 3$, $|E_1| = 3$. For the directed graph $G_2 = (V_2, E_2)$, where $V_2 = \{a, b, c\}$ and $E_2 = \{(a, b), (b, a), (b, c), (c, b), (a, c), (c, a)\}$, $deg^+(a)$, $deg^-(a)$, $deg^+(b)$, $deg^-(b)$, $deg^+(c)$ and $deg^-(c)$ are 2, $|V_2| = 3$, $|E_2| = 6$

In an undirected graph $G = (V, E)$, two vertices u and v are called *connected* if G contains a path from u to v, where $u, v \in V$. Otherwise, they are called *disconnected*. If there is a path between every pair of vertices in G, then the graph G is *connected*. In other words, in a connected graph, the vertices must be reachable. A connected component of an undirected graph is a maximal set of nodes such that each pair of nodes is connected by a path. Connected components form a partition of the set of graph vertices, meaning that connected components are non-empty, they are pairwise disjoints, and the union of connected components forms the set of all vertices.
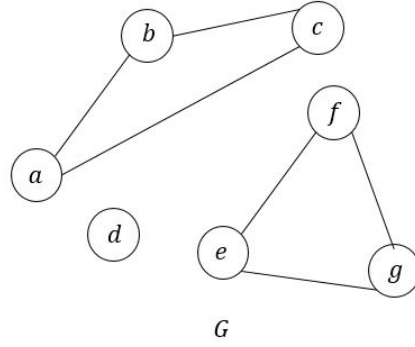
Figure 2.2: Example of a disconnected graph G, which consist of three connected components, $G_1 = (V_1, E_1)$, $G_2 = (V_2, E_2)$ and $G_3 = (V_3, E_3)$, where $V_1 = \{a, b, c\}$, $E_1 = \{(a, b), (b, c), (a, c)\}$, $V_2 = \{d\}$, $E_2 = \varnothing$, $V_3 = \{e, f, g\}$, $E_3 = \{(e, f), (f, g), (e, g)\}$.

Lemma 2.1 (Connected Graph). *A graph $G = (V, E)$ with $|V|$ vertices and $|E|$ edges, the number N_G of connected components at least must be :*

$$|V| - |E|$$

Proof. *By induction on E, if $|E| = 0$, then each vertex $v \in V$ is a connected component, so the corollary holds. If $|E| > 0$, arbitrarily pick an edge$(u, v) \in G$ and G' is the graph obtained by removing (u, v). Then G' has at most one component more than G. By induction hypothesis, G' has at least $v - (e - 1)$ connected components, so G has at least $|V| - (|E| - 1) - 1 = |V| - |E|$ connected components.*

Corollary 2.2 (Connected Graph). *A connected graph $G = (V, E)$ with $|V|$ vertices has at least $|V| - 1$ edges.*

Proof. *According to lemma 2.1. If graph G has less than $|V| - 1$ edges, then $N_G > |V| - (|V| - 1) = 1$. In other words, there must be more than 1 connected components in graph G.*

2.3 Electrical Network

Our work are formulated and analysed on the electrical network, which is often used to model real-world electrical transmission systems. The electrical network is an interconnected network connected by lines to carry power between nodes (generators and loads) to meet the demands and supplies of power.

Definition 2.3 (Electrical Network). *An electrical network $N = (G, b, cap, l, \bar{g})$ is an undirected graph $G = (V, E)$, susceptance function $b : E \to \mathbb{R}_0^+$, capacity function $cap : E \to \mathbb{R}_0^+$, load function $l : V \to \mathbb{R}_0^+$ and an upper bound of the generated power $\bar{g} : V \to \mathbb{R}_0^+$.*

The function *cap* represents the thermal line limits in electrical network, the function b represents the suceptance of each line, and here the symbol s is adopted, where $s_v = g_v - l_v$ potentially, where $v \in V$, g and l are function to represent the generated power and loaded power, respectively. If $s_v > 0$, then node v is a generator bus, if $s_v < 0$, v is a load bus, and we call node v transmission bus, when $s_v = 0$. Normally when we build model on the electrical network, it holds always with $\sum_v s_v = 0$, where $v \in V$. Through our work we write the function in short as the following rule: cap_{vw} for $cap(vw)$, l_v for $l(v)$, and $\bar{g}_v$ for $\bar{g}(v)$.

Before we model the transmission planning on the electrical network $N = ((V, E), b, cap, l, \bar{g})$, graph $G = (V, E)$ should be undirected. However, during modelling we will assign the value of power flow f for each edge $e \in E$ with flow function $f : E \to \mathbb{R}^+$, which need to fixed the orientation of its edge E. The directed edge $\vec{e} = (u, v)$, where $\vec{e} \in \vec{E}$ represents power flow carried from vertex u to v. We use $h(\vec{e})$ and $t(\vec{e})$ represent the head vertex v and tail vertex u of the directed edge $\vec{e}$, respectively.

Definition 2.4 (Kirchhoff's Current Laws). *A modeled electrical network $N = ((V, E), b, cap , l_1, \bar{g}_1)$ with a function $f : E \to \mathbb{R}$ must always hold power balance in each vertex $v \in V$, in short, we write $f_v w$ for $f(vw)$:*

$$s_v + \sum_{(u,v) \in \vec{E}} f_{uv} - \sum_{(v,w) \in \vec{E}} f_{vw} = 0 \tag{2.7}$$

Definition 2.5 (Kirchhoff's Voltage Law). *The Kirchhoff's voltage law is defined in an electrical network $N = ((V, E), b, cap, l_1, \bar{g}_1)$ with potential $\theta : V \to \mathbb{R}$ and with a function $f : E \to \mathbb{R}$, in short, we write θ_v for $\theta(v)$ and $f_v w$ for $f(vw)$:*

$$f_{vw} - b_{vw}(\theta_v - \theta_w) = 0 \tag{2.8}$$

Definition 2.6 (Power Flow Function f). *Given a electrical network $N = ((V, E), b, cap, l_1, \bar{g}_1)$, if there a function $f : E \to \mathbb{R}$, which can satisfy the Kirchhoff's current laws and the Kirchhoff's voltage law simultaneous. We call this function power flow function.*

Note that in the flow function $f : E \to \mathbb{R}^+$ we need to fixed the orientation of its edge E. The directed edge $\vec{e} = (u, v)$, where $\vec{e} \in \vec{E}$ represents power flow carried from vertex u to v. We use $h(\vec{e})$ and $t(\vec{e})$ represent the head vertex v and tail vertex u of the directed edge $\vec{e}$, respectively.
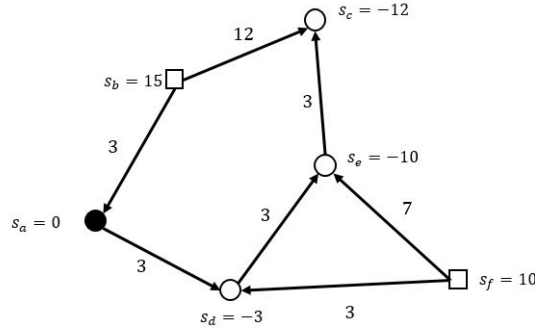
Figure 2.3: Example of a electrical network, boxes denote generator buses, circles denote load buses and filled dots denote transmission buses. The weights of edges represent the power flow, and the arrows display the direction of power flow.

2.3.1 Calculation of Susceptance

For a given electrical network $N = ((V, E), b, cap, l, \bar{g})$, the set of suscpetance b is not directly accessed, which is needed to be computed with the help of the available dataset of resistance r and reactance x. The follow formulation [EH72] is used through all the project to calculate the susceptance.

$$Y = \frac{1}{Z} = \frac{1}{r + jx} = \frac{r}{r^2 + x^2} + j\frac{-x}{r^2 + x^2} \tag{2.9}$$

$$b = Im(Y) = \frac{-x}{r^2 + x^2} \tag{2.10}$$

Y is the admittance, measured in siemens. Z is the admittance, measured in siemens. j is the imaginary unit. b is the susceptance, measured in siemens. r is the resistance, measured in ohms. x is the reactance, measured in ohms.

2.3.2 Per-Unit-System

When we analyze the electrical network in electrical engineering, in order to identify incorrect data more easily, the per-unit-system [EH72] can be adopted by absorbing large differences in absolute values into a relatively narrow range. In the per-unit-system voltage, current, power, and impedance values are normalized with a selected base, as shown in the computations for all related quantities below (measured by unit pu). At the beginning, we need assume two independent base: power base and voltage base:

$$P_{base} = 1pu \tag{2.11}$$

$$V_{base} = 1pu \tag{2.12}$$

Then, the I_{base}, Z_{base} and Y_{base} can be derived from power and voltage base.

$$I_{base} = \frac{P_{base}}{V_{base}} = 1pu \tag{2.13}$$

$$Z_{base} = \frac{V_{base}}{I_{base}} = \frac{V_{base}^2}{I_{base}V_{base}} = \frac{V_{base}^2}{P_{base}} = 1pu \tag{2.14}$$

$$Y_{base} = \frac{1}{Z_{base}} = 1pu \tag{2.15}$$

For example, given an electrical network $N = ((V, E), b, cap, s)$, there is one line (v, w) operating at $100kV$ with a nominal base rating of $200kV$. Likewise, the capacity cap_{vw} in the network is $80MW$ with a nominal base rating of $100MW$, where $(v, w) \in E$. So the line (v, w) has a per-unit voltage quantity of $0.5pu$ and a per-unit power quantity of $0.8pu$.

2.3.3 Cost of Electricity by Source

In our model we consider the operating costs of the generating units as one part of the objective function. The motivation is that we want to minimize the investment cost of circuits and the operating costs of the generating units simultaneously. The distinct ways of generation bus incur different costs, especially with the technology development and environment impact of different energy source can lead to the operating cost changed frequently in the future. In this case this can be an attractive topic of TNEP. Because of the difficulty of environment impact evaluating, we consider only the energy source price, when we simulate the operating costs of the generating units. There are several ways to calculate the cost, here we derive the levelized cost of electricity (LCOE) [ene05], which is an economic assessment of the average total cost to build and operate a power-generating asset over its lifetime divided by the total energy output of the asset over that lifetime. The factors are considered in the calculation are capital costs, fuel cost and other margin cost. The cost is given per kilowatt-hour.

The levelized cost of electricity (LCOE) is given by:

$$LCOE = \frac{\text{sum of costs over lifetime}}{\text{sum of electrical energy produced over lifetime}} \tag{2.16}$$

$$= \frac{\sum_{t=1}^{n} (I_t + M_t + F_t)/(1+r)^t}{\sum_{t=1}^{n} (E_t)/(1+r)^t} \tag{2.17}$$

I_t: investment expenditures in the year t.

M_t: operations and maintenance expenditures in the year t.

F_t: fuel expenditures in the year t.

E_t: electrical energy generated in the year t.

r: discount rate t.

n: expected lifetime of system or power station.

In our work we focus on the cost of electricity in Germany in the year 2018 [sou18]. The following Table 2.1 shows the price in eurocent per kilowatt, which is computed by above formula 2.16.

Table 2.1: Cost of electricity by source

source	price ($Eurocent/kW$)
Brown Coal	8.4
Gas	16.8
Geothermal	8.4
Hard Coal	11.6
Multiple	18
Nuclear	8.4
Oil	8.4
Other	8.4
Run of River	10.8
Solar	18
Storage Hydro	8.4
Waste	21.5
Wind Offshore	16.9
Wind Onshore	11.4

2.4 Mathematical Optimization Program

The mathematical optimization program is the selection from the candidate solution set, in order to find the best solution with regard to some constraints. In other words, it can be expressed a systematic procedure of choosing input values from an allowed set and computing the values to maximizing or minimizing a real function. The standard form of a optimization problem is expressed as follow [Bie09]:

$$\begin{aligned}
\textbf{minimize} \quad & z = f(x) \\
\textbf{subject to} : & \\
& g_i(x) \geq 0, \quad i \in \{1, \ldots, m\}, \\
& h_i(x) = 0, \quad i \in \{1, \ldots, p\}, \\
x \in W_1 \times W_2 \times \ldots \times W_n, \quad & W_i \in \{\mathbb{R}, \mathbb{Z}, \mathbb{B}\}, i \in \{1, \ldots, n\}, m, p \textbf{ and } n \in \mathbb{N}^+
\end{aligned}$$

where x is a vector of n decision variables $x_1, \ldots, x_n$, $f(x)$ is the objective function that is used to evaluate different solutions, and $g(x)$ and $h(x)$ are inequality and equality constraints on the variable x_i. $\mathbb{B}$ indicates the set of binary values $\{0, 1\}$. Throughout our work, we will use the sub-class of the optimization program: Mix Integer Non-linear Programming (MINLP), Binary Mix-Integer Linear Programming (BILP) and Mix-Integer Linear Programming(MILP). MINLP problems combine the combinatorial difficulty of optimizing over discrete variable sets with the nonlinear constraints. Compare to MINLP

problems, if the object function and constraints include the linear functions with discrete and continuous variable sets, then we can call this optimization program MILP, moreover, if there are binary variables in the variable set, then it can be named BILP.

2.5 Tools

Throughout our work, we need to solve non-linear problem and linear problem, respectively. So first the framwork CasAdi[AG18] is adopted in order to solve non-linear program. CasAdi a symbolic framework for algorithmic differentiation and numeric optimization. It can generating complete, large and sparse Jacobians and Hessians matrix, and support the interface to access Bonmin(Basic Open-source Nonlinear Mixed Integer programming) [Bon17], which is open-source project for solving MINLP problems. For solving the MILP and BILP problems, we use the Gurobi [GO16] Optimizer, which is a state-of-the-art solver for mathematical programming. For our work, CasAdi version 3.1.0, Bonmin version 1.8 and Gurobi version 7.0 has been used.

3. Related Work

Up to now, many approaches have been worked out in TNEP area. Most of them focus on the different aspects of the TNEP, like the view of modeling from the AC model and DC model, and the performance of the optimization methods. The Comprehensive review of the work on TNEP can be found in [HHK13a] or [HHK13b], which provide the past achievements and current stage to find out further ideas in these fields. The AC model is complete and practical, but it leads to complex and need an efficient optimization technique to find solution. In paper [YH89] the author propose the AC optimization model, which includes security constraints on bus voltage magnitudes, swing angles and the AC load flow solution. Compare with the AC model, the DC model contains simplifications, so it is simple to understand and modeling, but it can be hard to implement the result in practice. In paper [RBH14] the author confirms the existence of the gap between the optimal solution of TNEP produced by the DC power flow approximation and the AC power flow heuristics, and proposes how to bridge this gap with the LPAC power flow model. Even the complexity of TNEP is NP-hard due to the subproblem Steiner trees, which is discussed and proved in [MPS10], the optimization technique are still rapidly developed to find optimal solution of TNEP. The methods can be concluded into two methods: mathematical and Meta- heuristic optimization approaches. The main works on mathematical optimization are generalized in two area: 1. How to convert power system equations into optimization programming model, which is much complicate and difficult, especially in practice. 2. Because of non-efficiency calculation of non-linear programming, lots of research works have been carried out to linearized the non-linear model.

Compared with the mathematical optimization, it is easier to use and simplify by using the heuristic optimization, but the optimal solution is not stable and will possible fall into local optimal. Moreover, the cost of simulation time is often high. So some other published works try to combine the benefits of both mathematical and heuristic approaches, which is called Meta-heuristic. In this chapter the related works concerning TNEP are listed.

3.1 Mathematical Optimisation Methods

For the reason that the networks in TNEP are various types and with different levels of complexity, which may leads to a hight computational complexity, generally the simplified versions by linearization and relaxations of TNEP are proposed in research works. Four different mathematical formulations used in transmission expansion studies-transportation models, hybrid models, DC power flow models, and disjunctive models are summerized and

compared in [RMGH02], which are also adopted and reformulated in our work. Moreover, the general description of DC system and the well known Garver's six-bus network, which is used to illustrate the proposed methods, can be found in [RMGH02].

Villasana [VGS85] conceptualize the transmission networks and proposed how the TNEP is solved in simple, applicable way to distribution circuits. However, The downside of this work is that, it consider TNEP in a local engineering preliminary planning level.

In general, TNEP is large-scale, mixed-integer, nonlinear, and nonconvex. Natalia Alguacil [AMC03] derive a mixed-integer linear formulation that simplify the problem by linearizing and relaxing it into a MILP.

During the linearization of the NLP constraints, the big-M-notation is widely adopted for TNEP, a survey to compute the minimal values for the big-M-notation is included in [MPS10].

Silvio Binato [BPG01] and Shengjun Huang [HD17] have described using the Benders decomposition approach to solve TNEP problems, which ensures finding the optimal solution by using additional constraints, iteratively evaluated. Shengjun Huang [HD17] also include dynamic searching and parallel implementation.

3.2 Meta-heuristic Optimisation Approaches

The classical heuristic approach are included in [Gar85], it first uses the linear programming for network analysis, in order to find where the capacity shortages exist, then present the algorithm to determine, where the new lines should be add to relieve the shortages. The algorithm described in [Gar85] has a well computational speed and simplicity. However, the disadvantage in this paper is that no mathematical link for the electrical laws in the linear model.

The TNEP problem is considered a very complex problem due to its combinatorial and nonconvex features. Different from using the mathematical optimization methods, some meta-heuristic methods have been proposed to tackle it. In the paper [RGM95] simulated annealing is proposed as a means of solving large scale problems in TNEP. It also gives the experiment for the comparison in a larger example system between the annealing method and a mathematical decomposition based optimization technique.

In paper [Cad09] the TNEP is included with two objectives: the transmission reliability efficiency and the cost of installing transmission lines, and this multi-objective optimization is solved by three different genetic algorithm.

In paper [LSRH09] two optimization tools based on artificial immune systems and differential evolution to solve the multi-stage TEP problem are presented. Silvio Binato has proposed using the greedy randomized adaptive search procedure in [SBdA01], which a meta-heuristic algorithm commonly applied to combinatorial optimization problems, to solve the transmission network expansion problem.

3.3 Summary

From the above studies, it can be seen that they mainly focus on how to install new lines in the existing paths and find the global optimal solution under reducing the computing time of large scale electrical network. Our work will be unique to provide the railway network as the candidate networks for expanding the existing electrical network and apply the former research to the combined network. This provides a unique alternative to expand the power networks in the face of the growing energy demand.

4. Datasets Preparing

In the recent years most research works focus on optimal planning of power transmission, which mainly discussed how to build lines at the known candidate edges. They did not involved the approach about the generating process of candidate routes, along which we can build lines to transfer power. In this chapter we want to represent the process to extract and generate the datasets from the real word, which include the existing power grid $N_1 = ((V_1, E_1), b_1, cap_1, s_1)$, the railway network $N_2 = ((V_2, E_2), b_2, cap_2, s_2)$ and the connection network $N_3 = ((V, E_3), b_3, cap_3, s_3)$, where $V = V_1 \cup V_2$. The existing power grid N_1 support the job for transmitting the power in order to meet the current demand of energy. The railway network N_2 and connection network N_3 are adopted to be the candidate network, which can be used to install new lines to transmit the power, when the current transmission network N_1 cannot support the future transmission task. Our intuition is that we can take advantage of the similarity between the power grid and railway network, such as they both are designed to resource allocation, moreover, they extended in all directions. However, there are no connections between railway network and power grid, building connections between both networks is necessary, we call these connections the connection network N_3. Note that in this chapter we consider mainly about the topology graphs of the before mentioned three networks: $G_1 = (V_1, E_1)$, $G_2 = (V_2, E_2)$, $G_3 = (V, E_3)$. The electrical parameters of the networks will be discussed later, especially the network N_2 and N_3, which cannot be directly accessed from real world, will be assigned by default value. In this chapter we describe how these networks for Germany are obtained.

First, the datasets in our work builds mainly on railway network and power grid, which extracted from OpenStreetMap(OSM), so the first part of this chapter OSM is involved, see Section 4.1. Then Section 4.2 and Section 4.3 describe the approaches to export the power grid and the railway network respectively. Different from the railway network, which is generated by our work and the power grid can be obtained from project SciGRID [Med18], which is available and can be downloaded from the SciGRID project webpage. After extraction from OSM, the topology graphs G_1, G_2 and the related data (e.g.the length of edge, electrical information and geo-location) of the power grid and the railway network are available, we need ensure the railway network is connected, in order to transfer power between any two buses in power grid, which is discussed in Section 4.3.2. Now the job is to generate the connection graph G_3 between graph G_1 and G_2, see in Section 4.4, which is implement by KD-tree algorithm. However, to avoid unnecessary computing on our connected network, we need to remove the vertices with degree two in the candidate graph G_2 and G_3, see Section 4.4.4.

4.1 Data Source: OpenStreetMap (OSM)

The work discussed in this chapter base on the free data, which is available from the collaborative project OpenStreetMap (OSM)[Wik14], which creates and distributes free geographic data for the world. The data of the railway network and power grid can be downloaded in an XML format, which follows a certain schema definition. It has three data primitives: nodes, ways and relations. The node present the vertex in the graph, the way, which contains an ordered collection of nodes, shows the edge in the graph. The relations consists of a block of ways to describe the area of special meaning, such as the building, the park and so on. For each primitive it has special tag to describe it's meaning and attribute. For example, the geographical locations of the nodes are defined by tag lon and lat, which means latitude and longitude to sepcify the location in geographical system. For the dataset of the work we focus on the relevant for the power grid and railway network, which are filtered separately by using the "power" tag and the "railway" tag [Wik17].

Figure 4.1: A church presented in OSM, which consists of one outer and one inner ring. ((OSM-ID=1596371))

In Fig. 4.1 there is a church, which is random extracted from the OSM. It is represented as a multipolygon relation, which is the combination of the way ("114397140") in the role outer (the outline) and the way ("114397134") in the role inner (the holes), see in Fig. 4.2. The endpoints of the both ways are joined, form a closed polygon. The geographical information of the church can be accessed from the tag in nodes.

```
<relation id="1596371" uid="62423">
 <member type="way" ref="114397140" role="inner"/>
 <member type="way" ref="114397134" role="outer"/>
 <tag k="building" v="church"/>
 <tag k="type" v="multipolygon"/>
</relation>
```

```
<way id="114397140" uid="165061">          <way id="114397134" uid="165061">
 <nd ref="1296015964"/>                      <nd ref="1296015891"/>
 <nd ref="3091060136"/>                      <nd ref="1296015593"/>

        ...                                          ...
 <nd ref="1296015964"/>                      <nd ref="1296015891"/>
</way>                                       </way>
```

```
<node              <node                    <node              <node
id="1296015964"    id="3091060136"          id="1296015891"    id="1296015593"
lat="49.0328274"   lat="49.0327912"   ...   lat="49.0331935"   lat="49.0330027"   ...
lon="8.4406380"    lon="8.4407201"          lon="8.4409257"    lon="8.4413649"
/>                 />                       />                 />
```
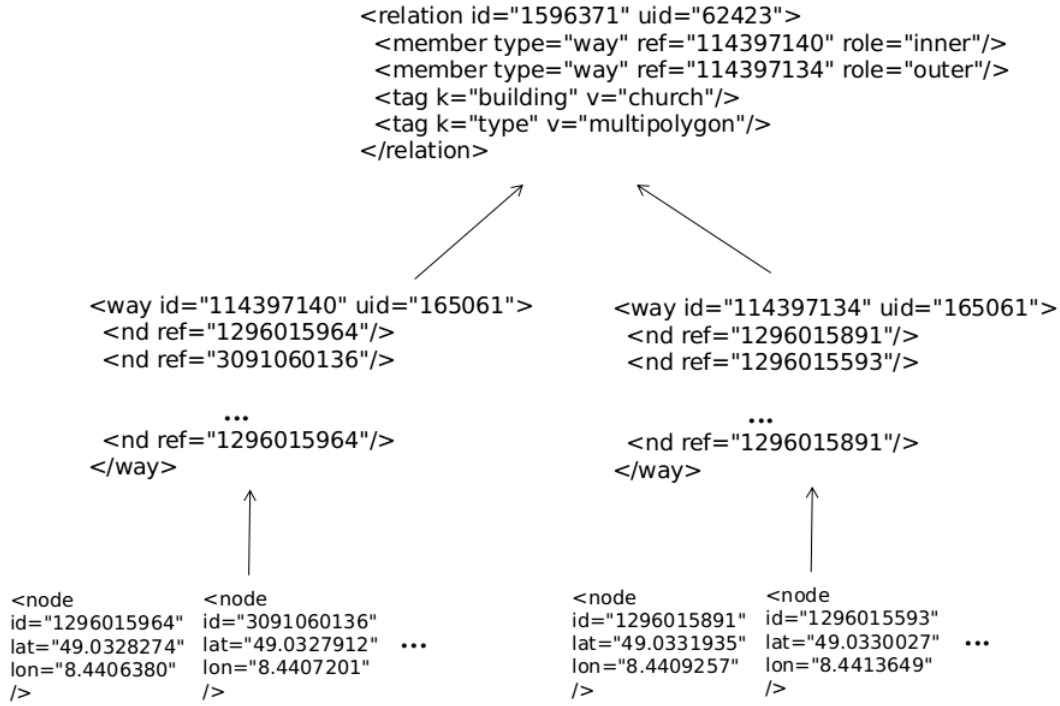
Figure 4.2: Display the church in OSM. (OSM-ID=1596371)

4.2 Reduced German Power Grid (SciGRID)

In this section we mainly explain how the power grid is generated by SciGRID project [Med18], which is a free, open source project and also builds on OSM data as the extraction of railway network. Since the current researches on the transmission network are integrated in some undisclosed internal energy system model, it is hard for people to discuss the underlying approaches, procedures, and results of the structure, simplifications, and the abstraction involved in the transmission network. In such a case, SciGRID aims at building an open source model of the electrical transmission network in Europe, which can provides the reliable and stable dataset of transmission network.

There are four steps in the SciGRID model: OSM data download, power data filtering, data export to a database, and data abstraction. As the approach to store the railway network, the topology structure of the power grid and the electrical parameters are exported as .csvdata files. In this section we explain the process to generate power grid by above mentioned four steps. Note that, this process are described by using German power grid.

4.2.1 OSM Data Download and Filtering

The OSM data in format PBF (Protocolbuffer Binary Format). PBF is primarily intended as an alternative to the XML format, which was designed to support future extensibility and flexibility and can be download from the website [osm18], as described in SciGRID, the status of the data is November 2015. After downloading the PBF file, the OSM data can be extracted and filtered by using the java tool osmosis [Wik18], which is a command line java application for dealing with OSM data, such as extracting data inside a bounding box and filtering the data with respect to the specific value of the tag. In SciGRID project, it filter the nodes, ways, and relations in OSM data, which have the "power" tag. Note that, only relations where the voltage tag has a value of $220kV$ and higher are considered,

and the relations has the tags "power=construction", "power=planned" or "power=fixme" or has no power tag are excluded from SciGRID. Since we focus only on German power grid, this can be done by setting the osmosis option –bounding box, which need to be indicated the the latitudes of the top, bottom bounding box, and the longitudes of the left and right edges of the bounding box, respectively. The German power grid is extracted in SciGRID project by following osmosis command: "–bounding-box top=56 bottom=46 left=5 right=16". Fig 4.3 and Fig 4.4 represent a nuclear generator in OSM and how
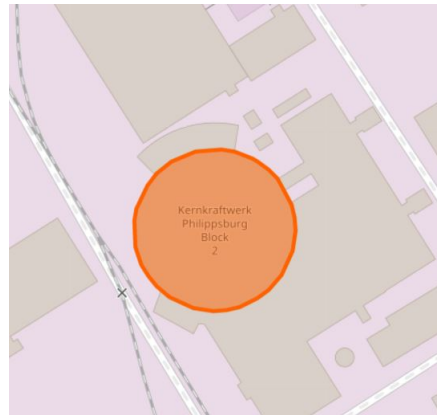
Figure 4.3: A nuclearplant in OSM (OSM-ID=94701596)

```
<way id="94701596" visible="true" version="17" changeset="27587097" timestamp="2
014-12-20T13:39:51Z" user="mapper999" uid="165061">
<nd ref="1099514272"/>
<nd ref="3244468654"/>
<nd ref="1099514248"/>
<nd ref="3244468646"/>
            ...
<nd ref="3244468648"/>
<nd ref="1099514296"/>
<nd ref="3244468651"/>
<nd ref="1099514272"/>
<tag k="building" v="yes"/>
<tag k="building:colour" v="#DDDDDD"/>
<tag k="frequency" v="50"/>
<tag k="generator:method" v="fission"/>
<tag k="generator:output:electricity" v="1468 MW"/>
<tag k="generator:plant" v="output"/>
<tag k="generator:source" v="nuclear"/>
<tag k="generator:type" v="PWR"/>
<tag k="height" v="50"/>
<tag k="name" v="Kernkraftwerk Philippsburg Block 2"/>
<tag k="operator" v="EnBW Kernkraft GmbH"/>
<tag k="power" v="generator"/>
<tag k="ref" v="KKP 2"/>
<tag k="start_date" v="1984-12-17"/>
<tag k="wikipedia" v="en:Philippsburg Nuclear Power Plant"/>
</way>
```

Figure 4.4: The OSM file for a nuclearplant in OSM (OSM-ID=94701596)

it is described in OSM file, respectively. This way is filtered, when we find the tag <tag k="power" v="generator"/>, which contains keyword "power". with the OSM file, the useful information of each node, way and relation can be accessed by iterate all its' tag.

Here we can also know this building can output $1468MW$ power and the power type is nuclear according to the key values "source" and "output".

4.2.2 Export to the Dtabase

After filtering the OSM data, the SciGRID project convert the OSM data to PostGIS-enable PostgreSQL database, in order to analyse the power data. The PostgreSQL is an object-relational database management system. PostGIS is a spatial database extender for PostgreSQL database, it adds support for geographic objects allowing location queries to be run in PostgreSQL. All these works are done by osm2pgsql, which is a command-line based program. osm2pgsql can define the settings of how the OSM data are store in database by providing a .style file, such as which columns in database are created for the tables of the power data, and which information need to be stored and which need to be ignore.

4.2.3 Abstraction

The SciGRID project is based on the "power" relations, which has a key "route" and value "power". Normally these relations consist of one or more substations and one or more transmission lines. Substations are defined in OSM by the key "power" and the values "substation", "sub_station", "station", "plant", "generator". Transmission lines are on the other side defined by the key "power" and the values "line" and "cable". Here is the table 4.2, which describe number of relations with key/value pair route/power in the dataset of Germany in OSM (Status: 09.11.2015).

Table 4.1: Number of relations with key/value pair route/power in the dataset of Germany in OSM (Status: 09.11.2015). The total number of relations is subdivided into sets of relations with different numbers of substations or categorized as discarded [Med18]

Description	Number
total number of "power" relations in Germany	826
relations to be fixed / being planned / being under construction	54
relations with 0 substation	2
relations with 1 substation	11
relations with 2 substations	690
relations with 3 substations	65
relations with 4 substations	4

The SciGRID project consider the following relations: 1. Relation with only two substations and one/several transmission lines linking them, example see in Fig 4.5. 2 Relations with three substations, example see in Fig 4.6. 3 Relations with three substations and a T-junction, example see in Fig 4.7. The reason to just select these kinds of relations is, that the calculating of the transmission lines length is more easy, when we can define which transmission lines are connected to which substation in above mentioned relations. Relations with zero or one substation are also not considered in SciGRID as they constitute incomplete electrical circuits.

In the abstraction of SciGRID the vertices represents the geographical center positions of the substations and the edges represents the transmission lines without considering the path in real word. In other words, we can get the topology of the electrical transmission network, which are store in SQL by using two tables: the vertices table, which contains the

Figure 4.5: Example of a relation constituted of two substations. The two red circles are the two substations contained in the relation (OSM-ID=1637161) and the black line are the transmission lines (this relation contains four transmission lines)[Med18].

Figure 4.6: Example of a relation (OSM-ID=339244) containing three substations (circles), a T-junction and seven transmission lines. Figure obtained using overpass-turbo.eu, credits: ©OpenStreetMap contributors[Med18].

vertices of the transmission network and the links(edges) table, which contains the links of the transmission network. Furthermore, in the links table the electrical properties of the transmission lines: resistance r, reactance x, capacitance c, and maximum current thermal limit $\bar{f}$. are calculated and insert to the links table for further research. In SciGRID, the electrical properties are calculated in per km units as follows[Med18]:

$$r_{ohmKm} = C_r/(\frac{wires}{wires_{typical}})/(\frac{cables}{3}) \tag{4.1}$$

$$x_{ohmKm} = C_x/(\frac{wires}{wires_{typical}})/(\frac{cables}{3}) \tag{4.2}$$

$$c_{nFKm} = C_c/(\frac{wires}{wires_{typical}})/(\frac{cables}{3}) \tag{4.3}$$

$$\bar{f}_A = C_I/(\frac{wires}{wires_{typical}})/(\frac{cables}{3}) \tag{4.4}$$

where, $wires_{typical}$ is the number of wires in a transmission cable. $wires_{typical}$ as is typically 2 for transmission lines of 220kV and 4 for transmission lines of 380kV. The coefficients used in Eq4.1-4.4 are listed in table 3 from reference [sci12].

Figure 4.7: T-connection from relation (OSM-ID=339244). The T-node represented by the blue filled circle, is the node at which the transmission line at the top part of the figure branch into two parts. Figure obtained using overpass-turbo.eu, credits: ©OpenStreetMap contributors[Med18].

Table 4.2: Electrical properties coefficients from reference[sci12]

Voltage level	C_r (ohm/km)	C_x (ohm/km)	C_c (nF/km)	C_I (A)
380kV	0.025	0.25	0.0137	2.6
220kV	0.080	0.32	0.0115	1.3

4.2.4 Conclusion

After abstraction, the SciGRID project output all the dataset in the .csvdata file. And it also provide the script to visualize the abstracted power grid. An example is shown in Fig 4.8. The SciGRID user can also find the power grid for any other region by using the previous steps.

4.3 Generation of the German Railway Network

Fortunately the German OSM data, which include the railway network, can also be download online in format PBF. With the help of software JOSM the data in format PBF can be transformed to XML. Then our job is to filter relative dataset (nodes, ways and geo-location) and generate the extracted railway network.

4.3.1 Filtering the Railway Network

In OSM the railway tag is used to tag the ways for many kinds of railways including light rail, mainline railways, metros, monorails and Trams. Besides the actual rails, it is also used for tagging railway infrastructure. It is necessary to pick out which railway tag can be used to filter the railway network, since the track paths are represented as the candidate edges, along which the lines may be built. We want, that the track paths are picked reasonable, such as it cannot be infrastructure. So the railway tracks should be picked under the consideration of two aspects: the environmental impact and the economic benefit. First, the citizens do not want, that in their living environment exists lots of power lines, which are difficult to maintain and easy to cause dangers from storms and trees. These are what kill most people in electrocution lawsuits. These are the reasons that most power lines fall and kill an unsuspecting homeowner, child or utility worker. Depending on these reasons we want to use the track paths, which is away from the cities. Second, the tracks under
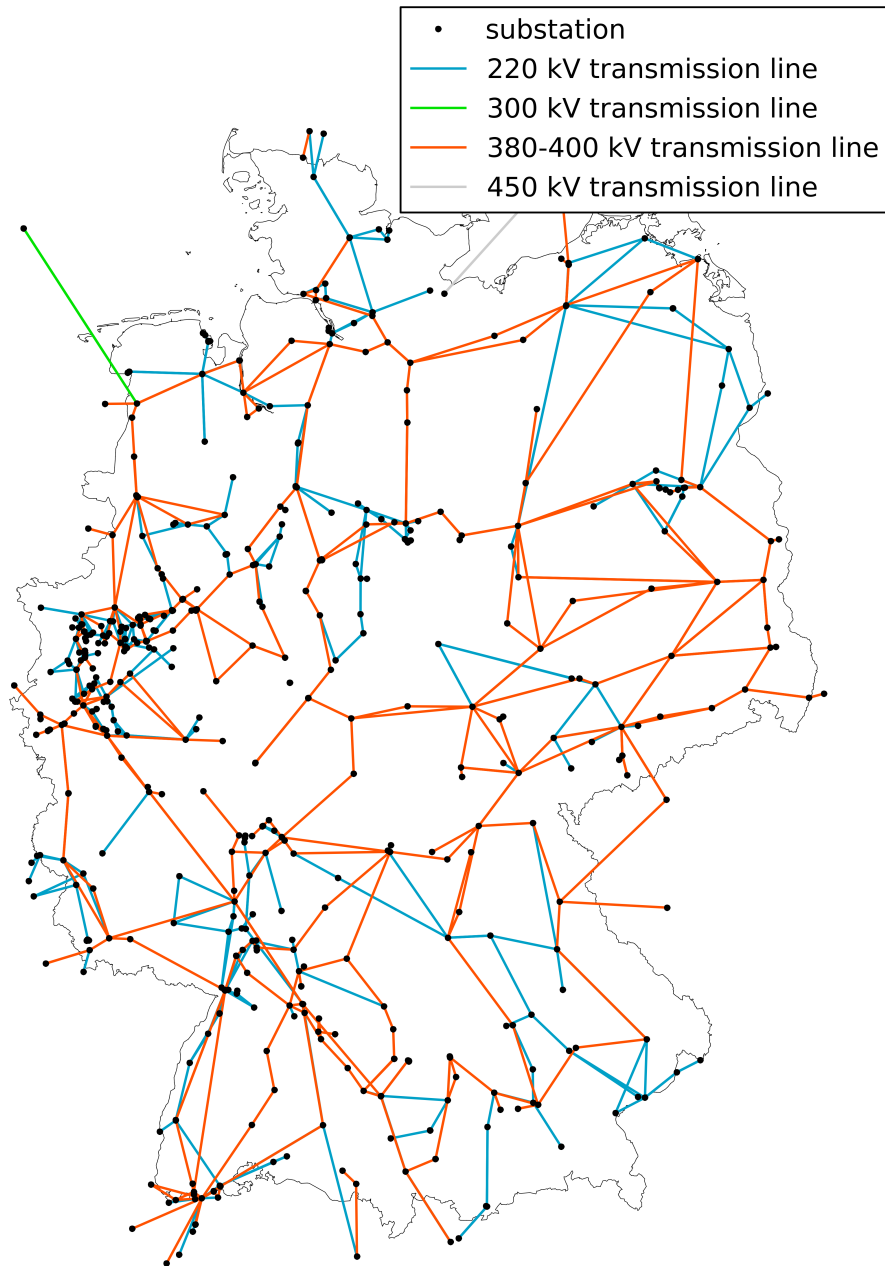
Figure 4.8: Example of the plotted SciGRID abstracted model network using the plot function available in SciGRID.py, status 09.11.2015. Credits: ©OpenStreetMap contributors[Med18].

construction or abandoned should not consider, while planning on the indeterminate paths would involved unnecessary analyze, such as why these paths are abandoned. So the criteria for picking tag are: 1. away from city, 2. already existed. All the tag value about railway network in OSM are shown in Table 4.3. Here we adopt tag "rail" to filter the railway network.

Table 4.3: Value of Tag "railway"

Value	Comment
abandoned	abandoned tracks and infrastucture removed.
construction	railway under construction.
disused	no longer used
funicular	cable driven inclined railways.
lightrail	a higher-standard tram system
miniature	miniature railways
monorail	a railway with only a single rail.
narrowgauge	Narrow-gauge passenger.
preserved	a historic railway.
rail	the standard railway for the country or state.
subway	A city passenger railway.
tram	One or two carriage rail vehicles.

During filtering the ways (edges) of railway graph, the way relations always consists of lots of nodes (vertices) to simulate the curve in real word. However, in our work we consider mainly about the topology structure of railway network. So for each way relations we simply remove all nodes between the header node and the tail node, meanwhile we sum the real length of the way. Example see Fig. 4.9. Note that in practice computation $len_{vw} = D_{vw} \quad v, w \in \{a, b, c, d\}$ (see Section 2.1.1), here we simply assume the length by default setting without consideration of latitude and longitude, which can be easy to understand.
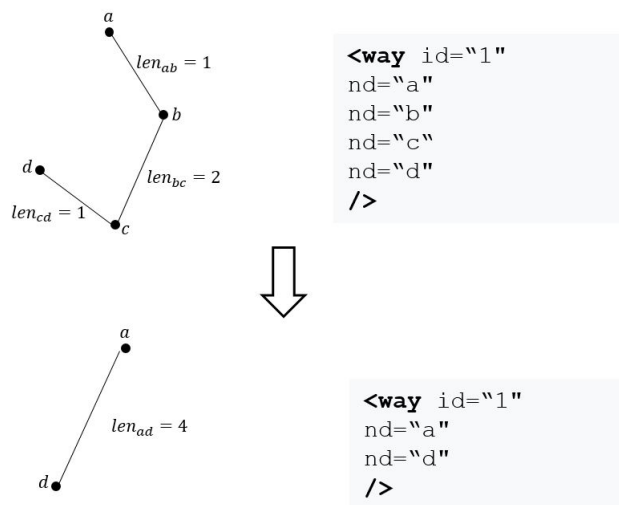
Figure 4.9: Contraction nodes during filtering way

Table 4.4: The number of edge and vertex of German railway graph

	Edge	Vertex
Before removing	307188	404044
After removing	53224	48263

Table 4.4 display the number of edge and vertex before and after above mentioned contraction. The extracted German railway graph G_2 by using above mentioned approach is shown in Fig. 4.10.
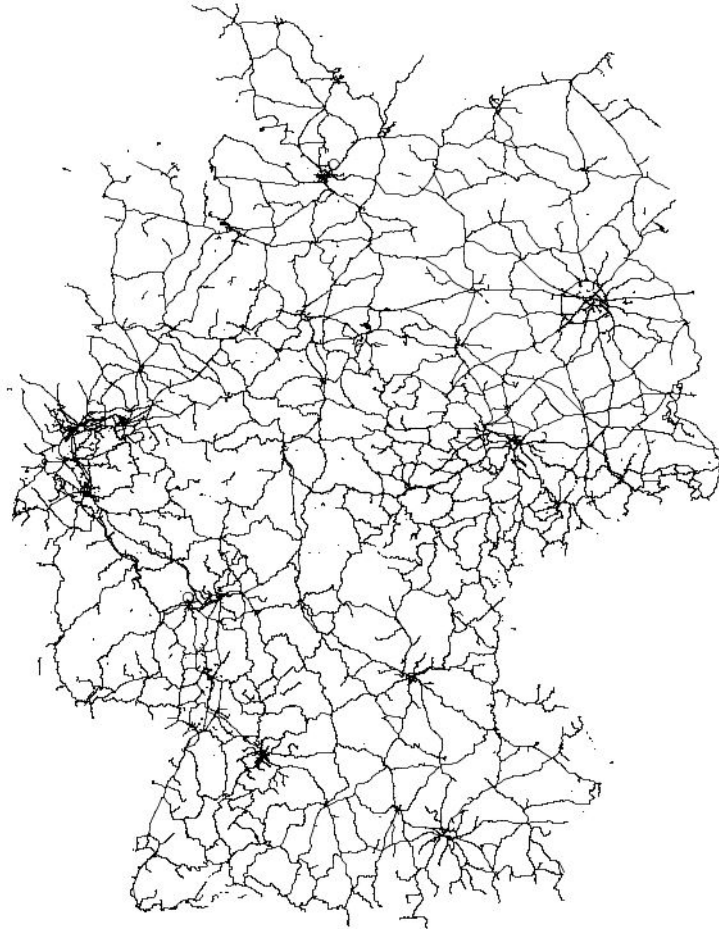
Figure 4.10: The German railway network by contract internal nodes after extracting from OSM

4.3.2 Find the maximum connected Sub-graph in Railway Graph

However, we can find the number of vertex is larger than the number of edges, in other words, according to lemma 2.2 the extracted railway network is disconnected. Our purpose is, that using railway network transfer power. Consider such extreme case: after building connections between power grid and railway network, power generators and loads are connected with two sub-graph of railway network, which are disconnected with each other. In other words, if we increase the demand of power load, there exists no path in railway network to transfer power from generator to load. This may live up to our intentions.
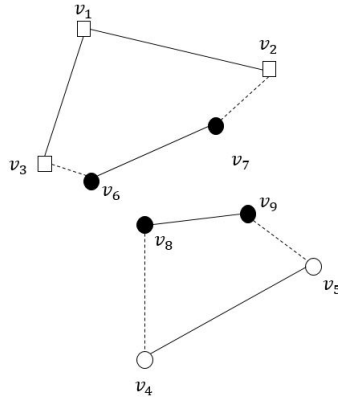
Figure 4.11: Example of the extreme case for the disconnected problem in the railway graph. Boxes denote generator buses (vertices in the power grid), circles denote load buses (vertices in the power grid) and filled dots denote transmission buses (vertices in the railway network). The whole graph consists of the power graph $G_1 = \{\{v_1, v_2, v_3, v_4, v_5\}, \{(v_1, v_2), (v_1, v_3), (v_4, v_5)\}\}$, the railway graph $G_2 = \{\{v_6, v_7, v_8, v_9\}, \{(v_6, v_7), (v_8, v_9)\}\}$ and the connection graph $G_3 = \{\{v_2, v_3, v_4, v_5, v_6, v_7, v_8, v_9\}, \{(v_3, v_6), (v_2, v_7), (v_4, v_8), (v_5, v_9)\}\}$.

In figure 4.11, we use the example to explain the disconnected problem in railway graph. For the existed power graph G_2, since the railway graph G_2 is disconnected, it cannot be used to transmit power from the generator buses v_1, v_2 and v_3 to load buses.

To avoid the above case, we need to find the maximum connected sub-graph in railway graph, which is connected meanwhile has largest numbers of vertices, since the network, which has more vertices and edges, can provide more candidate paths to build circuits. To discover the maximal connected sub-graph, we can use a traversal algorithm, either depth-first or breadth-first. First, we need to discover all connected sub-graph, then select one which has most vertices. To discover all connected sub-graph, we should set all vertices in railway network unvisited at the beginning. Then we do a traversal starting from a random vertex $v, v \in V_2$, which means we will visit all the vertices that can be reached from v. If there are other connected sub-graph, their vertices must still be unvisited after this traversal. We just need to start another traversal run from one of those unvisited vertices to find another connected sub-graph. When all the vertices in graph G_2 are visited, we will have discovered all the connected sub-graphs. During each traversal we can set a flag to record the number of vertices in the connected sub-graph to help us find the maximal connected sub-graph. In the algorithm 4.1 we use breadth-first search to do the traversal. The algorithm is given in pseudo-code, and the maximal connected sub-graph of German railway network is shown in Fig. 4.12. The table 4.5 display the number of vertex and edge in railway graph G_2 when it keeps the maximum connected sub-graph and removing other unnecessary sub-graphs.

Table 4.5: The number of edge and vertex of the maximum connected sub-graph in German railway graph

Edge	Vertex
30457	28839
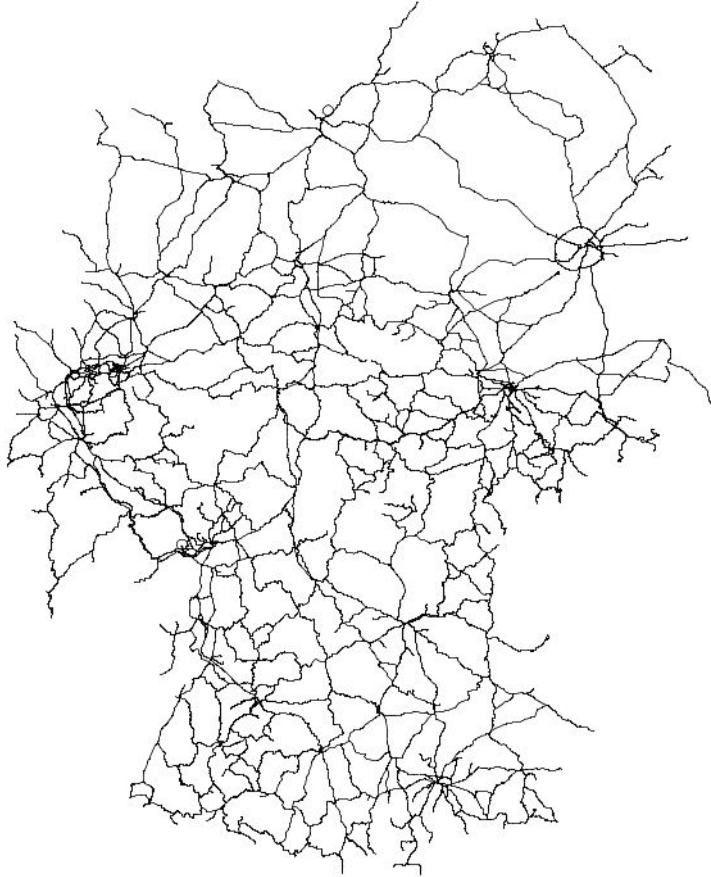
Figure 4.12: The maximal connected Sub-graph of the German railway network

4.4 Building Connections

As mentioned in the beginning, the purpose to build connections G_3 between the power graph G_1 and railway graph G_2 is to make the whole graph $G = G_1 \cup G_2 \cup G_3$ connected, and the disconnected problem in transmission network we have discussed when we extract the railway graph. However, the paths to built connections is difficult to extract from real world, in other words, these paths are not explicit existed as power grid and railway graph. So the connections need to be create under consumptions: 1. In order to avoid unnecessary cost, if we want to connect one vertex $v, v \in V_1$ to the railway network G_2, the most simply way is to find the nearest vertex $w, w \in V_2$, since the cost of building a new line along edge (v, w) depends largely on the length of the edge (v, w). In our work we do not consider about building the model to find the cost of installing a new line or building connections between the power grid and railway network, but the cost on installing a new line along connections should be larger than the cost on installing a new line along railway network, since railway network is already existed. 2. We have two choice to building connection, one is that for each vertex in power graph find the nearest neighbour in railway graph, another one is that for each vertex in railway graph find the nearest neighbour in power graph. Since the distribution of the vertex in power graph is separate, if we select second, it can create the connections, which is longer than the connections by the first choice. Our intuition is to keep the connection as short as possible, then the first choice seems more reasonable to generate connections between power graph and railway graph. 3. To keep the diversity of the selection in candidate path to transmit power by using railway network, only considering the relation of the nearest neighbour seems not enough, so we propose for each

---

Algorithm 4.1: CONNECTED SUBGRAPH

 Input: Graph $G = (V, E,)$
 Data: Queue Q
 Output: Flag flag(i), the number of vertices in the ith connected subgraph

  // Initialization
1 **forall** $v \in V$ **do**
2 | visited(v) = $false$ // Mark all vertices as unvisited

  // Main loop
3 $i = 0$
  // ith connected subgraph
4 **forall** $v \in V$ **do**
   | // If v is not yet visited, it's the start of a newly
   | // discovered connected component containing v
5 | **if** visited(v) **then**
   | | // traversal of vertex v
6 | | flag(i)++
7 | | Q.CLEARQUEUE()
8 | | Q.ENQUEUE(v)
9 | | visited(v)=true
10 | | **while** Q *is not empty* **do**
11 | | | $w \leftarrow$ Q.DEQUEUE()
12 | | | **forall** $(w, u) \in E$ **do**
13 | | | | **if** visited(u) **then**
14 | | | | | visited(u) = $true$
15 | | | | | Q.ENQUEUE(u)
16 | | | | | flag(i)++

17 | $i++$

---

vertex $v, v \in V_1$ in power graph find the k nearest neighbours $w_1, w_2...w_k, w_1, w_2...w_3 \subset V_2$ in railway graph.

4.4.1 K nearest Neighbours

K Nearest Neighbours query is given a query vertex and the positive integer K, from the dataset to find the K closest to the query vertex, when $K = 1$, it becomes the nearest neighbour query. In the general process to find K nearest neighbors (KNN) of one vertex $v, v \in V_1$, first this distances from v to all vertices in railway graph G_2 are to be calculated and then only nearest k-points can be discovered for voting. This approach is also known as brute-force approach. When the volume of data is huge, in hundreds or thousands, this repeated distance calculations can be very tedious and time consuming. To fasten up this process and so as to avoid measuring distances from all the points in the dataset, some prepossessing is necessary. This preprocessing helps to search vertices which are likely to be close to it. There are several indexing data structures can achieve this goal, such as vantage point tree, ball tree , k dimensional tree (kd-tree), through this work the KNN search is mainly based on kd-tree.

4.4.2 K dimensional Tree (kd-tree)

A kd-tree is a space-partitioning data structure for storing a finite set of vertices from a k dimensional space. A kd-tree is a binary tree. Each node in kd-tree represents a

hyperplane, which perpendicular to the axis of the current partition dimension, and divided the dimension space into two parts; one part in its left sub-tree, another part in its right sub-tree. In other words, when the current node division dimension is a, where $a \leq k$, the coordinate value in a dimension of all vertices in the left sub-tree are less than all vertices in the right sub-tree, then the process of building the tree iterates through the dimensions to partition the sub-tree and stops when a node has less than one. During the partitioning through the different dimensions for the different levels of the tree, normally we use the median point for the partition, in order to make the tree balance. In our work, since the nodes in the networks are located by the latitude and the longitude, the following explanation are based on 2d-tree. An example of two dimensional tree is shown in fig 4.13. It is showed that if the median of n elements were found in $O(n)$ time, a depth-balanced kd-tree can be built in $O(nlogn)$ time. We use the above mentioned process to construct the kd-tree for the set of vertices V_2 from the German railway graph.
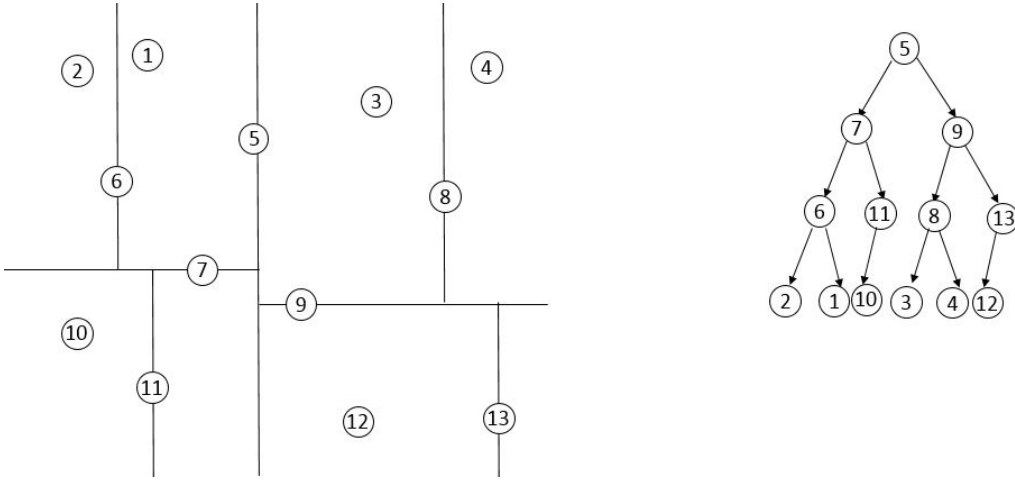
Figure 4.13: An example for two dimensional tree, here it is assumed, that the vertices have only two dimension: the horizontal and the vertical, the initial partition dimension is the horizontal.

4.4.3 Find k nearest Neighbors

In this section the process to find the k nearest neighbors $v \in V_2$ of each given vertex w, $w \in V_1$, is represented. We are interested to get k nearest neighbors, where k can be any value. However, we will first see how to get the closest one point. It can then be easily extended to understand how to get more.

Suppose here is the constructed kd-tree T for the set of vertices V_2 from the German railway graph. For a given vertex w, $w \in V_1$, the process of searching the closest neighbor is given in flow:

Step 1: The query start from the root node of T, according to the comparison of the query vertex and each node in T keep going down, until reaching the leaf node of T. The comparison between w and the node in T refers to the comparison of the value corresponding to the k dimension ($k = 0$, comparing the value of latitude, $k = 1$, comparing the value of longitude). If the value of vertex w small than the compared node corresponding to the k dimension, then the left subtree is visited, otherwise visit the right subtree. When the leaf node l is reached, the geo-distance $D(w, l)$ between w and l is calculated, and record l as the current nearest neighbor, $D(w, l)$ as current minimum geo-distance.
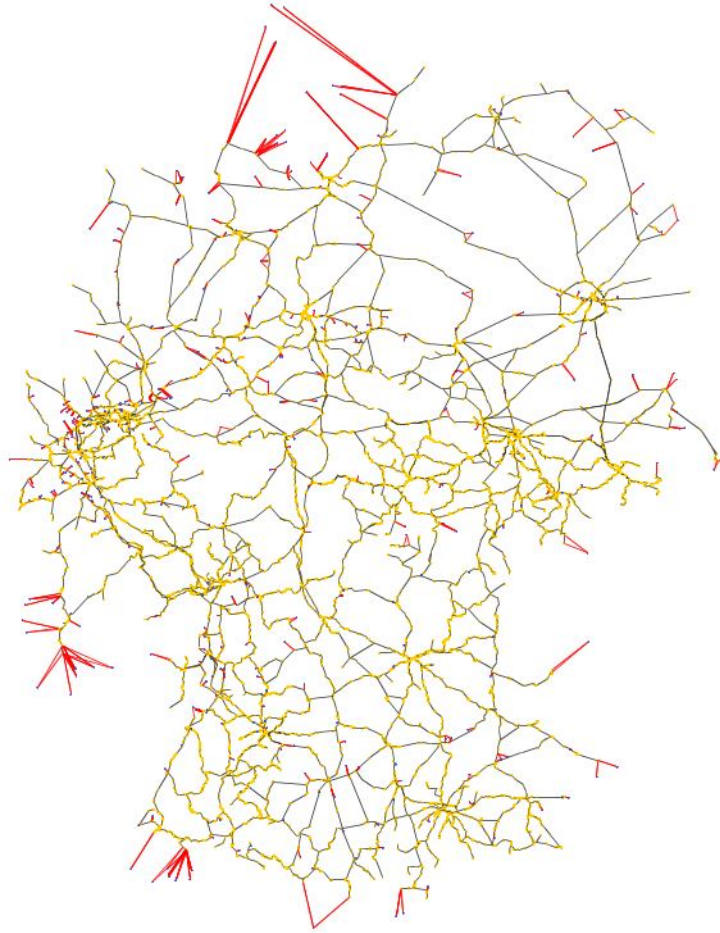
Figure 4.14: Building connections between buses and railway vertices, which are built between the power grid and the railway network. Here we set $k = 3$. The vertex in the power grid is drawn of the color blue, the vertex in the railway network is drawn of the color yellow and the connection is drawn of the color red.

Step 2: Backtracking the tree, backtracking is performed to find nearest neighbor of w. That is, if there is a vertex in the tree T closer to w in the unvisited branch, the distance between them is less than the $D(w, l)$. If the distance between the unvisited branch of the parent node of the current visited node and w is less than $D(w, l)$, we think that there are vertex much closer to w in the unvisited branch, then visit this branch, the same query process as Step.1. If find much closer vertex, then update the current nearest neighbor vertex l, and updates the distance $D(w, l)$. If the distance between the unvisited branch of the parent node of the current visited node and w is greater than $D(w, l)$„ it indicates that there is no vertex much closer to w in the branch. The judging process of the backtracking is from the bottom up, until it goes back to the root node.

Instead of keeping a single nearest vertex we could maintain a priority queue (max heap) to keep k nearest vertices. The first k vertices would be enqueued anyway. Onwards, a new point, if better, would replace the worst of the nearest neighbors found so far. That way we can maintain the k nearest neighbor easily.
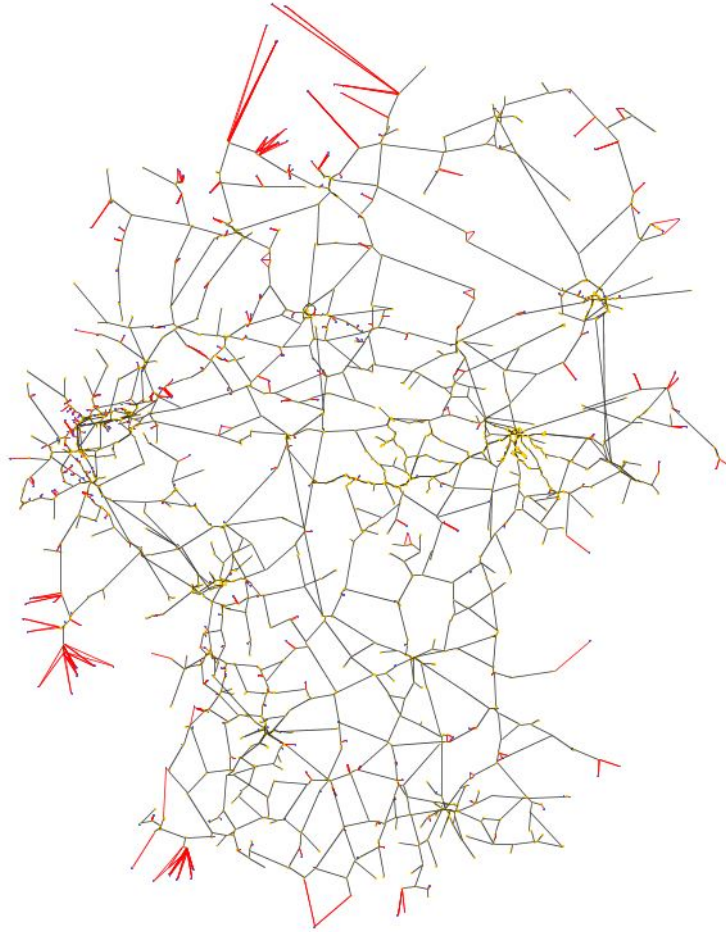
Figure 4.15: Remove the railway vertices with degree two in the candidate network

4.4.4 Two Degree Vertices Contraction

After building connections N_3 between the power graph N_1 and the railway graph N_2, to simplify the subsequent calculation, we contract the vertex $v, v \in V_2$, which degree is two. Because during the TNEP for the vertices in the railway network, they work as the transmission bus, which generate and load none power. If the degree of the railway vertex is two, it is not necessary to set a bus to shunt the power flow.

The Fig. 4.15 display the railway network and the connection network after contracting the railway vertices with degree two. Here even we remove the two degree vertices in the candidate network and the new neighborhood relations generate for the rest vertices, the Geo-distance between two vertex in the contracted candidate network keeps original.

4.4.5 Conclusion

Here all the dataset of the railway network and the connection network are also output in the .csvdata file, which store the topology of the candidate network and the weight of edges. The missing electrical parameter of the candidate network are assigned default value in the test case.

5. Mathematical Modeling

The transmission network expansion planning(TNEP) optimization is to determine optimal investments on transmission line additions from the candidate network to satisfy reliability criteria under future load and generation scenarios. In the last decades many mathematical programming approaches have been developed and well summarized. In this chapter we present the description of three different optimization models, which are frequently discussed for TNEP [RMGH02]. The first model is a non-linear mixed-integer model. However, this model has an innegligible problem, that non-linear model cannot be efficiently solved and find possible solution on large-scale problem in practice, which means a further simplification is needed. So another linearized mixed binary model is proposed determing which line should be built. Since we commonly want to built multiple lines along one edge in the candidate graph, the improved mixed-integer linear model is introduce in the end. In view of that TNEP should ideally be performed jointly with the investment of generation, the following models take both the cost of installing lines and power generation into account. In the modeling of TNEP, we have the following settings: first we fixed the loaded power of each vertex in the power grid; second the generation of each vertex is considered as a variable but with upper bounds; at last the new lines only installed in the candidate networks.

Note that the construction of all three models mentioned above are based on the graph $G = (V, E)$, which consists of the power graph $G_1 = (V_1, E_1)$, the railway graph $G_2 = (V_2, E_2)$ and the connection graph $G_3 = (V, E_3)$, where $V = V_1 \cup V_2$ and $E = E_1 \cup E_2 \cup E_3$. The process for generating the graph $G = (V, E)$ is described in Chapter 4. Here is the symbols description, which are used in this chapter for building the mathematical model. C_{vw} represents the cost of a line that can be built along edge (v, w). n_{vw} and Z_{vw} represent, respectively along edge (v, w) the number of lines built and the decision variable, whether line is built. The $\bar{n_{vw}}$ is the maximum value of n_{vw}, in other words, the maximum lines can be built along edge (v, w). g_v is the generation in bus v, whose maximum value is $\bar{g}_v$, P_v is the cost of the energy produced in bus v by the generating, which is measured in €$/MW$. the constant l_v represent the power loaded at bus v, which is already given before optimization. f_{vw} is the power flow from bus v to bus w, the capacity of power flow from bus v to bus w is represented by cap_{vw}. b_{vw} is the susceptance of a line along edge (v, w).

5.1 Mixed-integer non-linear Programming (MINLP Model)

In this section, a formulation of the TNEP problem is presented, which is a mixed-integer non-linear problem(MINLP) [RMGH02]. In this model the DC power flow equations are

represented with integer transmission circuit investment decision variables. Using this model the classical TNEP problem can be written in the following form:

Minimize:

$$h = (1 - \alpha) \sum_{(v,w) \in E_2 \cup E_3} C_{vw} n_{vw} + \alpha \sum_{v \in V_1} g_v P_v$$

Subject to:

$$g_v - l_v = \sum_{(v,w) \in E} f_{vw} \qquad \forall v \in V \qquad (5.1)$$

$$f_{vw} - n_{vw} b_{vw} (\theta_w - \theta_v) = 0 \qquad \forall (v, w) \in E_2 \cup E_3 \qquad (5.2)$$

$$f_{vw} - b_{vw} (\theta_w - \theta_v) = 0 \qquad \forall (v, w) \in E_1 \qquad (5.3)$$

$$0 \leq g_v \leq \bar{g}_v \qquad \forall v \in V_1 \qquad (5.4)$$

$$|f_{vw}| \leq n_{vw} cap_{vw} \qquad \forall (v, w) \in E_2 \cup E_3 \qquad (5.5)$$

$$|f_{vw}| \leq cap_{vw} \qquad \forall (v, w) \in E_1 \qquad (5.6)$$

$$|n_{vw}| \leq n_{vw}^- \qquad \forall (v, w) \in E_2 \cup E_3 \qquad (5.7)$$

$n_{vw} \in \mathbb{N}_0$ integer, $f_{vw} \in \mathbb{R}$, $g_v \in \mathbb{R}$ and $\theta_v \in \mathbb{R}$

The objective function consists of a weighted sum of the investment cost of new added lines and the operating cost of the generating units. The operating cost of the generating units is influenced by the power source type, the dataset of the generating units described later in Section 2.3.3. Depending on the above mentioned two factors of investment: the cost of new added lines and the operating cost of the generating units, the weighting factor α is used to adjust the objective function in our model. The Constraint 5.1 represents the conservation of power in each node. In other words, this is the formulation for Kirchhoff's current law (KCL) in the DC model. The Constraints 5.2 and 5.3 are an expression of Ohm's law. Constraint 5.2 models over the candidate graph, which includes connections graph $G_3 = (V, E_3)$ and railway graph $G_2 = (V_2, E_2)$. Constraint 5.3 models over the existing lines (power graph) $G_1 = (V_1, E_1)$. Note that 5.2 is a non-linear constraint. The Constraint 5.4 refers to generation limit. The Constraints 5.5 and 5.6 enforce the power flow limits in transmission lines over the candidate graph $G_2 \cup G_3$ and existing graph G_1 respectively. Constraint 5.7 represents the limitation of the allowable built lines.

The above formulated non-linear programming is a difficult combinational problem, especially when the problem scale is large using exhaustive search to find a feasible solution is not efficient, not speak of the global optimization. We test this model on a 8-bus test case the solver Bonmin [Bon17], see Section 7.1.

5.2 Binary Mixed-integer linear Programming (BLP Model)

Because of the computational ability of a nonlinear solver, we propose a binary mixed-integer linear model(BILP) [RMGH02] for the TNEP problem. As analysed before, the reason of the non-linearities of a DC flow model arise from the multiplication of bus voltage angles and new lines decision variables, see eq. 5.3. In this disjunctive model the Constraint 5.3 is reformulated by applying the disjunctive binary circuit decision variable. Note that in this model we suppose to make a further restriction of our problem, which means we consider whether build the line between two buses (vertices) over the candidate graph, without parallel lines. However, in the next model, this restriction can be relaxed, which allows to build multiple lines along one edge. This linear model for TNEP problem can be formulated as follows:

Minimize:

$$h = (1 - \alpha) \sum_{(v,w) \in E_2 \cup E_3} C_{vw} Z_{vw} + \alpha \sum_{v \in V_1} g_v P_v$$

Subject to:

$$g_v - l_v = \sum_{(v,w) \in E} f_{vw} \qquad\qquad \forall v \in V \qquad\qquad (5.8)$$

$$|f_{vw} - b_{vw}(\theta_w - \theta_v)| \leq M_{vw}(1 - Z_{vw}) \qquad \forall(v,w) \in E_2 \cup E_3 \qquad (5.9)$$

$$f_{vw} - b_{vw}(\theta_w - \theta_v) = 0 \qquad\qquad \forall(v,w) \in E_1 \qquad\qquad (5.10)$$

$$0 \leq g_v \leq \bar{g}_v \qquad\qquad \forall v \in V_1 \qquad\qquad (5.11)$$

$$|f_{vw}| \leq Z_{vw} cap_{vw} \qquad\qquad \forall v, w \in E_2 \cup E_3 \qquad (5.12)$$

$$|f_{vw}| \leq cap_{vw} \qquad\qquad \forall v, w \in E_1 \qquad\qquad (5.13)$$

$Z_{vw} \in \{0, 1\}$, $f_{vw} \in \mathbb{R}$, $g_v \in \mathbb{R}$ and $\theta_v \in \mathbb{R}$

Constraint 5.8 represents node power balance. Constraints 5.9 and 5.10 represent Ohm's law (KVL) for candidate and existing lines, respectively. Constraint 5.11 is the generation capacity limit. Constraints 5.12 and 5.13 are transmission capacity limits for the candidate and existing lines, respectively.

The Constraint 5.2 is linearized to Constraint 5.9 by applying the binary decision variable Z_{vw} and big-M constant in place of integer variable n_{vw}. Note that $Z_{vw}, (v,w) \in E_2 \cup E_3$ is equal to 0, if no circuit is built between bus v and w. On the contrary, if Z_{vw} is equal to 1, that only means the circuit along edge vw is selected to transfer power. M_{vw} is the parameter used to avoid the appearance of non-linear constraints.

$$Z_{vw} = \begin{cases} 1 & \text{install line at edge } (v, w) \\ 0 & \text{install no line at edge } (v, w) \end{cases}$$

So if at the candidate edge vw a line is built, the right hand side of Constraint 5.9 must simply be 0, then the Constraint should become:

$$|f_{vw} - b_{vw}(\theta_w - \theta_v)| \leq 0 \quad \forall(v,w) \in E_2 \cup E_3 \qquad (5.14)$$

The new constraint 5.16 fit the KVL over the new line correctly. In other words, no matter how large M_{vw} is, the Constrains 5.9 can represent the same meaning as Constraint 5.2. In the opposite case, the candidate edge vw is not used to transfer power, we can get follow derivation.

$$f_{vw} = 0 \qquad\qquad\qquad (5.15)$$

Then the constraint becomes:

$$|b_{vw}(\theta_w - \theta_v)| \leq M_{vw} \quad \forall(v,w) \in E_2 \cup E_3 \qquad (5.16)$$

So if we want, that Constraint 5.16 is equivalent to the Constraint 5.2, the parameter "big-M" M_{vw} must be big enough to avoid creating implicit limits over voltage angle differences, which can lead to additional investments. However, too large value result in numerical instabilities in practical implementation. The numerical instabilities can lead to undesirable results: slow performance, wrong answers or inconsistent behavior. So the problem is, how to derive the minimum value for the parameter M_{vw}. It must not impose an implicit limit, and meanwhile ensure that the optimal solution can be found.

5.2.1 Determination of a suitable "big-M"

As mentioned above, the choosing of coefficients "big-M" influence how efficiently can the solution of the disjunctive model be found. For this reason, the upper bound for the "big-M" coefficients should be found for any circuits in the candidate network. In paper [MPS10] involved the theoretical derivation and practical computation for "big-M". The derivation can simply begin with two conditions.

1. If we build lines at edge $(v, w) \in E_2 \cup E_3$, then we have:

$$|(\theta_w - \theta_v)| \leq |f_{vw}/b_{vw}| \leq \frac{\bar{f_{vw}}}{|b_{vw}|} \quad \forall (v, w) \in E_2 \cup E_3 \tag{5.17}$$

2. If we do not build any lines at edge $(v, w) \in E_2 \cup E_3$, with the equation 5.16 we get the follow expression:

$$|(\theta_w - \theta_v)| \leq \frac{M_{vw}}{|b_{vw}|} \quad \forall (v, w) \in E_2 \cup E_3 \tag{5.18}$$

From Equation 5.17 we can find, that in order to get the minimum value of M_{vw} of edge (v, w), the maximum possible angel difference need to be computed. Before we compute the upper bound of the maximum possible angel difference, we need to definite the longest path LP_{vw} between bus v and bus w.

Definition 5.1. *Given graph $G = (V, E)$, the length len_{vw} of each edge (v, w) in G is defined by:*

$$len_{vw} = \frac{\bar{f_{vw}}}{|b_{vw}|} \tag{5.19}$$

Definition 5.2. *Given vertex v and vertex w in graph G, LP_{vw} is the maximum length of a simple path π_{vw} in G from vertex v to vertex w. It consists of several edges, along which the lines are built. If the simple path π_{vw} is present as a set of lines installed at edge e $\{e_{v1}, e_{12}...e_{nw}\}$, the length of the longest path LP_{vw} can be computed as follow:*

$$LP_{vw} = \sum_{(a,b) \in \pi_{vw}} len_{ab} \tag{5.20}$$

Now if we know the longest path π_{vw} between bus v and bus w, we can get the follow equation:

$$|\theta_v - \theta_w| \leq \sum_{(a,b) \in \pi_{vw}} |\theta_a - \theta_b| \leq \sum_{(a,b) \in \pi_{vw}} \bar{f_{ab}}/|b_{ab}| = LP_{vw} \tag{5.21}$$

However, we can not forecast along which line the circuits are built. So we simply find the longest path in G by assuming, that along each line in G we build circuit. In other words, the minimal value to obtain M_{vw} is how to find the longest path LP_{vw}. As we know the longest path problem itself is NP-complete [Gol10], which is very difficult to solve. So we use the following approach [MPS10] to compute the alternative upper bound M for M_{vw}:

Step 1: Put all buses $v \in V$ in the set S_v, and initiate longest path $M = 0$;

Step 2: Select one bus v_i from the set S_v, and list all edges $E_i \subseteq E$ connected to bus v_i. Then find the edge $(v_i, v_j) \in E_i$, which has the maximum cost $c_{v_i, v_j} = \bar{f_{v_i, v_j}}/|b_{v_i, v_j}|$.

Step 3: Set $M = M + c_{v_i, v_j}$.

Step 4: Remove bus v_i from the set S_v, and remove the edge (v_i, v_j) from the set E to avoid picking up twice the same edge.

Step 5: If set S_v is not empty, go to Step 2, else end.

Now we can get the final upper bound for M_{vw}, by setting:

$$M_{ij} = b_{ij} \cdot M \tag{5.22}$$

This formulation of the MILP can be approached by linear solver, like Gurobi. However, it is apparent that this model has one main disadvantage: along one candidate edge non-multiple lines can be installed.

$$N_{vw}^1 = 0\ or\ 1$$

$$N_{vw}^2 = 0\ or\ 1$$

$$N_{vw}^3 = 0\ or\ 1$$

$$N_{vw}^4 = 0\ or\ 1$$

Vertex v   *Vertex w*
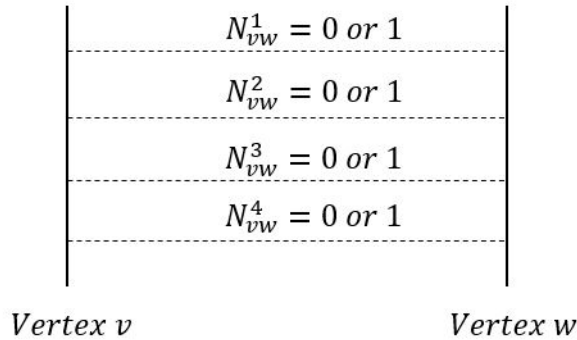
Figure 5.1: Install lines at edge (v, w)

5.3 Multiple Parallel Line Model (MPLP Model)

In the above section we have talked about the model, which can determine whether to install the circuit over the candidate network, but it only allows one circuit to be built along a candidate line. Since in construction of transmission grid may require build multiple parallel lines along one candidate line for reliability and economic purpose, in this section we propose an extended linear model, which allows to built multiple parallel circuit along one candidate line[HD17]. In this thesis we use binary variable n_{vw}^k to represent, whether the kth parallel line at edge (v, w) is built. Here is an example about building lines at edge $(v, w) \in E_2 \cup E_3$, see Fig. 5.1. In the following model we suppose to set $n_{vw}^- = 4$, and as described before variable n_{vw}^- means the maximum number of lines, so k can not bigger than n_{vw}^-. If we suppose to build 2 lines, any two of these binary variables($n_{vw}^1, n_{vw}^2, n_{vw}^3$ and n_{vw}^4) must be setted by 1, such as $n_{vw}^1 = 1$ and $n_{vw}^3 = 1$. The variable f_{vw}^k is the flow through the kth parallel line. So the reformulated model represents as follow.

In the objective function for the decimal disjunctive model, the unit investment cost of installing lines becomes:

$$\sum_{(v,w)\in E_2\cup E_3} \sum_{k=1}^{n_{vw}^-} C_{vw} n_{vw}^k$$

So the whole objective function is:

$$h = (1 - \alpha) \sum_{(v,w) \in E_2 \cup E_3} \sum_{k=1}^{n_{\overline{vw}}} C_{vw} n_{vw}^k + \alpha \sum_{v \in V_1} g_v P_v$$

Parameters f_{vw}^k, cap_{vw}^k and b_{vw}^k are the power flow, the capacity and susceptance of a single candidate line from vertex v to vertex w. the reformulation of the power balance 5.23, the KVL 5.24 and transmission capacity constraints 5.25 are represented as below:

$$g_v - l_v = \sum_{(v,w) \in E} \sum_{k=1}^{n_{\overline{vw}}} f_{vw}^k \qquad \qquad \forall v \in V \qquad (5.23)$$

$$|f_{vw}^k - b_{vw}^k(\theta_w - \theta_v)| \leq |M_{vw}(1 - n_{vw}^k)| \qquad \forall(v, w) \in E_2 \cup E_3 \quad 1 \leq n \leq n_{\overline{vw}} \qquad (5.24)$$

$$|f_{vw}^k| \leq n_{vw}^k cap_{vw}^k \qquad \qquad \forall(v, w) \in E_2 \cup E_3 \qquad (5.25)$$

Contrary to the disjunctive model, the decimal disjunctive model prepare parallel candidate lines to transfer power. At the same time this model can also support multiple types of transmission candidate lines to be invested in parallel along a certain candidate edge, since we assign independent variables for each line. However, this operation to support multiple parallel lines can greatly increase the size of the mixed integer optimization model, especially when the upper limit on number of parallel lines are large. The determination of the upper limit on number of parallel lines are also needed to consider, because we cannot install infinite lines along one candidate edge due to physical issues. It is discussed in Section 7.1.3.1, in oder to avoid missing the global optimal solution.

6. Heuristic Algorithm

In chapter 5 we have proposed the mathematical models(MINLP, BLP, MPLP) to install lines along candidate edges, in order to expand the existed transmission network. However, when we focus on building the mathematical models, it is difficult to quickly obtain the optimal solution by dealing with the large scale problem, even the non-linear model has been linearized, finding the optimal solution for the German network(German power grid and German railway network) would cost hours long. For this reason, in this chapter a heuristic algorithm is proposed to select the candidate edge from the candidate network to install lines, which is more efficient than mathematical optimization. As we know, when the power grid face of rising demand of energy, some transmission lines in the power grid must achieve restriction of transmission capacity, in this case, the power grid has reached its limit of transmitted power. So if we want to let the power grid transmit more power when it reaches the limit, the new lines should be installed. Under the condition of without considering the candidate network, our heuristic algorithm first discovers, how much energy the power grid can carry load, thus the transmission bottlenecks can be detected, then we can build the new lines along the edges, which are selected from the candidate network , to solve the transmission bottlenecks. So the process of the heuristic algorithm for TNEP can be summed up to three steps: 1. Design the transmission plan to solve the maximum load state problem for the transmission network. 2. Find the bottleneck edges of the designed the transmission network. 3. Find paths in the candidate network to substitute for the bottlenecks. We repeat the above mentioned three steps until the maximum load state of the transmission network can not be changed any more, even when more new paths in the candidate network are added in the transmission network. Through this chapter an simple example is given to illustrate the heuristic algorithm.

6.1 Find the maximum Load State of the Transmission Network

Before we discussed the method to solve the maximum load state problem for the transmission network, the definitions of the increased Power Factor ε and the maximum load state problem need to be given:

Definition 6.1 (The increased Power Factor ε). *Given a electrical network $N = ((V, E), b, cap, l, \bar{g})$ in DC model. The original loaded power is marked as $l_{original}$, then the current loaded power $l_{current}$ with an increased power factor ε can be represents as $l_{current} = \varepsilon * l_{original}$, ,and $\varepsilon \geq 1$.*

Definition 6.2 (Maximum Load State Problem). *Given a electrical network $N = ((V, E), b, cap, l, \bar{g})$ in DC model. Under the conditions to meet all electrical constraints, finding the maximum set value for the increased power factor ε, is the maximum load state problem.*

Now the question is how to find the increased power factor ε of the power grid N, according to the definition of the maximum load state, this problem can be solved by using following mathematical model(LP):

Maximize:

$$\varepsilon$$

Subject to:

$$g_v - \varepsilon * l_v = \sum_{(v,w)\in E} f_{vw} \qquad \forall v \in V_1 \qquad (6.1)$$

$$f_{vw} - b_{vw}(\theta_w - \theta_v) = 0 \qquad \forall (v, w) \in E_1 \qquad (6.2)$$

$$0 \leq g_v \leq \bar{g}_v \qquad \forall v \in V_1 \qquad (6.3)$$

$$|f_{vw}| \leq cap_{vw} \qquad \forall v, w \in E_1 \qquad (6.4)$$

$$\varepsilon \geq 1 \qquad (6.5)$$

$\varepsilon \in \mathbb{R}$, $f_{vw} \in \mathbb{R}$, $g_v \in \mathbb{R}^+$ and $\theta_v \in \mathbb{R}$

Because this model do not involved integer variables, it can be efficient to find the optimal solution. The objective function intuitively adopts the increased power factor ε and maximum it. Constraints 6.1 forces the power balance for each vertex in the power grid by adopting the increased power factor ε. Constraints 6.2 represents Ohm's law(KVL) for the transmission line. Constraints 6.3 is transmission capacity limits for the transmission lines. Constraint 6.4 is the generation capacity limit. Constraints 6.5 force the load increased.
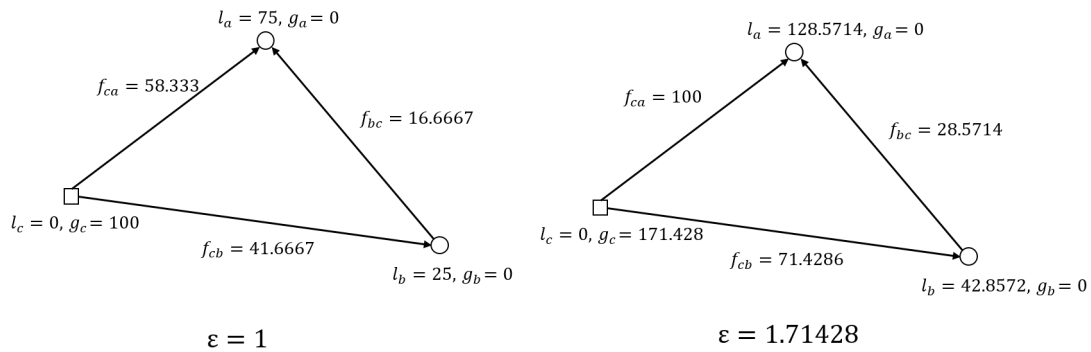
Figure 6.1: Example for the bottleneck detecting of the electrical network $N_1 = (G_1 = (V_1, E_1), b, cap, l, \bar{g})$, $V_1 = \{a, b, c\}$, $E_1 = \{(a, b), (a, c), (b, c)$, the susceptance (ohm/km) and capacity $(MW) : b_e = 1, cap_e = 100$, where $e \in E$. Left network is designed with original loaded power, and right network is designed with the maximum increased power factor $\varepsilon_{max} = 1.71428$. Boxes denote generator buses, circles denote load buses and filled dots denote transmission buses.

In Figure 6.1 the original power flow is displayed on the left, where the increased power factor $\varepsilon_{original} = 1$. Using the above mentioned linear model we find the maximum increased

power factor $\varepsilon_{max} = 1.71428$, and the corresponding power flow is shown in right network, from which we can find that the transmission lines (a, c) has reached the limit of the capacity. In other words, if we want to increase the loaded power by $\varepsilon > \varepsilon_{max}$, the current power grid needs to be expanded.

6.2 Bottlenecks Detecting

In this section, our goal is to find the bottlenecks in the power grid. In our heuristic algorithm the bottlenecks can be detected, when the power grid meet the maximum load state. So suppose the power grid is given as $N_1 = ((V_1, E_1), b, cap, l, \bar{g})$, and the generated power in the power grid is large enough for the increased demand of power. When we find the maximum setted value for the increased power factor ε of the power grid N_1, we could iterate all the transmission flow f_{vw} of each edge (v, w), where $(v, w) \in E_1$, to detect which transmission line is the bottleneck of the further increased power factor ε. Here the bottleneck line is defined as follow:

Definition 6.3 (The Bottleneck Transmission Line)**.** *Given an electrical network $N = ((V, E), b, cap, l, \bar{g})$ and its maximum set value for the increased power factor ε with a feasible power flow f, a small number δ is adopted, where $\delta \in \mathbb{R}^+$ and close to zero, for each transmission line $(v, w), (v, w) \in E_1$, if $|f_{vw}| > (1 - \delta)cap_{vw}$, then the transmission line (v, w) is a bottleneck in the network N.*

In figure 6.1 after iterating all the transmission lines in the electrical network, according to Definition 6.3 the transmission lines (a, c) can be the bottleneck edge of the power grid. Note that in our example $\delta = 0.05$.

6.3 Choosing Edges from the candidate Network and redesign the Power Grid

In this section the candidate network(the railway network and the connection network) $N_2 = ((V_2, E_2), b, cap, l, \bar{g})$, $N_3 = ((V_3, E_3), b, cap, l, \bar{g})$ are involved, in order to expand the bottlenecks in power grid N_1. Our idea is to find a shortest path in the candidate network which has the same endpoints as the bottleneck edge in power grid. Then along these edges in the path the new lines are installed and added in the power grid to expand the transmission network. Note that the shortest path consists of edges only from the candidate network. Here the shortest path for a bottleneck is given:

Definition 6.4 (The shortest Path for a Bottleneck)**.** *Given the electrical network $N_1 = ((V_1, E_1), b, cap, l, \bar{g})$ the railway network and the connection network) $N_2 = ((V_2, E_2), b, cap, l, \bar{g})$, $N_3 = ((V_3, E_3), b, cap, l, \bar{g})$ and the set of bottlenecks $E_{bottleneck} \subseteq E_1$, let $e_{u,w} = (u, w)$ be the edge incident to both u and w, for each edge $e_{u,w} \in E_{bottleneck}$, the shortest path for the bottleneck $e_{u,w}$ from u to w is the path $SP = (e_{v_1,v_2}, e_{v_2,v_3}, ...e_{v_i,v_{i+1}})$, where $v_1 = u$, $v_{i+1} = w$ $v_i \in V_2 \cup V_3$, $e_{v_i,v_{i+1}} \in E_2 \cup E_3$, which minimizes the $\sum_u^w W(e_{v_i,v_{i+1}})$ by adopting a weight function $W : E \rightarrow \mathbb{R}$, which represents the cost of installing new lines along each edge. Here we consider the investment of installing new lines as the weight.*
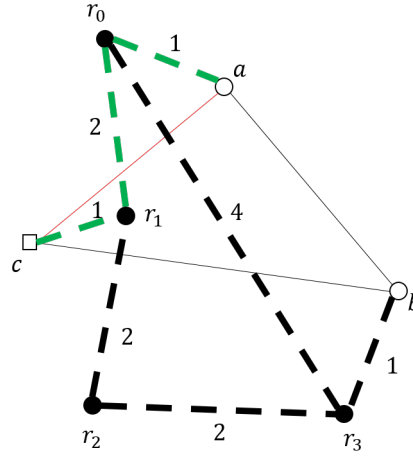
Figure 6.2: Example of choosing the candidate edges in the railway network N_2 and the connection network N_3 for the bottlenecks, where $G_2 = \{ r_0, r_1, r_2, r_3\}, \{(r_0, r_1), (r_1, r_2), (r_2, r_3), (r_3, r_0)\}\}$ and $G_3 = \{\{r_0, r_1, r_2, r_3, a, b, c\}, \{(r_0, a), (r_1, c), (r_3, b)\}\}$, the edges in the candidate network is drawn with the imaginary lines, the selected candidate edges are shown in green and the weight of each edge in G_2 and G_3 can be found in Figure.

In figure 6.2 shows the same power grid as in figure 6.1, As discussed in last section, the transmission lines (c, a) is the bottleneck edge for the current power grid. there are two paths from vertex c to vertex a in the candidate network: $SP_1 = (e_{c,r_1}, e_{r_1, r_0}, e_{r_0, a})$ and $SP_2 = (e_{c,r_1}, e_{r_1, r_2}, e_{r_2, r_3}, e_{r_3, r_0}, e_{r_0, a})$. The cost of installing new lines for SP_1 is obviously smaller than the path SP_2. So the shortest path for the edge (c, a) in candidate network is SP_1.

In this part the simple shortest path problem is solved by using Dijkstra algorithm[Dij59]. The Dijkstra algorithm thought is: given a graph $G = (V, E)$, the start vertex s and the target vertex t, where G is a weighted graph, the graph vertex set V is divided into two groups, the first group is the vertices of the known shortest path to the start vertex s(expressed in Q, there is only one vertex s in the set Q at the beginning), the second set consists of the vertices, which has not yet been found the shortest path to s, expressed by U, according to the shortest path length in order add the vertex v from the second set Q to S, until the target vertex t is added to the Q, the algorithm is over.

Now the power grid N_1 is expanded by adding the in candidate network new installed lines, see in Figure 6.3, however, since the power grid is changed, the new power flow in the power grid need to be computed, which can meet all the electrical constraints. Here we still use the LP model, which is described in section 6.2, to obtain the maximum load state.
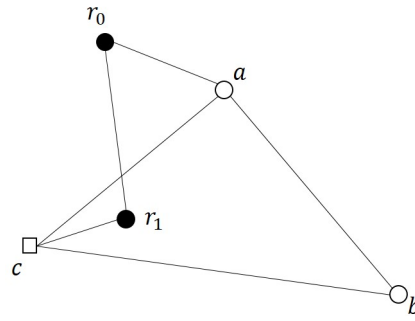
Figure 6.3: The expanded Power Grid by adding the in candidate network new installed lines

6.4 Conclusion

The heuristic algorithm repeat the above mentioned process, until the maximum load factor of the transmission network can not be increased any more, even new candidate edge from the railway network are added in the power grid. After 2 iterations the example in this chapter can obtain the maximum load factor of the transmission network as 2.857142 and the expanded transimission network is shown in figure 6.4.
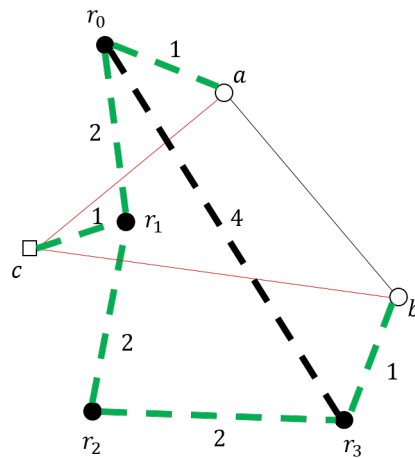
Figure 6.4: Example of the expanded power grid by using the heuristic algorithm, the edges in the candidate network are drawn with the imaginary lines, the selected candidate edges to build new lines are shown in green.

7. Evaluation

All results represented in this chapter was worked out from operation system Ubuntu 16.04. The computer has the follow configuration: CPU: Intel core (TM) i7-2600k 3.40Hz, Memory size: 8192 MB. The non-linear model is programed by python 2.7. The linear model was implemented in C++. The mathematical program for non-linear model were build and solved using the Bonmin [Bon17] and Casadi 3.2.0 package [And13]. The mathematical program for linear model were solved using the Gurobi 7.0 package [GO16].

There are several famous test cases, which are applied in different researches, such as Garver 6-bus, IEEE 24-bus, and Brazil South 46-bus. However, these test cases did not consider about, how to generate the candidate network. In our work we use railway network as the candidate network to expand the power grid, in other words, the test case should consist of two networks: existing power grid and the candidate network, which has totally different buses and lines with power grid. For this reason we design a simple and easily computed 8 buses test case, which has 4 power buses (vertices) power gird and 4 railway buses (vertices) railway network. The initial topology of 8 buses test case is shown in Fig. 7.1 and the branch data is displayed in table 7.3. Since this number of vertices in this test case is not large, the proposed non-linear model can be worked out in 5 minutes. Then we can use this 8 buses test case to make comparison between the non-linear and linear model. In the end the German power grid and the German railway network in practice are adopt for evaluating the multiple parallel line model.

Figure 7.1: The initial topology of 8 buses test case, boxes denote generator buses, circles denote load buses and filled dots denote transmission buses

7.1 Case Studies

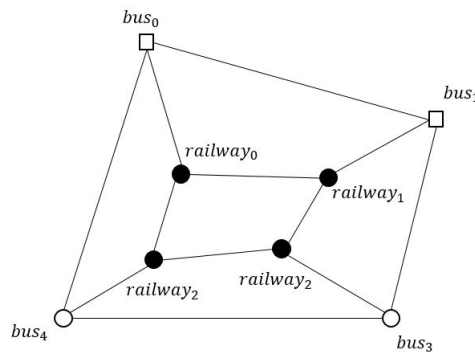The proposed models MINLP and MPLP in Section 5 has been applied to the 8 buses system and to the German realistic network. As shown in Fig. 7.1 the 8 buses system is a network with 4 buses, 4 railway vertices (buses), 4 existed edges and 8 candidate edges. The maximum number of the allowable installed lines for each candidate line in linear model is constrained no more than 5 lines to avoid long time calculation. However, the test cases with different maximum number of the allowable installed lines is adopted to discuss, how it influence the optimal solution when it changes. The default electrical parameters of the 8 buses system is given in Section 7.1. The cost of installing new lines is really hard to estimate, it is influenced by many factors, such as the environmental impact, the installed line types and the construction ways. So in our works the cost of installing new lines is simply set by $400 Euro/km$. The branch data of the 8 buses system is shown in Section 7.3. The generation and load data for 8 buses system is shown in Section 7.2.

Table 7.1: The default parameters

parameter	value	
investment of installing new line in the candidate network	400	$Euro/km$
impedance	0	ohm/km
reactance	1	ohm/km
capacity	50	MW

Table 7.2: The generation and load data for 8-buses system

Bus	Maximum Generation(MW)	Load(MW)	Energy Source Type
0	500	20	Gas
1	200	100	Wind Onshore
2	0	90	-
3	0	20	-

Table 7.3: The branch data of 8-buses system

line	impedance (ohm/km)	reactance (ohm/km)	length (km)	Capacity (MW)
$bus_0 - bus_1$	0	1	4	90
$bus_1 - bus_2$	0	1	4	90
$bus_2 - bus_3$	0	1	4	90
$bus_3 - bus_0$	0	1	4	90
$railway_0 - railway_1$	0	1	2	50
$railway_1 - railway_2$	0	1	2	50
$railway_2 - railway_3$	0	1	2	50
$railway_3 - railway_0$	0	1	2	50
$bus_0 - railway_0$	0	1	1	50
$bus_1 - railway_1$	0	1	1	50
$bus_2 - railway_2$	0	1	1	50
$bus_3 - railway_3$	0	1	1	50

For the German realistic network, which include the German power grid, the railway network and the connection network, the assumption of electrical parameters in the railway network and the connection network are given in Table 7.4. The impedance, the reactance and the capacity of the new installed line are adopted by the common value, which appears in the power grid data.

Table 7.4: The default setting of the candidate network

Parameter	Value
impedance (ohm/km)	0.025
reactance (ohm/km)	0.25
investment of installing new line in the candidate network (e/km)	400
flow capacity (MW)	800

The motivation of our work is using the candidate network to expand the power grid, which can meet the increasing demand of loaded power in the future. For this reason, our test cases will focus on, how our work performs during the increasing loaded power in the power grid. Here the symbol ε is adopted, where $\varepsilon > 1$, which means the increased factor. Let the original loaded power of each bus(vertex) in the power grid is presented as $l_{original}$, then in our test cases the loaded power must be $l_{test} = \varepsilon * l_{original}$. For example, in the 8 buses system the bus_1 load $100MW$ power, if in a test case the increased factor is set by 2, then the loaded power of bus_1 is $200MW$.

7.1.1 Comparison between MINLP and MPLP in 8-buses System

In this section we fixed the increased factor $\varepsilon = 2$ for 8-buses, since when $\varepsilon = 2$, the existed power grid cannot meet the demand of the power transmission, in other words, the new lines are needed to be installed to expand the existed power grid. Here we do not apply the MINLP model in the German realistic network, which can take intolerable time to find the optimal solution. The solutions of both models MINLP and MPLP are shown in Table 7.5. We can find, that the minimized cost of investment by NLP is larger than the one by MPLP. Since the current algorithm for solving non-linear model cannot efficiently find global optimization and the non-linear solver is not stable for numeric issues, the solution of MINLP easily fall into local optimization. In the comparison test case, the maximum allowable installed circuits can be large as 5, because the scale of the test network is not huge. Both tests are finished in one minute. In conclusion, the MPLP performs even better than the MINLP in obtaining the optimal solution.

Table 7.5: Comparison between MINLP and MPLP

NLP	Solution	MPLP	Solution
Best	7600	Best	6400
$n_{0-railway1}$	6	$n_{0-railway1}$	4
$n_{1-railway2}$	0	$n_{1-railway2}$	2
$n_{2-railway3}$	1	$n_{2-railway3}$	0
$n_{3-railway4}$	4	$n_{31-railway4}$	2
$n_{railway1-railway2}$	0	$n_{railway1-railway2}$	2
$n_{railway2-railway3}$	0	$n_{railway2-railway3}$	0
$n_{railway3-railway4}$	1	$n_{railway3-railway4}$	0
$n_{railway4-railway1}$	3	$n_{railway4-railway1}$	2

7.1.2 Increased Factor ε

In this section the MINLP model is adopted into this test, how the candidate network performs when the increased factor ε changes. This test involve the 8-buses system and the German realistic network, for the 8-buses system the optimal solution can be given quickly, however, for the German realistic network the test set the time limit in 3000 seconds. Even the global optimal solution can not be find in the short time, however, the keeping decreasing gap can be displayed.

7.1.2.1 Test in 8-buses System

The test begin with setting up $\varepsilon = 1.6$, when the current power grid need to be expanded to supply more power transmission, and ended with setting up $\varepsilon = 2.8$, when the network can not transmit any more power even all edge the candidate network is applied to install new lines or the sum of load is lager than the sum of generation. The solution is showed as following Table 7.6. All the tests finished in 1 second. Note that the maximum allowable installed lines n_{vw}^- is fixed by 5. When ε is set to 1.6, the current power grid can support the work to meet the demand of loaded power, then there are no new installed lines in the candidate network. Then with the increasing loaded power, in this situation, more and more new lines along railway edges and connections are installed. And we can find that, when ε is set to 2.8, $n_{bus_0-railway_0}$ and $n_{bus_2-railway_2}$ have reached the upper bound of installing new lines, which means the hole network included the power grid and the railway network is already saturated. And in our test, no feasible solution can be found, when $\varepsilon > 2.8$.

Table 7.6: Test case in 8-buses system by the different increased factor ε

$\varepsilon = 1.6$	$\varepsilon = 2.0$	$\varepsilon = 2.4$	$\varepsilon = 2.8$
Best 0	Best 6400	Best 9600	Best 12000
$n_{bus_0-railway_0} = 0$	$n_{bus_0-railway_0} = 4$	$n_{bus_0-railway_0} = 4$	$n_{bus_0-railway_0} = 5$
$n_{bus_1-railway_1} = 0$	$n_{bus_1-railway_1} = 2$	$n_{bus_1-railway_1} = 0$	$n_{bus_1-railway_1} = 0$
$n_{bus_2-railway_2} = 0$	$n_{bus_2-railway_2} = 0$	$n_{bus_2-railway_2} = 4$	$n_{bus_2-railway_2} = 5$
$n_{bus_3-railway_3} = 0$	$n_{bus_3-railway_3} = 2$	$n_{bus_3-railway_3} = 0$	$n_{bus_3-railway_3} = 0$
$n_{railway_0-railway_1} = 0$	$n_{railway_0-railway_1} = 2$	$n_{railway_0-railway_1} = 3$	$n_{railway_0-railway_1} = 2$
$n_{railway_1-railway_2} = 0$	$n_{railway_1-railway_2} = 0$	$n_{railway_1-railway_2} = 3$	$n_{railway_1-railway_2} = 2$
$n_{railway_2-railway_3} = 0$	$n_{railway_2-railway_3} = 0$	$n_{railway_2-railway_3} = 1$	$n_{railway_2-railway_3} = 3$
$n_{railway_3-railway_0} = 0$	$n_{railway_3-railway_0} = 2$	$n_{railway_3-railway_0} = 1$	$n_{railway_3-railway_0} = 3$

7.1.2.2 Test in the German realistic Network

Since the scale of the German network is large, which has 586 buses, 9011 railway vertices and 8702 candidate edges, to find a global optimal solution in the short time is impossible, here the time limit is set by 3000 seconds. The German power grid can support the power transmission, when $\varepsilon = 1.0174$, So the test begin with setting up $\varepsilon = 1.2$, ended with setting up $\varepsilon = 3.0$, the pace of the growth of the increased factor is 0.24. The end condition is set by 3.0, because when $\varepsilon = 3.2$, there is no feasible solution can be found in the limited running time. The maximum number of allowable installing lines n_{vw}^- is fixed by 3. The gab shown in results is defined as follow formula: $gap = |ObjBound - ObjVal|/|ObjVal|$, which represents, how close the current optimal solution to the global optimal solution.

The fig. 7.2 shows the results of the MPLP model, which is implemented with the German network. Even the increased factor $\varepsilon = 3.2$, the feasible solution can also be found in 3000 seconds, however, finding the global optimal solution through MPLP model still needs

hours long for the German realistic network, when the increased factor is large. Compared to the NLP model, which takes days long no feasible solution can be accessed, this MPLP model has done well in the time cost.

In fig. 7.3 the first time of finding a feasible solution with the different increased factor is represented. The gab at the end of the running time ($3000s$) with the different increased factor is shown in fig.7.4. From these figures we can estimate that the time of finding a feasible planning solution and the gab by MPLP model has no clear relation with the with the increasing loaded power in the power grid. However, from the most test cases, we can find that the gaps decrease rapidly in the first 2000 seconds, then the downward trend gradually eased. So the gaps shown in 3000 seconds can be a reference optimal solution.

7.1.3 Increase the maximum Number of allowable installing Lines n_{vw}^-

The above test cases only consider fixing the maximum number of allowable installing lines for each candidate edge in the railway network and connection network. Because, the upper bound of the new installed lines determine the scale of our model, in other words, if n_{vw}^- is set to a high value, which can cost long time to find a feasible solution. However, if the the maximum number of allowable built circuits is set to a small number, the optimal solution or feasible solution cannot be obtained. See example in Fig. 7.5. Before we illustrate the example, the definition of the saturated edge need to be proposed first.

Definition 7.1 (The saturated edge)**.** *Given the maximum number of allowable built lines n_{vw}^- and the electrical network N, in an optimal planning the candidate edge $e(v, w) \in E$ is saturated when the number of the installed lines n_{vw} equals n_{vw}^-. In other words, all candidate lines along the edge e have been invested to transfer power.*

In Fig. 7.5 suppose the maximum number of allowable installing lines $n_{vw}^- = 2$, the bus a generate $10MW$, the bus b and bus c load $2MW$ and $6MW$, respectively. The capacity of each line is $2MW$. Here without consideration of the constraint Ohm's Law, the optimal solution of the power transmission planning for the network N_1 must be: $n_{ab} = 2$, $n_{ac} = 2$, $N_{bc} = 1$, in other words, at least five lines need to be installed for meeting the power balance constraints. From the above definition of the saturated edge, we can find, the line $e_1(a, b)$ and $e_2(a, c)$ are saturated. If n_{vw}^- is increased by 1, a better solution can be found: $n_{ab} = 1$, $n_{ac} = 3$, $n_{bc} = 0$, which just need to install four lines. Hence, it is important to find a proper n_{vw}^- to find the optimal solution.

7.1.3.1 Optimal the maximum Number of allowable Installing Lines n_{vw}^-

However, before we compute out the solutions during optimal planning, we cannot ensure the maximum number n_{vw}^- of allowable built lines is large enough to find the global optimal solution. So the follow steps is used to find a largest enough n_{vw}^-:

Step 1: Initiate n_{vw}^- by a small number, for example $n_{vw}^- = 1$.

Step 2: Find the optimal solution for the electrical network after implement the mathematical model.

Step 3: Check the solution, if no feasible solution or there is any line $e(v, w) \in E$ saturated, set $n_{vw}^- = n_{vw}^- + \varphi$, where φ is the factor every run you wan to improve n_{vw}^-, then go to Step 2. If no line in the electrical network is saturated, then we find the optimal n_{vw}^-, this process end. Remark that large φ can accelerate the speed to find the proper n_{vw}^-, however, it might cause a computational issue for large scale problem in each run.
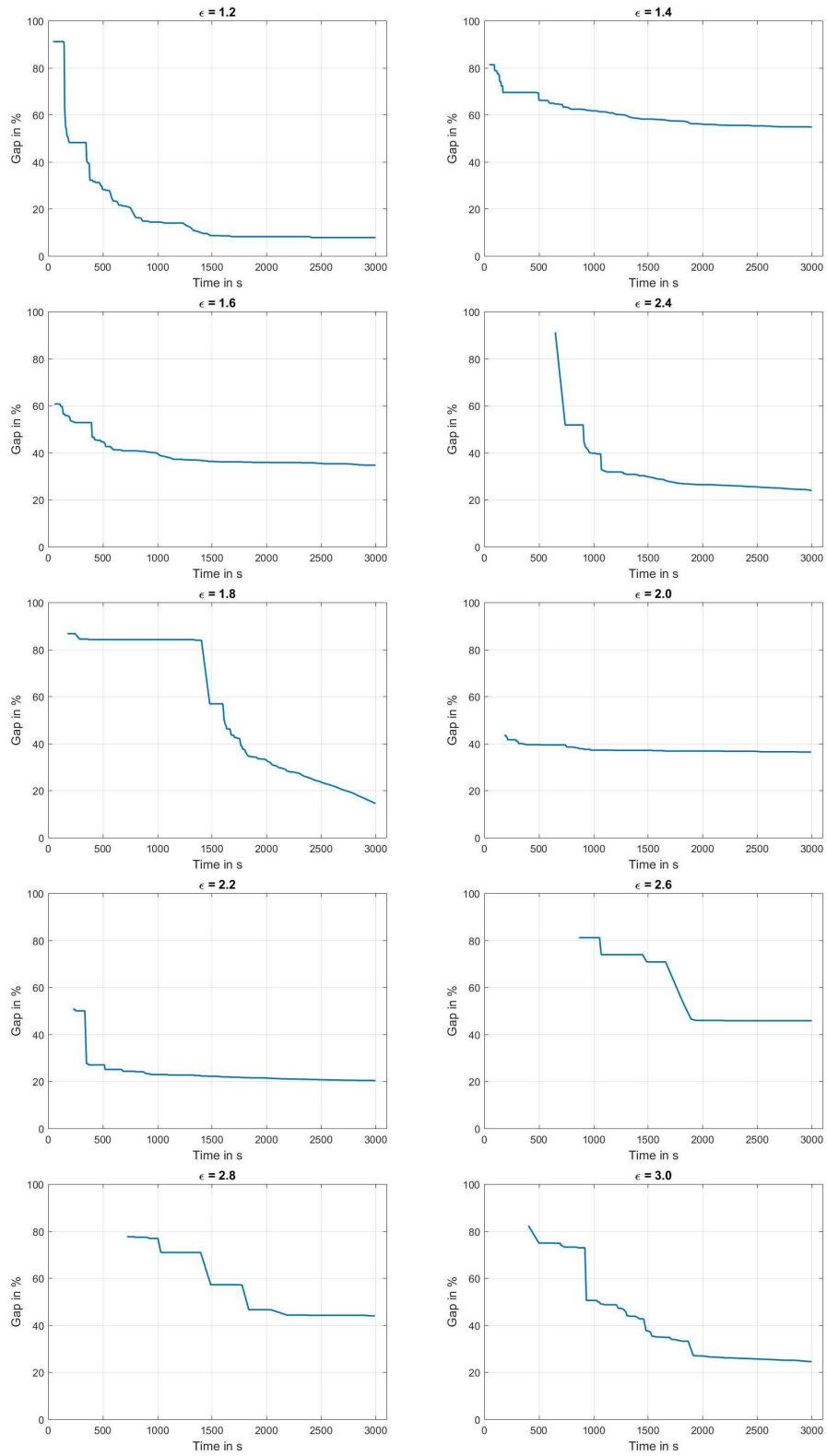
Figure 7.2: The variations of the gab, where $n_{vw}^- = 3$, the pace of the growth of the increased factor is 0.2
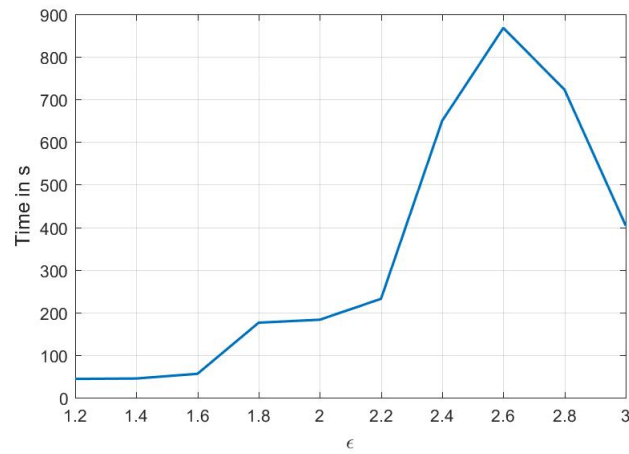
Figure 7.3: The first time of finding a feasible solution with the different increased factor, where $n_{vw}^- = 3$.
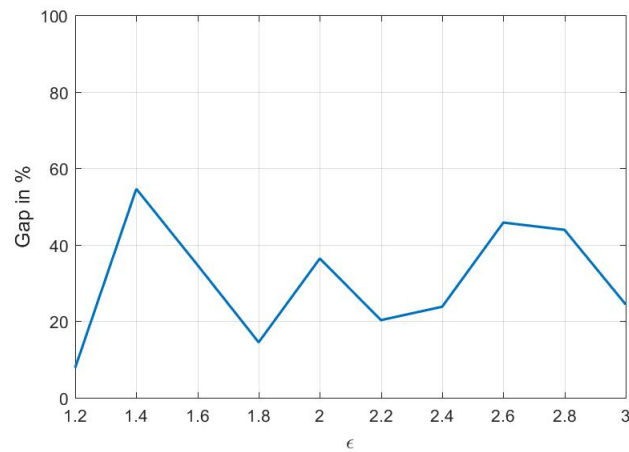

Figure 7.4: The gap at the end of the running time $(3000s)$ with the different increased factor, where $\bar{N}_{vw} = 3$.
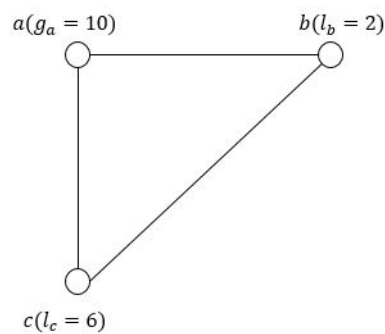

Figure 7.5: Electrical network N_1 for limitation test of maximum number of allowable built circuits.

7.1.3.2 Test in 8-buses System

In this test we set n_{vw}^- by different numbers in the 8-buses system, it begins with 1, and is increased by 1 till 6. The increased factor of loaded power ε here must be fixed on a certain number, here we select 2, in this case the railway network is used to expand the power grid and the feasible solution of the test can be access in one hour, then we can exclude the influence of the increased factor ε in our test, and focus on the the influence of the maximum Number of allowable Installing Lines. The results of the test cases in 8-buses System are given in Table 7.7. All the tests are finished in 1 second. From the results it is clear that, with the increasing n_{vw}^-, the optimal solution going to decrease and stay the same value in the end. So n_{vw}^-, which is set by 3, is already enough, when the the increased factor of loaded power $\varepsilon = 2$.

Table 7.7: Increasing the maximum allowable building lines n_{vw}^- in the 8-buses system, here infeasible means there are no feasible solution for n_{vw}^-

$n_{vw}^- = 1$	$n_{vw}^- = 2$	$n_{vw}^- = 3$	$n_{vw}^- = 4$	$n_{vw}^- = 5$	$n_{vw}^- = 6$
infeasible	infeasible	Best 7200	Best 6400	Best 6400	Best 6400

7.1.3.3 Test in the German realistic Network

The same test has also been done for the German realistic network, n_{vw}^- begins with 1, and is increased by 1 till 4, and the increased factor of loaded power ε is fixed by 2. The results of the test cases in the German network are given in table 7.8. The limitation of the time is set by 3000 seconds, so the optimal solution in the test are not guaranteed to be global optimal, but from the result it can be concluded that under 3000 seconds $n_{vw}^- = 2$ is large enough to avoid the influence of the number of the maximum allowable building lines, in other words, even we increase n_{vw}^-, the invest of installing new lines will not be decrease, only the cost of the calculation time is increased, since the scale of the mathematical model grows up.

Table 7.8: Increasing the maximum allowable building lines n_{vw}^- in the German network, here infeasible means there are no feasible solution for n_{vw}^-, the cost in table represents, how much money is invested in the solution, the time represents, how much time is spent on finding the first feasible solution.

-	$n_{vw}^- = 1$	$n_{vw}^- = 2$	$n_{vw}^- = 3$	$n_{vw}^- = 4$
cost (*Euro*)	Best 36724	Best 33378	Best 33378	Best 33378
time (*s*)	35	95	201	936

7.1.4 Comparison between MPLP and the heuristic Algorithm

Since in the heuristic algorithm we only consider installing one line at each edge in the candidate network, so in this test we set $n_{vw}^- = 1$, when we compare the MPLP and the Heuristic algorithm. Here for the German realistic network we compare the maximum load can be increased, the running time and the investment of finding optimal solution by fixing the same load. The result is shown in Table 7.9. Based on the experimental data, we can get the following conclusions: First, comparing to the heuristic algorithm, the MPLP model can fit higher energy demand with the same candidate network, since when we find the short path in the candidate network for the bottleneck edges, it might happen that for the same bottleneck they share the same edge in the short path. However, in

MPLP model, the solution not only focus on the cost, but also focus the feasible solution. Second, when setting the same increased factor ε, the heuristic model shows significant higher computational efficiency than the MPLP model. At last, the investment to expand the power grid usually higher for the heuristic model than the MPLP model.

Table 7.9: Comparison between MPLP and the Heuristic Algorithm

-	the Heuristic Algorithm	MPLP
the maximum increased loaded factor ε_{max}	1.7649	2.3675
time (s), $\varepsilon = 1.7649$	15	3000 $(gap = 6.4\%)$
cost$(Euro)$, $\varepsilon = 1.7649$	657690	Best 52126

7.1.5 Introduce the operating Cost of the generating Units

In the section, we want to observe the effects of the weighting factor α to the MPLP model, with α is used to adjust the objective function, by varying the ratio of the cost of new installed lines and the operating cost of the generating units. To proceed the experiment, we start by setting the factor α as 0, and increasing with a step of 0.2, till $\alpha = 1$. The starting point $\alpha = 0$ means we only consider the cost of installing new lines, while $\alpha = 0$ represents the cost of the generating units. In this process, we observed the number of the new installed lines in the candidate network with the increasing of the weighting factor. The result is shown in Table 7.10. Note that the limit of the running time is set 3000 seconds.

Table 7.10: Changing the weighting factor α

α	0	0.2	0.4	0.6	0.8	0.95
the number of new installed lines	44	66	85	100	125	354

From the table, the number of new installed lines in the candidate network shows positive relativity to the weighting factor. Since the cost of installing a new line is cheaper than the operating cost of the generating units, when increase the weighting factor, we tend to install more lines in the candidate network to fulfill the transport of the energy from the cheap generating units. And in this test, if we α to 1, the feasible solution can not be accessed in 3000 seconds, because in this situation installing new lines is costless, the model will the hole transmission network redesign.

8. Conclusion

This thesis introduces a study framework and methodology of solving TNEP by using the railway network. The framework has been implemented on the German realistic networks to facilitate different types of future demand of loaded power. We adopt three mathematical modeling approaches: MINLP, BLP and MPLP and innovative transmission candidate selection algorithm (the Heuristic algorithm) for TNEP problem. The mathematical modeling approaches, which are introduced in Chapter 5, has been applied to model the German transmission expansion optimization problem. The heuristic algorithm also been included for this case, resulting in finding the feasible solution more efficiently, but not optimal than the mathematical modeling approaches. Based on the study results and analysis in Chapter 7, we summarize the following observations and conclusions:

First, we compared the non-linear mathematic model (MINLP) with the linear mathematic model (MPLP) by applying on the 8-buses system. Both of the model can support installing parallel lines along one edge in the candidate networks (the railway network and the connection network). the MINLP in contrast to MPLP is much easier to build and understand. However the MPLP has a higher computational performance (within seconds) with a global optimal solution than the MINLP (within minutes), which may stop when reaching a local optimal solution.

Second, we continue apply the MPLP on the German realistic network (German power gird, German railway network and Connection network) to study the computational performance for the TNEP problem with different increased power factor, and the gap between the current optimal solution to the global optimal solution is computed from Gurobi. When the consumed power increased, the relative gap decreased rapidly in the first 2000 seconds, then the downward trend gradually eased. In the most of our experiments, the gap can minimize to lower than 30% within 3000 seconds.

Third, we proposed new method using the heuristic algorithm to solve the TNEP problem with German realistic network. By comparing with the MPLP model, the heuristic algorithm shows significant higher computational efficiency than the MPLP model. When the increased power factor is set to 1.7649, the heuristic algorithm can receive the optimal solution within 15 seconds, while it takes the MPLP model 3000 seconds to reach a minimum gap at 6.4%. However the cost for the optimal solution is around 10 times more expensive for the heuristic algorithm than the MPLP model.

At last, we studied the influence of the German realistic network by introducing the operating cost of the generating units into the MPLP model. When the weighting factor α

increases, in other word, when the operating cost of the generating units becomes more important in the objective function of the MPLP model, the optimal solution computed from Gurobi tends to install more lines in the candidate network to fulfill the transport of the energy from the cheap generating units.

8.1 Future Work

Further work can be proposed as followed: First, in this thesis, there is a challenge to apply the solution into the practical situations as our model is based on the DC model with simplifications by not considering the reactive power and the power losses. These two AC operations can be further studied by reinforce the resulted plan from our model. Second, for both mathematic model and our heuristic algorithm, we only consider installing new lines in the candidate network. But the existing power grid is also playing a important role in the model and can be considered to expand the transmission network by installed new parallel lines. Third, as the cost of installing lines in railway network is a influencing factor to evaluate our model, we can build a separate model to estimate the total cost. Last but not least, our heuristic algorithm can be improved by installing parallel lines in the candidate network.

Bibliography

[AG18]   Joel Andersson and Joris Gillis. Homepage — casadi, 2018. [Online; accessed 3-April-2018].

[AMC03]   Natalia Alguacil, Alexis L Motto, and Antonio J Conejo. Transmission expansion planning: A mixed-integer lp approach. *IEEE Transactions on Power Systems*, 18(3):1070–1077, 2003.

[And13]   Joel Andersson. *A General-Purpose Software Framework for Dynamic Optimization.* PhD thesis, Arenberg Doctoral School, KU Leuven, Department of Electrical Engineering (ESAT/SCD) and Optimization in Engineering Center, Kasteelpark Arenberg 10, 3001-Heverlee, Belgium, October 2013.

[Bal12]   R Balakrishnan. *A Textbook of Graph Theory.* Springer-Verlag New York, 2012.

[Bie09]   Daniel Bienstock. *Mathematical Programming Computation-A Publication of the Mathematical Optimization Society.* Springer, 2009.

[Bon17]   Bonmin. Basic open-source nonlinear mixed integer programming, 2017. [Online; accessed 5-February-2018].

[BPG01]   Silvio Binato, Mário Veiga F Pereira, and Sérgio Granville. A new benders decomposition approach to solve power transmission network design problems. *IEEE Transactions on Power Systems*, 16(2):235–240, 2001.

[Cad09]   F. Cadini. Optimal expansion of an existing electrical power transmission network by multi-objective genetic algorithms. *Reliability Engineering and System Safety*, 2009.

[Dij59]   E. W. Dijkstra. A note on two problems in connexion with graphs. 1959.

[EH72]   O. I. Elgerd and H. H. Happ. Electric energy systems theory: An introduction. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-2(2):296–297, April 1972.

[ene05]   Nuclear energy agency/international energy agency/organization for economic cooperation and development projected costs of generating electricity. 2005.

[Gar85]   L. Garver. Transmission network estimation using linear programming. *Power Apparatus and Systems, IEEE Transactions*, 1985.

[GO16]   Inc. Gurobi Optimization. Gurobi optimizer reference manual, 2016.

[Gol10]   Oded Goldreich. *P, NP, and NP-Completeness: The Basics of Computational Complexity.* Weizmann Institute of Science, 2010.

[HD17]   S. Huang and V. Dinavahi. A branch-and-cut benders decomposition algorithm for transmission expansion planning. *IEEE Systems Journal*, pages 1–11, 2017.

[HHK13a] R. Hemmati, R. A. Hooshmand, and A. Khodabakhshian. Comprehensive review of generation and transmission expansion planning. *IET Generation, Transmission Distribution*, 7(9):955–964, Sept 2013.

[HHK13b] Reza Hemmati, Rahmat-Allah Hooshmand, and Amin Khodabakhshian. State-of-the-art of transmission expansion planning: Comprehensive review. *Renewable and Sustainable Energy Reviews*, 23:312–319, 2013.

[LSRH09] Armando M. Leite da Silva Leandro S. Rezende and Leonardo M. Honorio. Artificial immune systems and differential evolution based approaches applied to multi-stage transmission expansion planning. *Intelligent System Applications to Power Systems, 2009. ISAP '09. 15th International Conference on*, 2009.

[Med18] Dr. Wided Medjroubi. Scigrid, 2018. [Online; accessed 29-March-2018].

[MPS10] Moulin, Michael Poss, and Sagastizábal. Transmission expansion planning with re-design. *C. Energy Syst (2010)*, 2010.

[osm18] Index of mirrors openstreetmap.org. openstreetmap.org planet data., 2018. [Online; accessed 3-April-2018].

[RBH14] Rodrigo R. E. Gumucio Russell Bent, Carleton Coffrin and Pascal Van Hentenryck. Transmission network expansion planning: Bridging the gap between ac heuristics and dc approximations. *Power Systems Computation Conference (PSCC)*, 2014.

[RGM95] R Romero, RA Gallego, and A Monticelli. Transmission system expansion planning by simulated annealing. In *Power Industry Computer Application Conference, 1995. Conference Proceedings., 1995 IEEE*, pages 278–283. IEEE, 1995.

[RMGH02] R Romero, A Monticelli, Ae Garcia, and S Haffner. Test systems and mathematical models for transmission network expansion planning. *IEE Proceedings-Generation, Transmission and Distribution*, 149(1):27–36, 2002.

[SBdA01] G.C. de Oliveira S. Binato and J.L. de Araujo. A greedy randomized adaptive search procedure for transmission expansion planning. *IEEE Transactions on Power Systems*, 2001.

[sci12] Deutsche energy-agentur (dena). ausbau- und innovationsbedarf der stromverteilnetze in deutschland bis 2030. *Technical report*, 2012.

[sou18] Studie stromgestehungskosten erneuerbare energien. *Fraunhofer ISE*, March 2018.

[VGS85] R Villasana, LL Garver, and SJ Salon. Transmission network planning using linear programming. *IEEE transactions on power apparatus and systems*, (2):349–356, 1985.

[Wik14] OpenStreetMap Wiki. Main page — openstreetmap wiki,, 2014. [Online; accessed 5-February-2018].

[Wik17] OpenStreetMap Wiki. Key:railway — openstreetmap wiki,, 2017. [Online; accessed 5-February-2018].

[Wik18] OpenStreetMap Wiki. Osmosis — openstreetmap wiki,, 2018. [Online; accessed 30-April-2018].

[YH89] H.K. Youssef and R. Hackam. New transmission planning model. *IEEE Transactions on Power Systems*, 1989.