

BESCHRÄNKUNG VON CLUSTERN
EINE BERECHNUNGSGRUNDLAGE DES ROUTINGMODELLS

by

Dietmar Pichler

Bachelorarbeit im Bereich
Telematik / Netzwerktechnik
zur Erlangung des akademischen Grades

Bachelor of Science

Studiengang Netzwerktechnik / Telematik

FH Kärnten

January 2011

Fachhochschulstudiengang
Telematik / Netzwerktechnik



B A C H E L O R A R B E I T

Beschränkung von Clustern Eine Berechnungsgrundlage des Routingmodells

zur Erlangung des akademischen Grades
Bachelor of Science

Autor: Dietmar Pichler
Matrikelnummer: 0810286040

Erstbetreuer: FH-Prof. Dipl.-Ing. Dr. Peter Bachhiesl

Zweitbetreuer:

Tag der Abgabe:

BESCHRÄNKUNG VON CLUSTERN
EINE BERECHNUNGSGRUNDLAGE DES ROUTINGMODELLS

by

Dietmar Pichler

Bachelorarbeit im Bereich
Telematik / Netzwerktechnik
zur Erlangung des akademischen Grades

Bachelor of Science

Studiengang Netzwerktechnik / Telematik

FH Kärnten

January 2011

Name: Dietmar Pichler
Matrikelnummer: 0810286040

Geburtsdatum: 17.07.1981

Adresse: Dr. Franz Palla Gasse 17, 9020 Klagenfurt

E i d e s s t a t t l i c h e E r k l ä r u n g

Mit dieser Erklärung versichere ich die erstmalig vorgelegte Bachelorarbeit selbständig erstellt zu haben.

Es wurden nur die von mir angegebenen Hilfsmittel und Quellen verwendet.

Datum: 04.04.2011
Ort: Klagenfurt

Unterschrift:

ABSTRACT

BESCHRÄNKUNG VON CLUSTERN

EINE BERECHNUNGSGRUNDLAGE DES ROUTINGMODELLS

In dieser Arbeit wird ein Modell beschrieben, welches ein bestehendes und bereits implementiertes Programm (RTR_R2008a) zu einer verbesserten Performance führen soll. Das Programm wurde von FH-Prof. Dipl.-Ing. Dr. Peter Bachhiesl geschrieben und berechnet eine Netzwerkinfrastruktur von Grund auf neu. Die Netzwerkinfrastruktur bildet das Telekommunikationsnetz der Telekom Austria ab und ermittelt die Längen der Kabel bezüglich der Trassenführung. Diese Längen, werden Versorgungslängen genannt.

Es wird angestrebt, dass bestehende Programm hinsichtlich der maximalen Versorgungslängen, bereits in einem frühen Stadium der Modellierung einzugrenzen. Somit kann in weiterer Folge sichergestellt werden, dass keine unzulässigen Versorgungslängen im Modell weiter berechnet werden. Werden diese Längen nicht eingeschränkt, kann es passieren, dass im letzten Schritt der Modellierung zu einer Fehlkalkulation kommt.

Daraus resultierend, muss der gesamte Vorgang der Modellierung nochmals durchlaufen werden. Es kann nicht ausgeschlossen werden, dass wiederum ein Fehler auftritt. Dieses Ergebnis soll erreicht werden, durch eine Implementierung eines Testbeds. Darum wird das Ziel verfolgt, dass jeder Modellierungsvorgang exakte Resultate liefert.

Schlüsselwörter: Versorgungslängen, Kanten, Knoten, Steiner Baum, minimal spannender Baum, Algorithmen

This thesis describes a model which has already been implemented by FH-Prof. Dipl.-Ing. Dr. Peter Bachhiesl. The program simulates a new data network which is based on an existing network infrastructure from the Telekom Austria. However a few weaknesses are included. The main aim of this employment is to rise the performance by purging the weaknesses. In fact, it is to find the maximum length from the distribution center to the customer. The goal is to isolate higher-than-average lengths in an early part of the modeling process. This to make sure that every calculation is able to provide correct results. As a result of not limiting these lengths it could occur that the whole procedure has to be continued. To reach this outcome, a testbed has to be implemented. Therefore, every single modeling process will deliver exact findings.

keywords: distribution lengths, edges, nodes, Steiner tree, minimal spanning tree, algorithms

Inhaltsverzeichnis

Inhaltsverzeichnis	ii
Abbildungsverzeichnis	iii
1 Einleitung	1
1.1 Allgemeines	1
1.2 Umfeld	1
1.3 Prozessablauf	2
1.4 Aufgabenstellung	3
2 Methoden zur Abschätzung der maximalen Versorgungslänge	4
2.1 Modelle	4
2.1.1 Gesamtmodell-Input Netzwerkgraph	4
2.1.2 Normierte Geodaten	5
2.1.3 Cluster-Modell	6
2.1.4 Routing-Modell	7
2.2 Mathematisches Konzept	8
2.2.1 Baum	8
2.2.2 Steiner Baum	9
2.2.3 Vorabschätzung über MST Ansatz	9
2.2.4 Leaf Backward Algorithmus	10
2.2.5 Implementierungskonzept	10
2.2.6 Implementierung	11
3 Ergebnisse	16
3.1 Auswertung der Resultate	16
3.2 Resultate Steiner Baum	16
3.3 Resultate MST	16
3.3.1 Statistische Auswertung der Resultate	18
4 Conclusio	21
4.1 Zusammenfassung der Resultate	21
4.2 Interpretation der Ergebnisse	21
Literaturverzeichnis	22

Abbildungsverzeichnis

2.1	Digitale Katastralmappe	5
2.2	Bereichsunterteilung in Cluster	7
2.3	Minimal Spannender Baum-MST	10
2.4	Ablaufdiagramm für Einbindung in bestehendes Programm	11
2.5	Ablaufdiagramm der Realisierung	12
2.6	Ablaufdiagramm-Detail der Realisierung	12
3.1	Gegenüberstellung der Resultate von MST und Steiner Baum	17
3.2	Abweichung in gültigen Wahlamtsbereichen	20

Kapitel 1

Einleitung

1.1 Allgemeines

Aufgrund des täglich steigenden Datenaufkommens und der damit verbundenen benötigten Infrastruktur, besteht ein Wettbewerb zwischen den Datennetz-Anbietern. Existierende Datennetze werden von Errichtern vielfach auch an konkurrierende Anbieter vermietet, da es nicht lukrativ wäre, wenn jeder Anbieter sein eigenes Netzwerk errichten würde. Hierfür werden Mietgebühren verlangt, welche von der Austrian Regulation Authority for Broadcasting and Telecommunications (RTR), festgelegt werden. Da die tatsächlichen Kosten der bereits bestehenden Infrastruktur, die für diese Verkabelung von der Telekom Austria angefallen sind, nicht mehr nachvollziehbar sind, werden Modelle benötigt, welche eine völlig neue Verkabelung oder verschiedene Szenarien eines Ausbaues simulieren, um einen fairen Preis festlegen zu können. Bislang existierende Modelle besitzen einen hohen Abstraktionsgrad, das bedeutet, dass die Ergebnisse nur schwer in die realen Gegebenheiten übertragen werden können. Deshalb sind sie weder für Netzwerk-Errichter noch für Regulatoren von großem Nutzen, da sie nicht für eine Grobplanung oder eine Modellierung herangezogen werden können. Die Ergebnisse solcher Programme sollten für die Festlegung der Mietkosten herangezogen werden können oder einer Investitionsabschätzung dienlich sein.

Darum wurde ein Programm, namens RTR_R2008a, entwickelt und implementiert. Dieses liefert Ergebnisse, die den realen Umständen nahekommen. Jedoch beinhaltet das Programm noch Schwachstellen, wie zum Beispiel eine nicht vorhandene Begrenzung der maximal zulässigen Kabellängen, die es zu beseitigen gilt [1].

1.2 Umfeld

Das Programm RTR_R2008a unterteilt Wähleramtsbereiche in Sub-Bereiche. Ein Wähleramtsbereich der Telekom Austria ist ein geographisches Gebiet, welches von einem Verteilzentrum, dem Wähleramt, Endteilnehmer mit Datenleitungen versorgt.

Die Größenbestimmung dieser Sub-Bereiche geschieht in einem parametrierbaren Cluster-Modell des Programms. Diese Parameter sind der maximale/absolute Informations(gehalt) bzw. -bedarf und die maximale/absolute Dilatation. Der Informationsbedarf beschreibt die Anbindung an das Netzwerk. Diese kann mit Kupfer-oder Lichtwellenleitern erfolgen. Wird die Anbindung mit Kupferkabeln vorgenommen, entspricht der Informationsbedarf der Anzahl der benötigten Kupferleiter. Werden Glasfasern verwendet, so gibt der Informationsbedarf die Anzahl der be-

nötigten Wellenlängen eines Lichtwellenleiters wieder. Bestimmt wird der Informationsgehalt auf Basis der Mikrozellendaten.

Die Dilatation eines Wählamtsbereiches ist die maximale, euklidische Distanz, über alle möglichen Anschlussobjekt-Paarungen. Nicht berücksichtigt wird die maximal mögliche Versorgungslänge von der Quelle (Distribution Center) zur Senke (Anschlussobjekt). Es besteht ein Zusammenhang zwischen Parametereinstellungen im Cluster Modell und der daraus resultierenden, maximalen Versorgungslänge in den Sub-Bereichen. Jedoch können über entsprechende Modellparameter keine Einschränkungen bezüglich der maximal möglichen Versorgungslängen im Cluster Modell getätigt werden [2].

Es wird ein Abbild der Trassenführung mittels eines Objektes, bestehend aus Knoten und Kanten erzeugt. Dieses Objekt ist ein Graph, welcher die Wählamtsbereiche der Telekom Austria simuliert und die entsprechenden Anschlussobjekte aufzeigt [1].

Um das bestehende System zu optimieren, sollten die Beschränkung der maximalen Anschlusslängen bereits im Clustermodell getroffen und nicht erst im rechenintensiven Routingmodell durchgeführt werden.

1.3 Prozessablauf

Der Ablauf ist in mehrere Schritte und Teil-Modelle gegliedert. Die Informationen für folgendes Kapitel wurden aus [1] entnommen, sofern nicht anders angegeben.

Geodaten

Vorab muss eine Aufbereitung der Geodaten [8] erfolgen. Diese werden in einem Graphen dargestellt und als *.ist* Dateien abgespeichert. Sie dienen dem Clustermodell als Input.

Clustermodell

Im nächsten Schritt wird der Netzwerkgraph (Wählamtsbereich) in Teilgraphen aufgespalten um unterschiedliche Hierarchie-Ebenen zu erhalten. Dieser Prozess wird als *clustern* bezeichnet und im gleichnamigen Modell durchgeführt (Kapitel 2.1.2).

Als Ergebnis erhalten wir unterschiedliche Hierarchieebenen:

Citynetzebene: Entspricht dem Wählamtsbereich (gesamter Netzwerkgraph)

Accessnetzebene: Teilgebiet des gesamten Wählamtsbereich

Accesszellebene: Gibt Gebiet der anzubindenden Endteilnehmer zu den Netzwerkverteilern wieder

Die erhaltenen Ebenen sind wie folgt miteinander verbunden:

Citynetzebene und Accessnetzebene: Das Wählamt wird mit Verteilern in der Accessnetzebene (Distribution Center) verbunden

Access(netz)-und Zellebene: Ist die Verbindung von Distribution Center zu den Endrichtungen, auch als *last mile* bezeichnet

Routingmodell

Der Datenoutput des Clustermodells dient als Input für das Routing Modell (Kapitel 2.1.4). In diesem werden die tatsächlichen Wege für die Trassenführung aufgrund der Faktoren Flächennutzung und Bebauungsgrad nun abstrahiert und dargestellt.

Der Output des Routingmodells bildet Trassenwege inklusive Anschlussobjekte ab, welche unter Rücksichtnahme den optimalen Kosten genügen. Des weiteren wird eine Auflistung der Kosten und Mengen zum Beispiel der benötigten Leitungen generiert, die als Grundlage einer Investitionsabschätzung dienen [1].

Derzeit kann es passieren, dass sich erst nach einem gesamten Modellierungsdurchlauf eine überdurchschnittlich lange Versorgungslänge ergibt.

1.4 Aufgabenstellung

In dieser Arbeit wird eine Abschätzung getroffen, bei welcher die maximale Versorgungslänge noch vor dem Routing Modell passiert. Bisher werden im Clustermodell die maximalen Anschlusslängen nicht direkt, sondern über den Informationsbedarf und die maximale Dilatation eingeschränkt. Gezeigt wurde dies im Projekt *Versorgungslängensensitivität* Semester IV. Es wird eine Implementierung benötigt, welche die maximale Anschlusslänge unmittelbar nach dem Clustermodell ermittelt und dem Routingmodell einen geeigneten Input liefert.

Ziel ist es, vermeintliche Fehler bezüglich der Anschlusslängen bereits vor dem rechenintensivem Routingmodell zu erkennen und diese zu vermeiden. Das ist ein weiterer Schritt, um das bestehende Programm zu optimieren. Somit ist sichergestellt, dass jede Modellierung korrekte Ergebnisse liefert.

Kapitel 2

Methoden zur Abschätzung der maximalen Versorgungslänge

2.1 Modelle

In diesem Kapitel wird auf die Modelle und Methoden eingegangen, die bei einem Modellierungsvorgang durchlaufen werden.

2.1.1 Gesamtmodell-Input Netzwerkgraph

Die Ausführungen zum Thema Gesamtmodell, welche in diesem Kapitel behandelt werden, beziehen sich auf [1], sofern nicht anders angegeben.

Ein Anschlussbereich wird wie folgt als Netzwerkgraph dargestellt. $G = (V, E)$ mit V als Menge von Knoten und E als Menge der Kanten. Die Erstellung von Knoten und Kanten wird unter der Berücksichtigung der Geodaten-Attribute aus Kapitel 2.2.1 realisiert.

Jeder Knoten $v \in V$ besitzt die folgenden Attribute:

- geographische Position x – *Koordinate*, y – *Koordinate in der digitalen Katastralmappe*
- Informationsbedarf $\pi(v) \in \mathbb{N}_{(0)}$, $v \in V$

Der Informationsbedarf wird auf Basis der Mikrozellendaten identifiziert. Besitzt ein Knoten keinen Informationsbedarf, dann wird er mit $\pi(v) = 0$ belegt. Man spricht von einem spatialen Knoten, welcher nur aufgrund der Graphen-Konstruierung in der digitalen Katastralmappe (DKM) entstanden ist.

Besitzt ein Knoten einen Informationsgehalt, gilt die Forderung $\pi(v) > 0$, es handelt sich um ein Anschlussobjekt. Es gilt zu beachten, ob das Anschlussobjekt mit Kupferkabeln oder aber mit Glasfaserkabeln versorgt wird. Wird das Anschlussobjekt mit Kupferkabeln in das Netzwerk integriert, dann gibt der Informationsbedarf die Anzahl der Teilnehmeranschlusseinrichtungen wie zum Beispiel von Telefonen an. Diese Teilnehmer werden mit kurzen Stichleitungen, sogenannten *drop-wires* versorgt. Wird der Teilnehmer mit Lichtwellenleitern angebunden, so gibt der Informationsbedarf die Anzahl der Wellenlängen wieder, für welche eine entsprechende Anzahl von *drop-wires* gedemultiplexed werden müssen. Kanten $e \in E$ können auch Verbindungstrassen von spatialen Knoten und Anschlussobjekten sein.

Workflow

Nachfolgend wird der Workflow eines Modelldurchlaufes beschrieben. Dieser wird grundsätzlich in zwei Schritten ausgeführt. Im ersten Schritt wird ein Top-down Clustering durchgeführt, auf welches eine Bottom-up Optimierung folgt. Um einen Modelldurchlauf zu ermöglichen, müssen die Inputdaten erstellt werden. Dies wird im nachfolgenden Kapitel 2.1.2 erläutert.

2.1.2 Normierte Geodaten

Normierte Geodaten stellen die generalisierten, topologisch korrekten Nutzungsflächen der digitalen Katastralmappe (DKM) dar, die um für die Graphgenerierung relevante Flächen wie Straßenkreuzungsflächen und Querungsflächen (Eisenbahnübergänge, Brücken über Gewässer) automatisiert bzw. semiautomatisiert erweitert werden [8].

Aufgrund von räumlichen Analysen und auf Vertriebsinformationen basierend, werden branchenspezifische Umsatzpotentiale für Standorte der Anschlussobjekte bestimmt. Hierbei werden die Faktoren der bestehenden Netzinfrastruktur und die nutzbare Infrastruktur in den Geobasisdatensatz integriert. Als bestehende Netzwerkinfrastruktur versteht sich das Strom- und Glasfasernetz, und unter nutzbarer Infrastruktur werden unter anderem existierende Leerverrohrungen, welche es zu nutzen gilt, verstanden. Die Umsetzung erfolgt mittels eines FME (Feature Manipulation Engine)-Modells, in welchem eine automatisierte Aufbereitung des jeweiligen Gebietes auf Basis der definierten Geobasisdaten erfolgt [8].

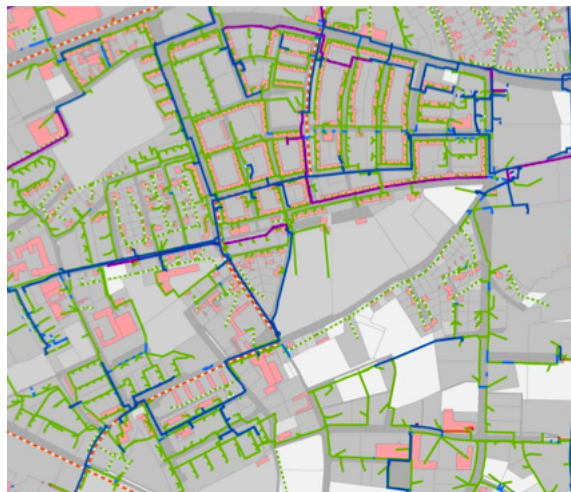


Abbildung 2.1 Ausschnitt der digitalen Katastralmappe [8]

Top-Down Clustering

Um Netzwerke zu optimieren ist es sinnvoll diese in Hierarchien einzuteilen. Eine Unterteilung in ein hierarchisches Netzwerk wird in Folge beschrieben. Dafür wird der Netzwerkgraph G in Teilgraphen $H_i \subset G$, $i = 1, \dots, m$ zerteilt. Jeder dieser entstandenen Teilgraphen spiegelt nun einen Accessnetzbereich (eine Ebene tiefer) wieder. Weiters wird jeder entstandene Teilgraph H_i weiter zerteilt in $K_{i,j} \subset H_i$, $j = 1, \dots, n_i$ und es ergibt sich nun das Accesszellennetz. Man erhält $K_{i,j}$ für die j -te Accesszelle im i -ten Accessnetz vom Teilgraphen H_i .

Somit wurde ein Netzwerkgraph in eine Drei-Ebenen-Hierarchie unterteilt. Dieser Vorgang wird als Top-Down Clustering des Anschlussbereiches G bezeichnet.

In diesem Abschnitt der Modellierung wird die Implementierung eingebettet. Es wird ein zusätzlicher Parameter eingeführt, welcher die Begrenzung der maximalen Anschlusslänge reguliert. So kann der Input des Routing-Modells kontrolliert beziehungsweise reguliert werden.

Bottom-up Optimierung

Um eine Simulation bezüglich der Kosten optimieren zu können, wird das Netzwerk vom Grund auf neu, also Bottom-up simuliert. Die Teilgraphen T' stellen kostenoptimierte, mit dafür entsprechenden Modellen berechnete, Netzwerktrassierungen dar.

2.1.3 Cluster-Modell

Wählamtsbereiche, die als Ausgangspunkt dieser Berechnung dienen, werden durch Zerlegung des Ausgangsgraphen in Beziehung zur den vorhin definierten Ebenen (Kapitel 1.3) unterteilt. Mit Hilfe einer Toolbox, welche im bestehenden Programm implementiert ist, werden Wählamtsbereiche in kleinere Subgebiete, den Accessnetzen, unterteilt. Die sich ergebenden Accessnetzebenen werden abermals unterteilt, wobei man als Output die Accesszellebenen erhält. Der gesamte Vorgang der Zerlegung eines Gebietes in kleinere Subgebiete wird als Clustern bezeichnet und verläuft in einer Top Down Strategie.

Hierbei wird für einen Netzwerkgraphen $G = (V, E)$ eine Zerlegung in Teilgraphen vorgenommen. Teilgraphen werden als $G' = (V, E)$ bezeichnet. Durch die Dilatation eingeschränkt, wird diese Graphenzerlegung durchgeführt.

Die Dilatation wird formal wie folgt beschrieben:

$$\sum_{i=1}^m \max \{ \|u - v\|_2 : u, v \in V_i \times V_i \text{ mit } \pi(u) > 0, \pi(v) > 0 \} \quad (2.1)$$

Es gilt die Teilgraphanzahl m und den minimalen Teilgraphdurchmesser zu finden, wobei die maximale Versorgungslänge auf den Teilgraphdurchmesser beschränkt ist.

$$\max \{ \|u - v\|_2 : u, v \in V_i \times V_i \text{ und } \pi(u) > 0, \pi(v) > 0 \} \leq d_{max} \text{ für } i = 1, \dots, m \quad (2.2)$$

Weiters gilt es, die Bedingung des Gesamtinformationsbedarfs zu berücksichtigen. Die Summen der Teilinformationsbedürfnisse aller Teilgraphen dürfen den Gesamtinformationsbedarf nicht überschreiten.

$$\sum_{v \in V_i} \pi(v) \leq \pi_{max} \text{ für } i = 1, \dots, m \quad (2.3)$$

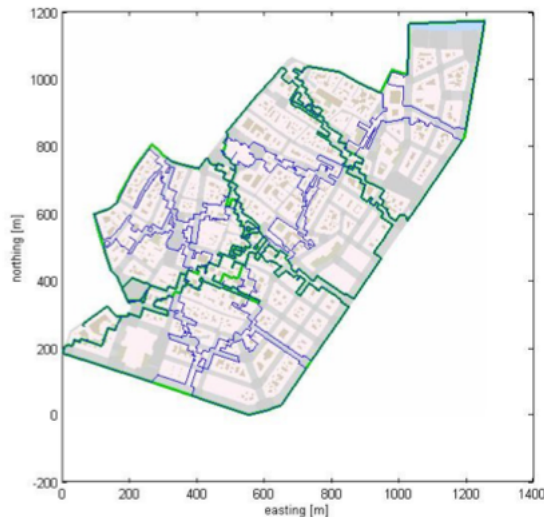


Abbildung 2.2 Bereichsunterteilung nach dem Cluster-Modell, ersichtlich ist die Einteilung in Accessnetze (grün) und Accesszellen (blau) [2]

Es gibt zwei unterschiedliche Möglichkeiten das Clustern durchzuführen:

- Barrier Clustering
- Dens Clustering

Wird die Forderung (2.1 Dilatation) dadurch ersetzt, dass alle Kanten, welche die Teilgraphen H_i miteinander verbinden, zu einer oder auch mehreren Nutzungsklassen gehören müssen, so wird vom Barrier-Clustering Modell gesprochen. Für dieses Modell ist die Lösung heuristisch. Beim Verwenden von Barrier-Clustering werden die Teilgraphen durch Breitensuche solange initialisiert, bis man auf Kanten mit vorgegebenen Nutzungen stößt.

Verwendet man Dens-Clustering, kann für ein gegebenes m eine exakte Lösung gefunden werden. Wird die Lösung von den Forderungen 2.2 oder 2.3 verletzt, dann wird m solange erhöht, bis sie wieder erfüllt sind. Im Anschluss werden die Teilgraphen vereinigt, bis eine der Bedingungen, 2.2 oder 2.3, verletzt wird.

2.1.4 Routing-Modell

Nach dem Clustern der unterschiedlichen Ebenen in Citynetz, Accessnetz und Accesszellen ist es nun möglich Trassen in den einzelnen Teilbereichen zu berechnen. Beeinflussend auf die Wahl der Trassierung wirken Faktoren wie zum Beispiel Kosten für Trassenführung. Diese Kosten setzen sich aus mehreren Faktoren, wie zum Beispiel unter anderem den Widmungsklassen, zusammen.

Ergo ist es günstiger eine Verlegetrasse durch ein unbebautes Gebiet zu erstellen, während im innerstädtischen Bereich die Kosten dafür wesentlich höher sind. Diese sind in den Bereichen ländlich und innerstädtisch signifikant unterschiedlich und wirken nun auf die Erstellung eines Trassenverlaufes ein.

Besondere Rücksicht wird unter anderem auf Gewässer genommen. Die Kosten, um diese zu

durchqueren, sind im Vergleich zu einer Trassenführung auf festem Boden, wesentlich höher. Um dem entgegenzusteuern wird dieser Parameter per default hoch angesetzt. Aufgrund dessen kann es vorkommen, dass es einzelne Anschlussobjekte gibt, welche über einen überdurchschnittlich langen Weg angebunden werden müssen [2].

Um solche Anschlusslängen zu vermeiden, werden die Versorgungslängen bereits im Cluster-Modell berücksichtigt. Somit können sich im Routing-Modell solch lange Trassenführungen nicht mehr ergeben, da das Längenmaximum bereits zuvor festgelegt wurde. Die Anschlussobjekte müssen nun einem anderen Cluster zugeteilt werden, um eine Optimierung bezüglich der Längenbegrenzung zu gewährleisten.

Nachdem die Routen für alle Teilbereiche (Bottom Up) kostenoptimiert, bezüglich der Trassierungslänge, erstellt wurden, sind die Routinginformationen in vorgefertigten Dateien aufbereitet.

Das Routing-Modell berechnet die Trassierung auf Basis des gegebenen Graphen. Dieser Graph wird vom Cluster-Modell aufbereitet und bereitgestellt. Es ist nun möglich eine kostenoptimierte Trassierung in den einzelnen Subgebieten eines Wähleramtsbereichs zu simulieren. Dabei wird eine Bottom Up – Strategie verfolgt. Das bedeutet, dass zuerst alle Accesszellen, anschließend alle Access- und danach Citynetze geroutet werden. In jeder dieser Hierarchien werden die Kanten so ausgewählt, dass in jedem Anschlussbereich

- a. genau eine Verbindung zwischen jedem enthaltenen Anschlussobjekt und einem Versorgungszentrum oder
- b. zwei knotendisjunkte Verbindungen für jedes Paar von Anschlussobjekten zum Versorgungszentrum existiert

und die Summe der Kosten der gewählten Kanten ein Minimum ist.

2.2 Mathematisches Konzept

Für nachfolgendes Kapitel ist es notwendig einige Begriffe vorab zu erklären.

2.2.1 Baum

Ungerichtete Graphen können in Form eines Baumes dargestellt werden.

Ein ungerichteter Graph G ist dann ein Baum, wenn:

1. G ist zusammenhängend und $m = n - 1$;
2. G enthält keinen geschlossenen Weg und $m = n - 1$;
3. In G gibt es zwischen jedem Paar genau einen Weg;

Definition: Ein ungerichteter Graph $G = (V, E, \gamma)$ heißt Wald, wenn G kreisfrei ist, also keinen elementaren Kreis besitzt. Falls G zusätzlich zusammenhängend ist, so heißt G Baum [6].

Desweiteren besteht ein Baum aus folgenden Elementen:

- Kanten
- Knoten

Knoten sind über Kanten miteinander verbunden. Wurzel und Blätter sind Sonderformen von Knoten. Die Wurzel ist der Startpunkt eines Graphen. Blätter hingegen sind die Enden eines Baumes, zu ihnen führt nur eine Kante.

2.2.2 Steiner Baum

Ist $G = (V, E)$ ein vollständiger Graph mit Kantengewichten $c : E \rightarrow \mathbb{R}_+$, so scheint bei flüchtiger Betrachtung das Problem, einen gewichtsminimalen Steiner Baum für die Terminalmenge K zu finden, identisch damit zu sein, einen minimalen spannenden Baum im (vollständigen) induzierten Subgraphen $G[K]$ zu bestimmen. Eine nähere Betrachtung zeigt aber, dass ein minimaler spannender Baum in $G[K]$ nicht zwangsweise ein minimaler Steiner Baum ist.

Definition: Sei $G = (V, E)$ ein ungerichteter Graph und $K \subseteq V$ eine beliebige Teilmenge der Eckenmenge. Ein Steinerbaum in G für die Menge K ist ein Teilgraph $T \subseteq G$, der ein Baum ist und dessen Eckenmenge K umfasst: $K \subseteq V(T)$. Die Elemente von K nennt man Terminale, die Ecken aus $V(T) \setminus K$ *Steinerpunkte* [6].

2.2.3 Vorabschätzung über MST Ansatz

Derzeit wird ein Steiner Baum zum Bestimmen der Trassierung verwendet. Mittels MST soll nun eine Abschätzung noch vor dem Routing-Modell durchgeführt werden.

Unter einem MST versteht sich ein minimal spannender Baum eines Graphen.

Definition: Sei $G = (V, E, \gamma)$ ein ungerichteter Graph. Ein Partialgraph $H = (V, E', \gamma)$ von G ist ein spannender Baum, wenn H ein Baum ist. Jeder spannende Baum von G enthält mindestens $|V(G)| - 1$ Kanten [6].

Jeder Graph besitzt auch ein Gewicht. Das Gewicht setzt sich aus der Summe der Kantengewichte zusammen. Alle Knoten des Hauptgraphen müssen auch im MST enthalten sein (Abbildung 2.3). Ein Baum heißt minimal spannend, wenn kein anderer Spannbaum im selben Graphen, mit einem geringeren Gewicht existiert [6].

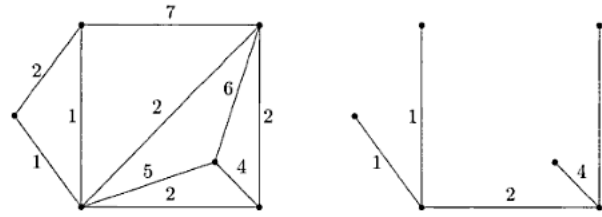


Abbildung 2.3 Hauptgraph [6] und der Minimal Spannende Baum zu diesem [6]

Mittels des Prim-Algorithmus wird ein minimal spannender Baum erstellt. Als Output ergibt sich ein Untergraph $G'(V', E')$, mit den günstigsten Kanten und Knoten, welche benötigt werden, um den minimal spannenden Baum für den ursprünglichen Graphen $G(V, E)$ bestimmen zu können [6].

2.2.4 Leaf Backward Algorithmus

Der Output aus Kapitel 2.2.2 stellt nun einen minimal spannenden Baum (MST) dar.

Spannende Bäume existieren nur in zusammenhängenden ungerichteten Graphen.

In diesem MST sind noch Knoten enthalten, welche keinen Informationsbedarf beinhalten und auch nicht als Verbindungsknoten dienen.

Nun müssen diese Bedingungen zutreffen um solche Knoten zu entfernen:

1. ist der Informationsgehalt $\pi(v) = 0$
2. hat der Knoten nur einen Nachbarknoten, handelt es sich um ein Blatt

Es gilt für $v \in V$ folgende Bedingungen zu berücksichtigen:

1. *v hat mehr als einen Nachbarn*
2. *v ist $\pi(v) > 0$, also ein Anschlussobjekt*
3. *v ist als Distribution Center gekennzeichnet*

Verletzt ein Knoten im Graphen eine oder mehrere dieser Bedingungen, so wird er und die zuführende Kante aus dem Graphen entfernt.

Der Output-Graph entspricht nun einem Steiner Baum.

2.2.5 Implementierungskonzept

Die Implementierung setzt sich aus den folgenden Teilschritten zusammen.

- Berechnen der günstigsten Wege im Graphen - MST Algorithmus
- Reduzieren der redundanten Knoten mittels Leaf Backward Algorithmus
- Berechnen des Flächenschwerpunktes und bestimmen des Distribution Centers

- Vergleich von Output Steiner Baum und MST

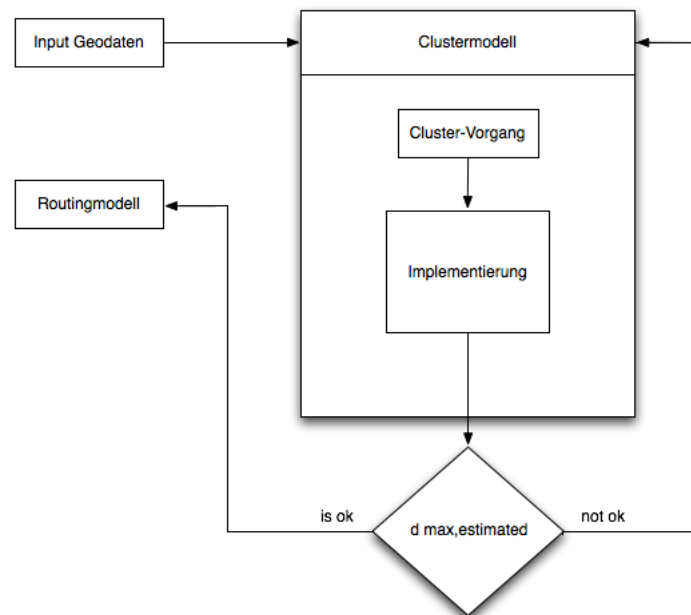


Abbildung 2.4 Ablaufdiagramm für Einbindung in bestehendes Programm

2.2.6 Implementierung

Da die Einbindung in das Programm RTR_R2008a nicht Teil dieser Arbeit ist, wird der Vorgang abgeändert durchgeführt. Es wird ein Testbed erstellt, in welchem die Resultate des MST-Ansatzes mit den Ergebnissen des Steiner Baumes verglichen werden. Wie im Ablaufdiagramm (Abbildung 2.5) ersichtlich, werden nun die Schritte in Pseudo-Codes beschrieben. Da das Cluster-Modell und die Aufbereitung der Geodaten bereits implementiert sind, werden diese im ersten Schritt nicht weiter beachtet. Somit wird mit der Berechnung des MST begonnen. Als Input dient hierfür ein *.ist* File, in welchem sich die Geodaten befinden.

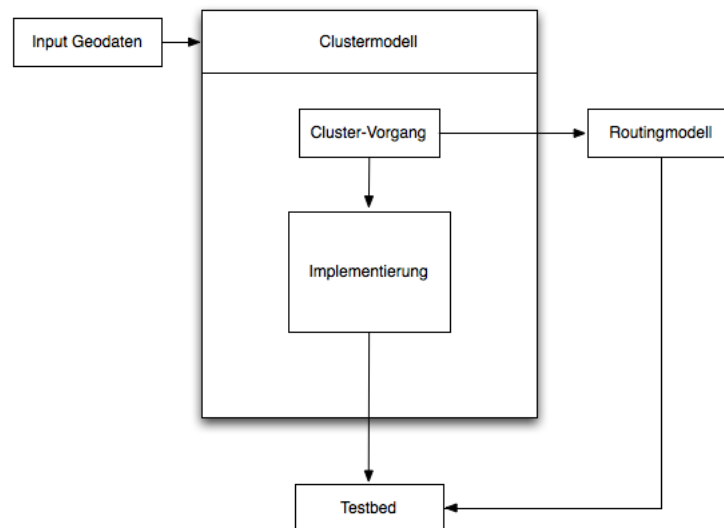


Abbildung 2.5 Ablaufdiagramm der Realisierung

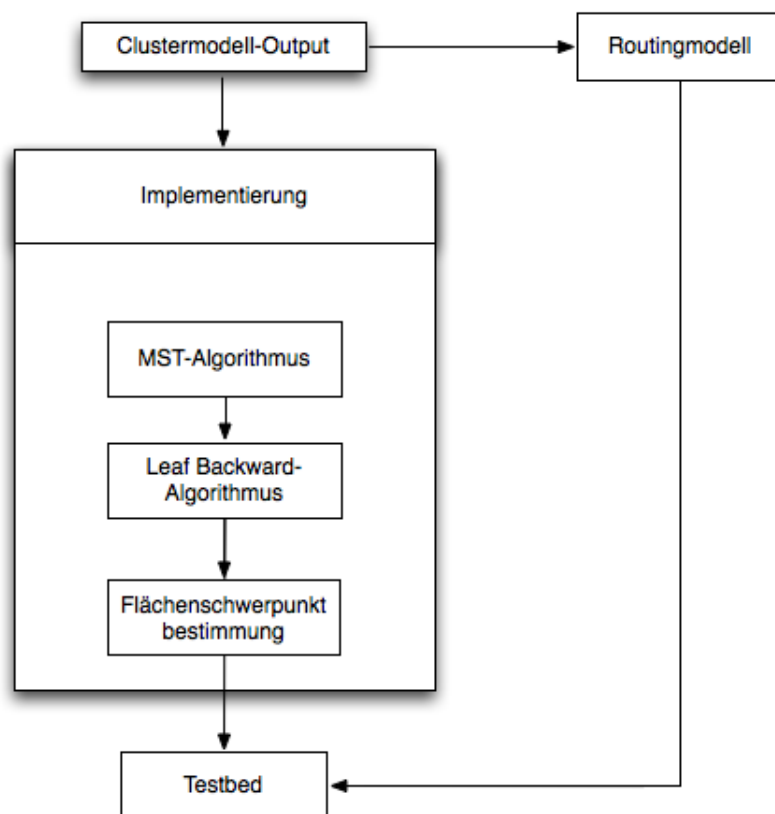


Abbildung 2.6 Detailansicht der realisierten Implementierung

MST Algorithmus

Um einen minimal spannenden Baum zu erhalten gibt es mehrere Möglichkeiten. In dieser Arbeit wird der Algorithmus von Prim verwendet, um einen MST zu erhalten. In diesem Algorithmus bilden die ausgewählten Kanten zu jedem Zeitpunkt einen Baum $G = (V, E)$. Zu

Beginn wird ein beliebiger Knoten $v \in V$ gewählt. Durch weiteres Hinzufügen der günstigsten Kanten, entsteht der gesuchte Baum [9].

Die Auswahl der Kanten wird dabei wie folgt getroffen:

1. Sei V die Menge der Ecken des gesuchten Baumes G und E die Menge der Ecken des Graphen.
2. Man wähle unter den Kanten (u, v) , deren Anfangsecke $u \in V$ und deren Endecke $v \in E \setminus V$ liegen, diejenige mit der kleinsten Bewertung aus.
3. Diese wird dann in G eingefügt.
4. Weiters wird v in V eingefügt.
5. Dieser Schritt wird solange wiederholt bis $V = E$ ist.

Beim Algorithmus von Prim wird das Sortieren der Kanten vermieden, jedoch ist der Aufwand der Kanten aufwändiger. Die Auswahl und Verwaltung der Menge V kann mit Hilfe eines Feldes erfolgen. Im Feld wird für jede Ecke $e \in E \setminus V$ die kleinste Bewertung unter den Kanten (e, v) mit $v \in V$ gespeichert. Es wird auch die Endecke dieser Kante abgespeichert. Wurde eine Kante (e, v) ausgewählt wird die Ecke e als zugehörig von V markiert. Die Laufzeit der Prim-Prozedur beträgt $n - 1$ Iterationen, da bei jedem Durchlauf genau eine Ecke eingefügt wird. Der Aufwand für die Auswahl der Kante und das Aktualisieren des Feldes beträgt $O(n)$. Das bedeutet, der Gesamtaufwand beträgt $O(n^2)$. Der Primalgorithmus sollte Verwendung finden, wenn die Anzahl der Kanten m in etwa der Größenordnung von n^2 entspricht [9].

Algorithmus 2.1: Prim-Algorithmus

```

var
  U: set of Integer;
  u, e: Integer;
begin
  B.initBaum(G);
  U := {1};
  while U.anzahl  $\neq$  n do begin
    Sei (e, u) die Kante aus G mit der kleinsten Bewertung,
    so dass  $u \in U$  und  $e \in E \setminus U$ 
    B.einfuegen(e, u);
    U.einfuegen(e);
  end
end

```

Der Output ist, wie in Kapitel 2.2.2 beschrieben, der kostengünstigste Untergraph. Es folgt der Schritt des Leaf Backward Algorithmus, in welchem die redundanten Knoten des Graphen entfernt werden.

Leaf Backward Algorithmus

In diesem Algorithmus wird ein schrittweises Terminieren der Knoten v mit $\pi(v) = 0$ und der inzidenten Kanten e , zu diesem Knoten, verfolgt. Neben Entfernen der Knoten aus dem MST, muss auch der Grad des Vorgängerknotens vermindert werden.

Algorithmus 2.2: Leaf-Backward-Algorithmus

```

Für alle  $v \in V$  führe aus:
var
  g: Integer;      // Grad des Knoten
  v, e: Integer;

while
  if ( $Nachbar(v) = 1$  und  $\pi(v) = 0$  und  $istDistr(v) = false$ )
    g.Nachbar(v) - 1;
    delete v;
    delete e;
  end
end

```

Gewichteter Flächenschwerpunkt

In jedem Teilbereich ist ein Wähleramt, oder auch Distribution Center genannt, notwendig. Um einen geeigneten Standpunkt für dieses festlegen zu können, muss der gewichtete Flächenschwerpunkt ermittelt werden. Ein ungerichteter Graph $G(V, E)$ bildet einen Teilbereich unserer Netzstruktur ab. Hierbei kann sich um eine Accesszelle oder ein Accessnetz eines Wähleramtsbereiches (Kapitel 1.2) handeln. Für die Ermittlung des Schwerpunktes muss ein Startknoten $u \in V$ gewählt werden [6].

Für seine Wahl sollte die Position so gewählt werden, dass er bestmöglich einem Distribution Center entsprechen kann. Zu diesem Zeitpunkt der Modellierung befindet man sich im Top-Down Clustering (Kapitel 2.1.3), darum kann die Wahl des potentiellen Distribution Centers nur abgeschätzt und noch nicht errechnet werden. Die exakte Bestimmung eines Distribution Centers kann erst im Routing-Modell (Kapitel 2.1.4) bestimmt werden.

Die Position eines Knoten $u \in V$ im Graph G besitzen x - und y -Koordinaten, welche Teil aus \mathbb{R} sind. Diese werden in einer Matrix dargestellt.

$$y = [v_x^1, v_y^1, v_x^2, v_y^2, \dots, v_x^n, v_y^n] \quad (2.4)$$

Mit der Anzahl n der Anschlussobjekte ($\pi(v) > 0$) $u \in V$. Darum werden alle x - und y -Koordinaten von allen Anschlussobjekten des Graphen in den Vektor y gespeichert.

Es wird eine Matrix $C \in Mat(2n \times 2)$ definiert.

$$C = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \end{bmatrix} \quad (2.5)$$

Es wird eine Gütefunktion $F(p)$ mit gegebenem Vektor y und gegebener rechteckiger Matrix C formuliert. Gesucht wird ein Vektor $p \in \mathbb{R}^2$, welcher die Gütefunktion

$$F(p) = \frac{1}{2} \|y - Cp\|^2 = \frac{1}{2} \sum_{i=1}^n [(v_x^i - p_x)^2 + (v_y^i - p_y)^2] \quad (2.6)$$

minimiert. Der Vektor $y - Cp$ entspricht der zu minimierenden Abweichung. Zur Lösung der

Problemstellung wird die *Methode der kleinsten Quadrate* angewandt [7]. Dieses Problem kann optimal gelöst werden.

$$p^* = (C^T C)^{-1} C^T y \quad (2.7)$$

Dieser Vektor $p^* = \begin{bmatrix} p_x^* \\ p_y^* \end{bmatrix}$ enthält die x- und y- Koordinaten des geographischen Schwerpunktes des Graphen G , bezüglich aller enthaltenen Anschlussobjekte. Somit kann die Position des Distribution Centers optimal bestimmt werden.

Kapitel 3

Ergebnisse

3.1 Auswertung der Resultate

Es werden in diesem Kapitel die erhaltenen Resultate mit der Variante des MST behandelt. Diese werden mit den Ergebnissen des Steiner Baums verglichen.

3.2 Resultate Steiner Baum

Das Programm RTR_R2008a verwendet zur Ermittlung der optimalen Trassierung einen Steiner Baum, welcher in Kapitel 2.2.2 kurz beschrieben wird. Mit diesem wird der kostenoptimierte Weg für eine Trassierung als Graph abgebildet. Die Modellierung von nicht redundanten Netzwerktrassierungen (Bäumen) erfolgt mittels Steiner Bäumen. Diese Probleme werden mit Hilfe von primaldualen Approximationsmethoden gelöst. Trassenredundante Netzwerkstrukturen werden als Generalized Steiner Baum-Problem modelliert, wobei wiederum primal-duale Approximationsmethoden oder Augmentierungsansätze zur Lösung der Probleme verwendet werden [2].

3.3 Resultate MST

Wie in Kapitel 2.2.3 erläutert, handelt es sich in dieser Arbeit um einen wissenschaftlichen Versuch einen Steiner Baum nachzubilden. In Abbildung 3.1 werden die Resultate beider Baumalgorithmen gegenübergestellt. Sofort erkennbar ist, dass mit dem MST die Längen in nahezu jedem Wählamtsbereich wesentlich länger sind. Im Durchschnitt ergeben sich um 68,78% längere Anschlusslängen, als beim Steiner Baum Algorithmus. Bei genauerer Betrachtung ist erkennbar, dass in nur einem der gewählten Wählamtsbereiche die maximale Anschlusslänge kürzer ausfällt.

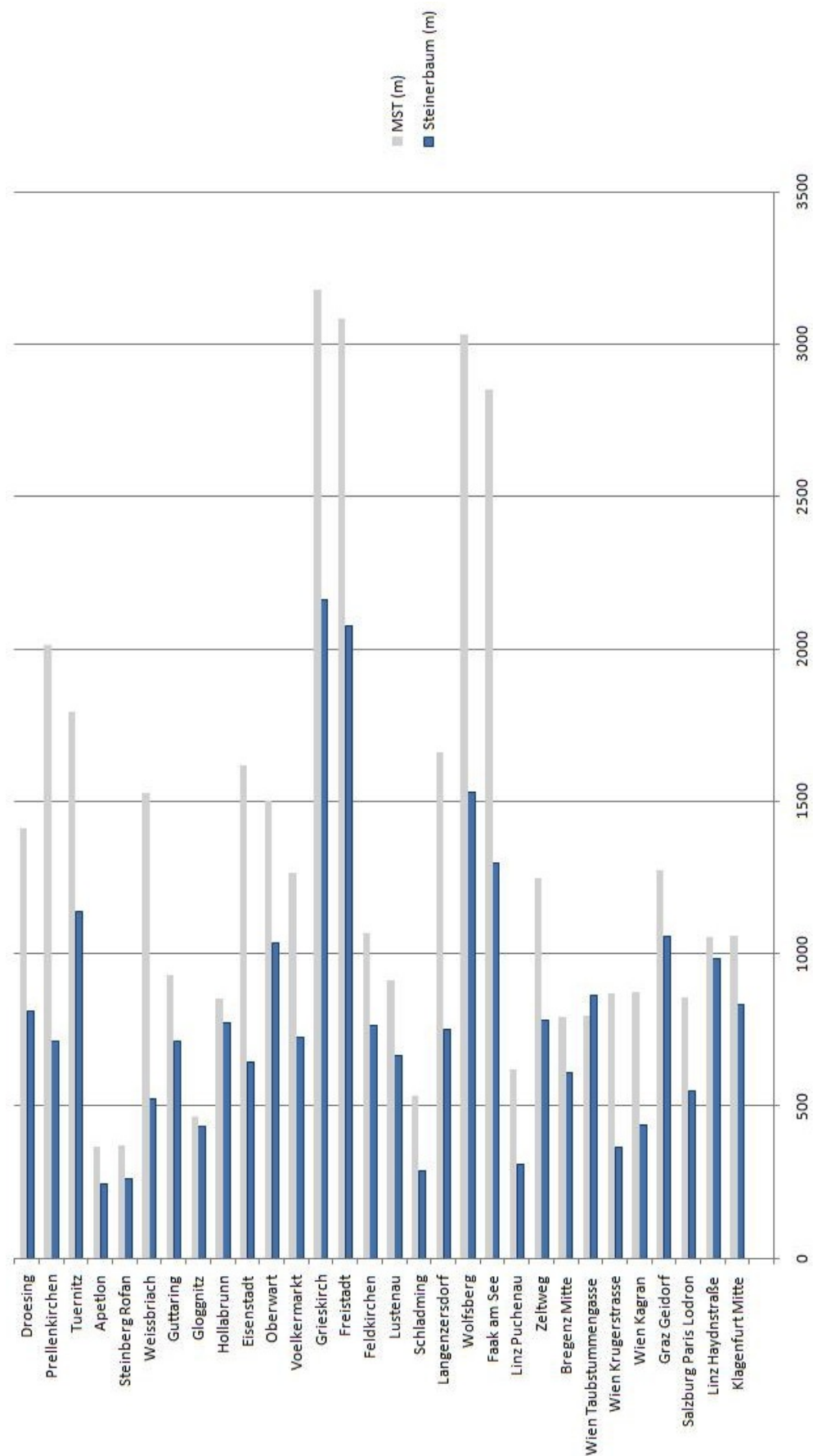


Abbildung 3.1 Gegenüberstellung der Resultate von MST und Steiner Baum in Meter

Um die Ergebnisse besser interpretieren zu können, muss eine statistische Bewertung durchgeführt werden.

3.3.1 Statistische Auswertung der Resultate

Die erhaltenen Ergebnisse liegen in einem breit gestreutem Bereich. Nun muss ein Bereich definiert werden, in welchem die Anschlusslängen liegen dürfen. Realisiert wird diese Berechnung, indem ein Konfidenzintervall (darauffolgendes Unterkapitel) berechnet wird.

Mit Hilfe von diesem wird eine Unter- und Obergrenze ermittelt. Über- oder unterschreiten die erhaltenen Anschlusslängen (Abbildung 3.1) eines Wähleramtsbereiches den definierten Bereich, werden sie verworfen.

Bemerkung: Es ist möglich, dass es zu von uns genannten *Ausreißern* kommt. Diese überdurchschnittlich langen Versorgungslängen können entstehen, wenn die Durchquerung eines Gewässers vermieden wird (Kapitel 2.1.4). Auch in den Wähleramtsbereichen, die für diese Arbeit herangezogen wurden, war einer dieser *Ausreißer* vorhanden. Jedoch verfälschen sie in Folge das Resultat und müssen deshalb ausgeschlossen werden.

Beurteilende Statistik hat das Ziel, eine Aussage über eine Grundgesamtheit (Population und Merkmale) zu gewinnen. Dies wird mit Stichproben durchgeführt.

Definition: Eine Stichprobe vom Umfang $n \in N$ ist eine Folge von zufällig gewählten Objekten aus der Population - also eine Folge von n Zufallsvariablen $(X_i)_{i \in N}$ [3].

Konfidenzintervall

Gesucht wird ein Intervall, um den realisierten Schätzer des Erwartungswertes \bar{x} , sodass der exakte Erwartungswert μ der Zufallsvariable X mit Q -prozentiger Wahrscheinlichkeit in diesem Intervall liegt. Es wird angenommen, dass die Variablen $X \in N(\mu, \sigma)$ wobei μ und σ unbekannt sind. Unter Zuhilfenahme einer realisierten Stichprobe x_1, \dots, x_n vom Umfang n , können die unbekannten Parameter geschätzt werden.

Der Erwartungswert μ kann mit \bar{x} abgeschätzt werden.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (3.1)$$

Das erhaltene \bar{x} entspricht einem realisierten Schätzer des Erwartungswertes μ . Nun kann die Standardabweichung σ geschätzt werden.

$$s_{\bar{x}} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n(n-1)}} \quad (3.2)$$

Mehrere Stichproben vom Umfang n führen zu mehreren Werten für \bar{x} und $s_{\bar{x}}$ und damit auf mehrere Konfidenzintervalle.

Beispiel: $Q=0.9$. Werden 100 Stichproben gezogen so folgen 100 Konfidenzintervalle K . Es darf erwartet werden, dass 90 den zu schätzenden Parameter μ enthalten und 10 nicht. Man beachte in diesem Zusammenhang: je grösser Q gewählt wird, desto breiter die Intervalle [3].

In dieser Berechnung sind X die maximalen Versorgungslängen (ersichtlich in Abbildung 3.1). Diese Längen sind die gewählten Stichproben x_1, \dots, x_n mit dem Umfang n (Anzahl der Wählerbereiche).

Um das Konfidenzintervall bestimmen zu können, muss ein zweiseitiges Quantil der $t_{n-1, \alpha}$ -Verteilung bestimmt werden, wobei $n - 1$ der Freiheitsgrad und $\alpha = Q - 1$ die Wahrscheinlichkeit der t -Verteilung ist. Q bezeichnen wir auch als Vertrauenswahrscheinlichkeit. Es wird eine Stichprobe realisiert und ein Konfidenzintervall KI ernannt, wobei μ und σ unbekannt sind. Das Konfidenzintervall KI für μ sei mit einer Vertrauenswahrscheinlichkeit Q gegeben:

$$KI = [\bar{x} - s_{\bar{x}} t_{\alpha/2, n-1}, \bar{x} + s_{\bar{x}} t_{\alpha/2, n-1}] \quad (3.3)$$

Das Ergebnis des Konfidenzintervalls KI für diese Realisierung beträgt:

$$[573, 360 \text{ m}; 1024, 500 \text{ m}] \quad (3.4)$$

Der erste Wert ist die Untergrenze, der zweite gibt die Obergrenze an. Die Lösung der Dichtefunktion der t -Verteilung ergibt $t = 2,75$ [4]. Es wurde eine Vertrauenswahrscheinlichkeit von 99% gewählt.

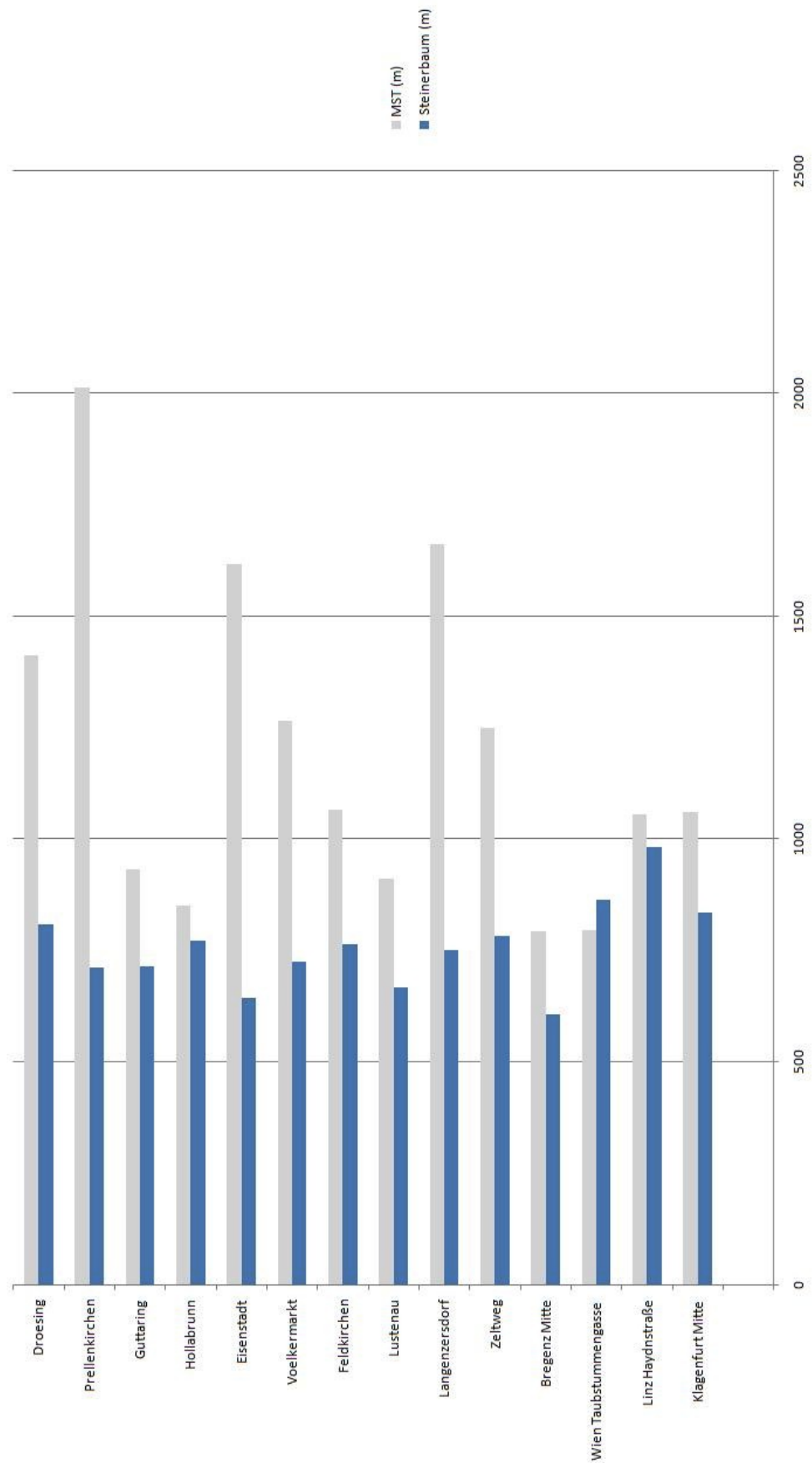


Abbildung 3.2 Abweichung der im Konfidenzintervall liegenden Wählamtsbereiche

Kapitel 4

Conclusio

4.1 Zusammenfassung der Resultate

Ziel dieses wissenschaftlichen Versuches war es, einen Steiner Baums [5] zu approximieren. Hierfür wurde ein minimal spannenden Baum [6] konstruiert und die Resultate beider Ansätze wurden abschließend verglichen.

4.2 Interpretation der Ergebnisse

Die Untersuchungen ergaben, dass es mit dem Ansatz des MST zu großen Abweichungen kommen kann. Die erhaltenen Versorgungslängen verhielten sich im Durchschnitt um 61,029% länger, als die Versorgungslängen welche mittels Steiner Baum errechnet wurden. Dies ist in Abbildung 3.2 ersichtlich.

Nachfolgend wurde eine statistische Auswertung, unter Ernennung eines Konfidenzintervalls durchgeführt. Von den ursprünglichen Wähleramtsbereichen, welche für die Untersuchung herangezogen wurden, lagen 46% im berechneten Intervall. Mit einem Ergebnis, dass sich mehr als die Hälfte der untersuchten Gebiete nicht in das Modell aussagekräftig einbinden lassen und somit nicht als Vergleichswert herangezogen werden konnten.

Literaturverzeichnis

- [1] Bachhiesl,P.,et al;*Bottom up Kostenmodellierung von Hybridnetzen*
- [2] Bachhiesl,P.,et al; *Model-and parameter description for the bottom-up network simulaion tool RTR_R2008a-Technical Report*, (Peter Bachhiesl)
- [3] Bachhiesl,P.; *Skriptum zur Lehrveranstaltung Mathematik III*, (Peter Bachhiesl)
- [4] Bronstein,et al; *Taschenbuch der Mathematik*,(Bronstein,) 7.Auflage, Frankfurt am Main, 2008
- [5] Jungnickel, D.; *Graphs networks and algorithms*, 3. Auflage, Berlin Heidelberg, 2008
- [6] Krumke,S.; Noltemeier, H.; *Graphentheoretische Konzepte und Algorithmen*, 2. Auflage, Wiesbaden, 2009
- [7] Papageorgiou,M.; *Optimierung*, 2. Auflage, München, 1996
- [8] Paulus,G., et al; *Entwicklung von kostenoptimierten räumlichen Szenarien für den strategischen Ausbau der Glasfasernetz-Infrastruktur am Beispiel eines Multi-Utility Unternehmens*, 2. Forschungsforum der österreichischen FHs, FFH 2008, 26-27 Mär 2008, Wels, Austria.
- [9] Turau,V.; *Algorithmische Graphentheorie*, (Volker Turau), 3. Auflage, München, 2009