Westfälische
Wilhelms-Universität
Münster

microm
Consumer Marketing

Area segmentation algorithms - A geoinformatic approach to Geomarketing strategies

Sabine Schmidt
Master of Geoinformatics
Matriculation register
In cooperation with microm

5th of December 2015

First supervisor:
Dr. Torsten Prinz (WWU)
Second supervisor:
Dipl. Ing. Sven Reifegerste (microm)

Declaration

I hereby declare that I completed this work without any improper help from a third party and without using any aids other than those cited. All ideas derived directly or indirectly from other sources are identified as such. This declaration also refers to the representation of figures and visual material.

5th of December 2015 | Sabine Schmidt

Abstract

Already since several years companies are doing Geomarketing analyses to support their decisions raising their profit and minimizing risks during their product and location planning. Nowadays Geomarketing strategies getting more and more important because more companies realize that they are a helpful tool. To make them more effective concerning time and cost the analyses are done more and more by using computers. But at the moment just a small amount of tools are available which offer functionality of Geomarketing strategies to their costumers. Additionally often the available software considers not all requirements that are defined by Geomarketing experts during the application of area segmentation. Area segmentation processes are clustering small geographic areas to bigger units which are called territories. Doing so the most common conditions are that the created territories need to be balanced, compact and contiguous. Motivated by the small amount of available software and that the most important conditions of area segmentation are not kept in mind the goal of this thesis is the development of an algorithm supporting area segmentation while considering balanced, compact and coherent territories. Therefore previous work will be analysed for determining promising algorithms. These algorithms will be implemented in order to compare them and determining the most promising approach. The algorithm with best results will be applied to three Geomarketing strategies: area segmentation and optimization, Greenfield analyses and Whitespot analyses. Finally the results will be evaluated and compared to previous work showing advantages and possible disadvantages if exist.

Acknowledgement

First of all I thank Microm Micromarketing-Systems and Consult GmbH which offer the possibility to do my master thesis within that interesting field of Geomarketing. They have been supported me during the whole time of implementation very well. Additionally microm provided all data which were necessary to do the implementation. Especially I would like to thank Andreas Strade and Sven Reifegerste.

Furthermore I thank my first supervisor Dr. Torsten Prinz who helped in case of any questions or problems.

Alexander Butsch from KIT Institute of Operations Research provided some information about compactness measures and answered my question quite well. Therefore I would like to thank you.

Without the help of my boyfriend and my family the realisation of this thesis would not have been possible. Throughout the entire period producing the thesis they have been supported me very well. I thank all of them very much with special thanks to my boyfriend.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Definition and aim of Geomarketing

The term of Geomarketing has established more and more in the field of marketing within the last years. A first approximation to the notion of Geomarketing was done in 1995 by Frühling and Steingrube [FS95]. They had explained that Geomarketing is just a genus for several instruments within the field of marketing. This shows that Geomarketing is no methodology but rather a discipline. Although some definitions had occurred in the 90s, the first use of Geomarketing analysis went back to the 50s. Already 1952 the first map showing the distribution of purchasing power in Germany was published. In 1982 several companies were founded, that have offered tools and possibilities for their costumers to practise geographic analysis. As a result these researches got easier more and more. Consequently the comprehensive application of Geomarketing was born [HM08]. Within the early 90s approaches and fundamentals of geographic analysis and the governance of marketing and distribution has been described within Geomarketing publications. The central idea of Geomarketing is that marketing composed of price, product, distribution and communication will be complemented by space. Consequently operating numbers can change dependent on the spatial location by spatial phenomena of productions and logistics.

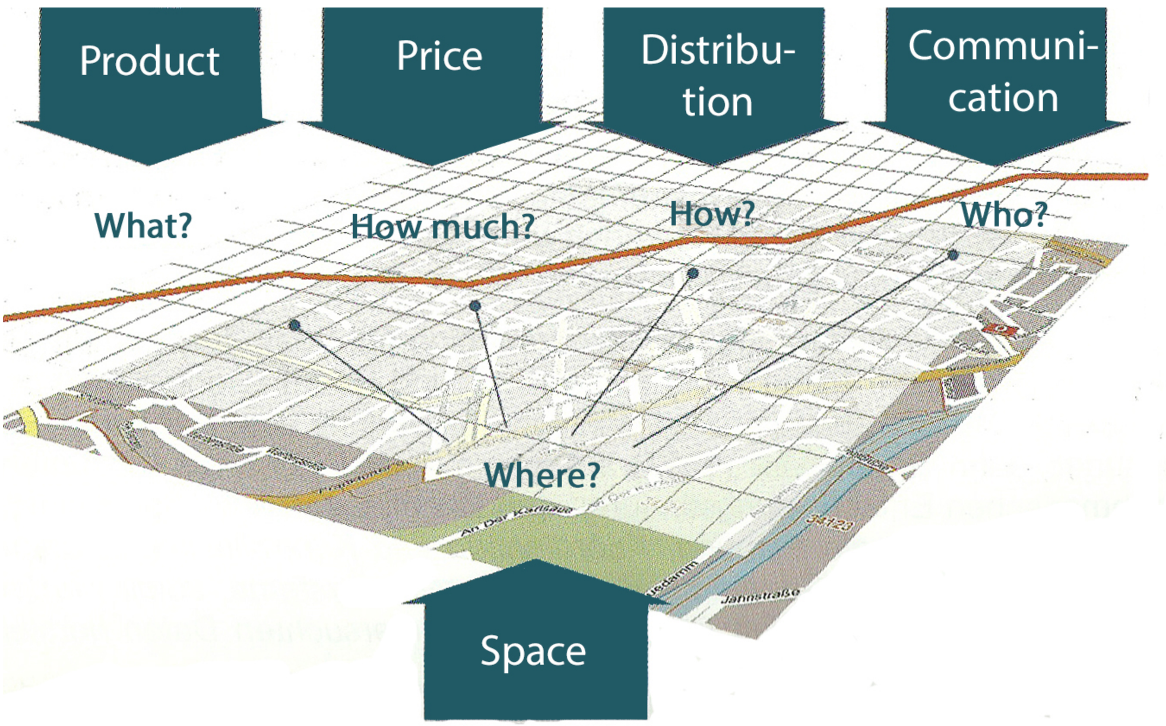


Figure 1: *Generic consideration of Geomarketing aspects complemented by space [HM08]*

Although the central idea of Geomarketing is known in the 90s often incorrect explanations are written down while defining the notion of Geomarketing. In several publications it is readable that Geomarketing is a spatial analysis of the market using a geographic information system. But by particularly consideration it is recognizable that this statement is not correct because a geographic information system is just a tool which supports Geomarketing analyses. Herter and Mühlbauer [HM08] made a often quoted definition of Geomarketing. They have defined that Geomarketing analyses examine current as well as potential markets considering spatial structures to make the planning of product sales more effective. Additionally the control of the markets should be more quantifiable. That means that all available information about the market are connected to a spatial reference system to make dependences, potentials and other properties visible. The application of Geomarketing analysis have their origins in the minimization of entrepreneurial risks by making the market more transparent so that an purposeful acting is possible. In the course of this, Geomarketing was established as a sub-discipline of the field of marketing. During the application several benefits can be achieved. By knowing potential costumers and competitors the marketing and distribution of products from a company can be done more dedicated so that efficiency enhancement and cost reduction can be caused. Additionally analysis can achive a lead to competitor companies. Furthermore inquests of the market may be helpful during the planning of new locations to determine a site with a high potential so that the risk of a malinvestment can be minimized. It is recognizable that as higher as the number of costumers is, none the worse the benefit of Geomarketing analyses are. During the surveys of the market several principles are utilized. One of them is the spatial factor of the market. Using spatial data (e.g. density of customers, locations, branch offices etc) imporant dependences can be visualized. The spatial data are almost given using addresses. Additionally a spatial heterogeneity can be recognized during analyses. This means that the market differentiates in space. In conclusion a third principle is generated by the mentioned fact. It describes the spatial segmentation of the market. It is recognizable that the higher the number of customers is, the higher the benefits from Geomarketing analyses are. As a prominent example and general speaking consumer in the western part and the eastern part of Germany differentiate in some aspects. In contrast an identification of consumers with similar affectations within a small range of space is possible. From this it follows the neighbourhood principle which explains that neighboured customers have a similar behaviour considering marketing aspects as product purchase. This fact has two reasons. On the one hand costumers with an analogical lifestyle life in the same space and consequently show common characters in costumer behaviour. On the other hand these people share the same infrastructure which takes influences to their purchasing habits as well. Additionally the distance to the location of a company affects the costumers in their decisions whether going to this location or not. Geomarketing is based on three stacks: information of the market, geodata and analyses.

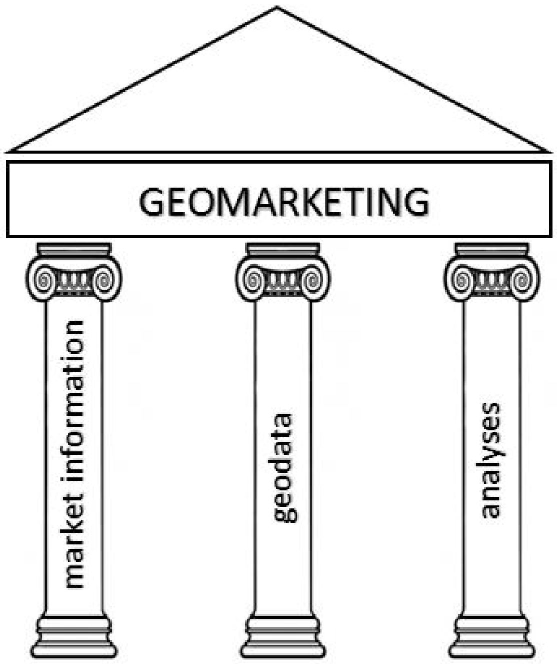


Figure 2: *Basic stacks of Geomarketing*

Market information are qualitative facts about costumers, competitors etc. within a regional (economic) zone. The data contain information about their socio-demographic, psychosocial, economic and consumption properties like income, product affinity, gender and household size [Tap07]. Geodata are information with a spatial reference like addresses, sales areas, locations and catchment areas. The boundaries of these regions may be administrative borders like federal states or townships as well as street sections, coordinates of houses and individual created areas for example a subdivision of postal code areas. By connecting market information to these geodata, analyses are granting important knowledge about the market which are helpful to support companies in their marketing decisions. These facts show that Geomarketing is an instrument for analysing, planning, checking and controlling the market. In the meantime Geomarketing is grown up to one of the most important approaches within the field of marketing to support companies during the accomplishment of their strategies. Consequently it is getting more essential to have systems providing functions and tools which are making these analyses easier and more efficient.

1.2 microm Micromarketing-Systems and Consult GmbH

Microm Micromarketing-Systems and Consult GmbH is a company in Germany which provides Geomarketing analyses to their costumers. It was founded in 1992 and since 1997 it is a subsidiary of the Creditreform. During the last decades microm grew up to one of the biggest providers of Micromarketing and Geomarketing within Germany. It offers possibilities and tools to do anaylses of Geomarketing data. This approach offers the advantage that the company can use all the knowledge which is provided by the employees of microm to control further steps of the marketing decisions. Besides that procedure microm offers additionally a web tool to their costumers so that they can do the analyses by theirself. The software is called mapChart Manager and is accessible with the help of a web browser like Firefox or Google Chrome.

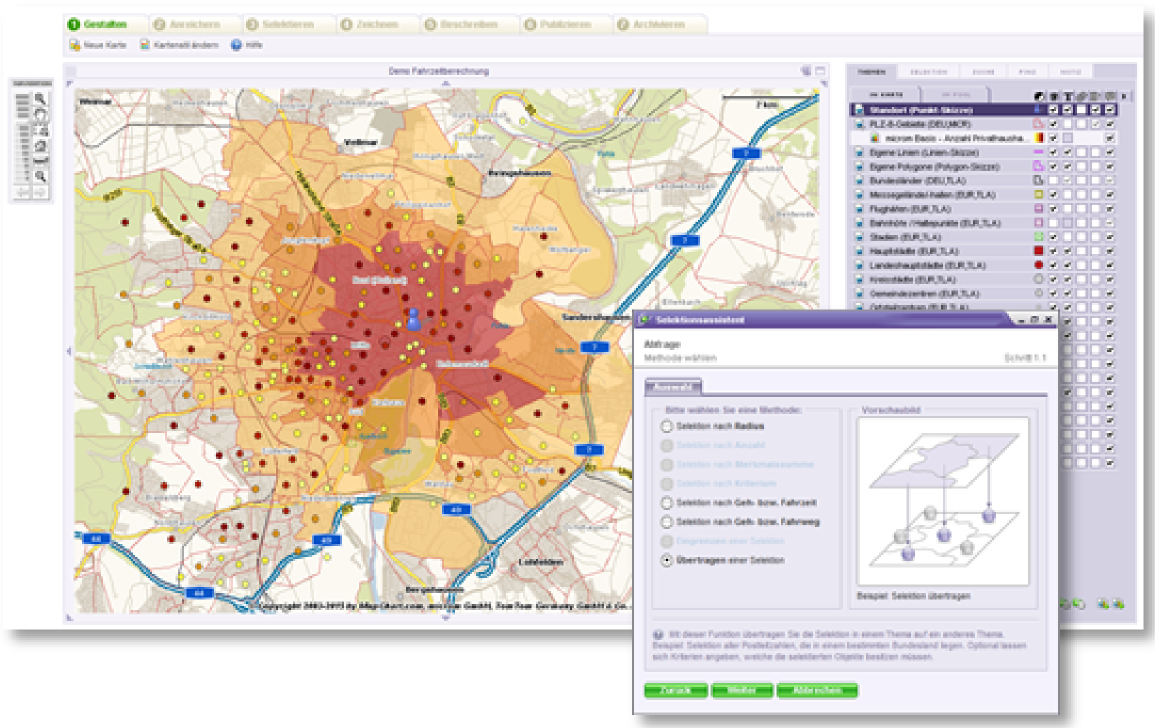


Figure 3: *Screenshot of mapChart Manager*

The advantage of a web tool like the mapChart Manager is that the users can have access to their data and maps from all over the world. Consequently sharing results and working independently from a computer and location makes the application of Geomarketing analyses easier. The mapChart manager offers functionality like the import of data, geocoding of addresses and do anaylses like catchment areas and driving distance zones. To do all that analyses a lot of data is indispensable like routing networks or information about the behaviour of potential costumers. All these data are offered by microm so that their costumers can buy the information they need. Doing so microm profits from their affiliation to Creditreform which collects costumer data from different resources among other things. As a subsidiary of Creditreform microm can use the information which Creditreform have been collected.

1.3 Motivation and Research Question

During the last years the innovation of computer systems raised up so that today a lot of business processes are done by computers to make them more effectiv concerning time and costs. This progress is considerable in the field of Geomarketing, too. More and more software is implemented helping sales managements and business analysts in their decisions. Although already some software packages which provide analyses for Geomarketing strategies are available, the amount of them is really small. Additionally the most of the algorithms containing restrictions concerning several parameters. Per example it may be important that created territories within an area segmentation process become well balanced. Additionally some other constraints may be considered at the same time. The most important condition during area segmentation processes are well balanced, compact and contiguous territories. But no software exists which considers all three parameters. Always just one constraint is kept in mind. That is why the aim of that master thesis is the implementation of an algorithm which considers the three mentioned conditions. Therefore existing algorithms will be considered in detail to determine whether it is possible to find approaches that are promising for the application to Geomarketing strategies. If approaches can be determined the question needs to be answered which algorithm yields the best results concerning balanced, contiguous and compact territories. Afterwards an application of the chosen algorithm will be done to three Geomarketing strategies: area segmentation and optimization, as well as Greenfield analyses and Whitespot analyses.

1.4 Methods

For finding a promising algorithm that can be applied to Geomarketing strategies at first previous work will be considered. Therefore algorithms of the past has been analysed to define approaches which will be implemented. As soon as algorithms are determined they will be implemented in JAVA. The results are displayed with QGIS. Afterwards a comparison of the algorithms will be possible using different defined parameters that concern the predefined conditions from the field of Geomarketing. The result of the comparison will show the most promising algorithm that will be applied to area segmentation and optimization, Greenfield analyses and Whitespot analyses. Therefore the algorithm will be enhanced so that the requirements of the analyses will be satisfied. Afterwards an evaluation of the results is done considering advantages and disadvantages as well as algorithms of related work.

1.5 Outline

Within the last decades several institutes and companies implemented area segmentation algorithms for doing their analyses. Getting an insight into some related works at first three examples will be considered within the following chapter. Some common used termini related to Geomarketing will be introduced in section 3 "Fundamentals of area segmentation". This section also includes notions and use cases which explain processes of area segmentation in more detail. Afterwards the owned knowledge will be used to consider algorithms from past in more detail to make a selection of promising algorithms possible. Comparing different types of approaches supports the election of approaches. The comparison of approaches and the selection of promising algorithms will be done in section 4 ”Selecting approaches for implementation”. As soon as the selection is done the implementation of the algorithm will be done in section 5. Thereby parameters and data are considered first. Afterwards each algorithm will be explained in more detail, showing their procedure and presenting the results. Using these information a comparison of the implemented algorithms will be done in section 6. Consequently the most promising algorithm can be determined to apply it to Geomarketing strategies. The application of the algorithm to area segmentation and optimization, Greenfield analyses and Whitespot analyses will be shown within the following section. Based on the implementation an evaluation will be accomplished in section 8. Finally the work will be summarized to discuss the results and to state a perspective of further work.

2 Related Work

2.1 KIT - Institute of Operations Research: discrete optimization and logistic

The Karlsruhe Institute of Technology (KIT) was founded in 2006 by the amalgamation of the university of Karlsruhe and the research center Karlsruhe. Together both institutes form one of the biggest research and education constitution in the world which concentrates on selected research areas. On the Institute of Operations Research the staff members delve into the field of discrete optimization and logistic which includes districting, too. In progress of studies they have been implemented a library called "Lizard" which includes an algorithm for the solution of area segmentation problems. The library can be downloaded for free, so an execution on given example data is possible. In addition to this, it is implemented into a Geographic Information System in the web browser. The algorithm of Lizard is based on the so called "Recursive Partitioning Algorithm" which was designed by Kalcsics et al. [KNS05]. The algorithm is a geometrical approach which divides the problem into sub problems to align territories that are balanced with respect to an activity value. The partition is done by placing a line into the dataset. Consequently a left problem (data left from line) and a right problem (data right from line) are created. Within the next steps additional lines are placed into the sub problems to divide them again into smaller ones. The number of sub problems depends on the number of distracted areas which should be created. The lines may have different directions which will be considered. These are considered all so that different partitions of sub problems exists. The sub problems are described by a binary tree whose root element complies the problem that should be solved. The different partition for the subdivisions are ranked by a heuristic measure for balance and compactness so that the partition with the best measure value can be chosen as result.

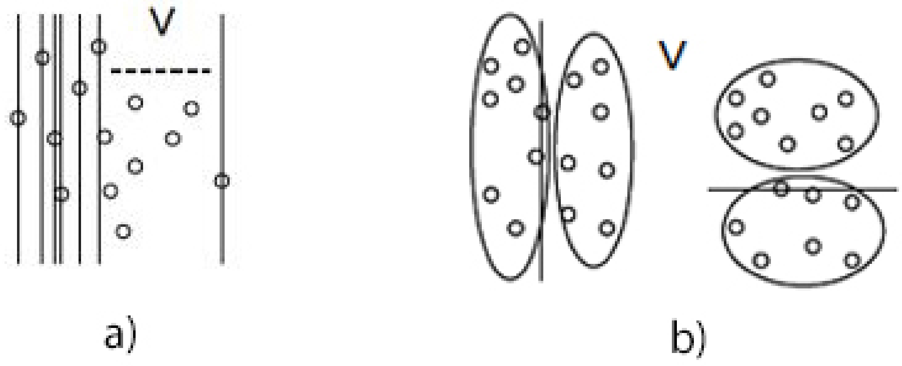


Figure 4: *a) Possible partitions using vertical lines in "Recursive Partitioning Algorithm", b)Two possible partitions with different compactness [KNS05]*

The steps of the partitions are presentable within Lizard. The download package of the algorithm already includes such example files which can be executed. The examples contains either point coordinates or line segments for visualizing street segments. Both input types are linked to an activity measure which will be used to achieve balanced territories. It is recognizable that no given centres exists. Consequently the area segmentation is not done considering given locations like. Before the algorithm in Lizard can be started some parameters need to be chosen. There exist input options for:

- Number of areas: This number defines how many territories will be created.
- Number of line directions: This value is necessary to define the allowed positions of each line which will be placed into the problem for dividing into smaller parts.
- Balance Tolerance: The value determines the range which needs to be satisfied comparing the activity measure of the territories.
- Weight Balance: The value needs to be between 0 and 1 to determine the importance of a balanced activity measure of the territories. In case of 1 balance is most important, in case of 0 the balance is not important. If balance is defined as 0, then compactness is most important.
- Compactness Measure: There exist different options how the compactness value is calculated. The name of some options are ConvexHullIntersection, MaximumDistance and WeightedPairwiseDistance. ConvexHullIntersection for example uses the circumference of the convex hulls of all territories during the calculation of the compactness [PDS14]. Consequently every mentioned option defines another approach of the calculation of the compactness rate.
- Bisecting Partition: With the help of this option the geometric object is selectable which divides the problem. The predefined option are lines like it is defined within the "Recursive Partitioning Algorithm". Additionally FlexZone and a combination of both can be chosen.

The visualisation of the result of the districting problem can be done step by step or by showing just the final result. Showing the results in partial processes represents the approach of the algorithm dividing the problem into sub problems. The calculation of the tree is not recognizable, but the leaf with the best measuring value and thus the best calculation will be shown. Additionally information about other partition solutions are shown, too. In Figure 5 the algorithm is visualized for a area segmentation problem where street segments should be divided into four balanced territories. The feasible balance tolerance is 0.1, the number of search directions is 8, the weight for the balance is 0.5 and the bisecting partition is done by a line. The results of the activity value for every subregions are shown in Table 1.

Blue territory	Red territory	Violette territory	Orange territory
12224	12210	12216	12208

Table 1: *Activity measures of resulting areas after doing an area segmentation using "Lizard"*

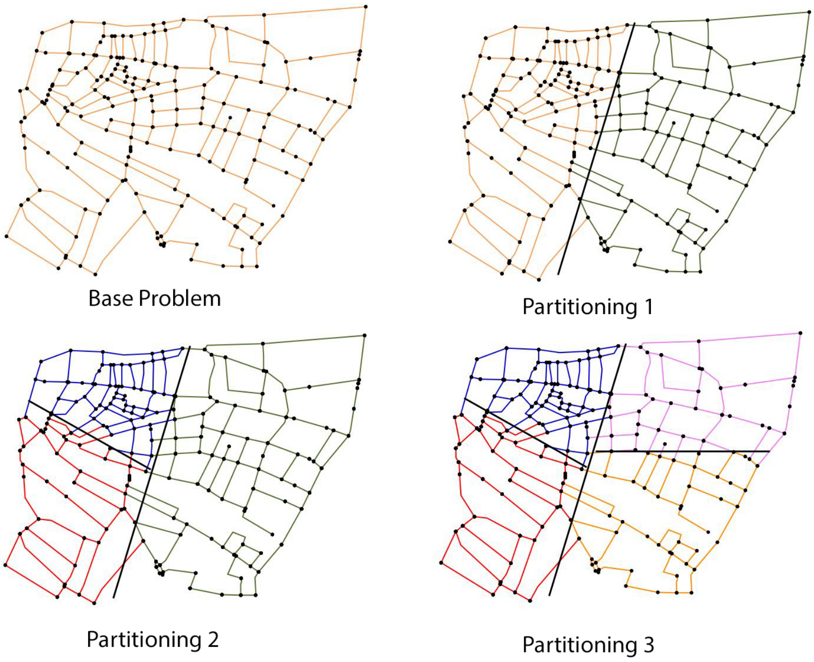


Figure 5: *Visualization of area segmentation using "Lizard"*

Dependent on different settings concerning the available options during the calculations different results can be achieved. Some examples are shown in the following figure. All calculations are done on the same data set which was already used in Figure 5. To emphasize the impact of the input parameters they have been changed. Their results are displayed in the sub figures of Figure 6. In Figure a) the importance of balance was set to 0, instead the compactness of the resulting territories should be most important. In Figure b) the options are chosen the other way around, consequently the compactness is less important and the balance is most important. Figure c) shows the result of the disctricting problem if just one search direction for the lines which were put into the dataset is allowed.

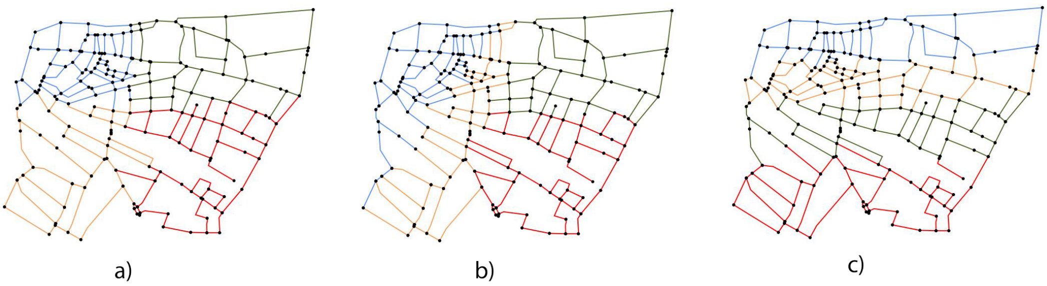


Figure 6: *Results of area segmentation by "Lizard" using different options*

2.2 Easy Map District Manager

The Lutum + Tappert DV-Beratung GmbH is one of the biggest companies in Germany which provides tools, data and services to do Geomarketing analyses. The company was founded in 1982. Already 1986 they sold for the first time a software for doing Geomarketing analyses, which is called Easy Map District Manager. Thenceforth the software was emphasized by a lot of functions. Today Lutum + Tappert provide different attendances from the field of Geomarketing. The main field is the sale and application of their software "Easy Map District Manager". With the help of this data of costumers, potentials of markets and data of the population can be visualized for example. Thus a costumized analysis is possible to realize marketing strategies, control the market and check some statistics. Additionally to the software the company offers data which may be necessary to acquire new costumers or to do more detailed analyses. Furthermore they offer services to help companies during their researches.
The most important tool considering related work is the software Easy Map District Manager. This one is provided in two different versions, which include different options and analysis functionalities. The Easy Map Standard Desktop Edition (EMSDE) is useful to create maps and to do some geographical analyses. However the EMSDE may be necessary to do area segmentation and planning of locations. Both editions are available as a demo version, consequently a deeper insight into the functionality of the software is possible. After the installation of the District Manager some sample data can be used to do some analyses. Therefore zip-code areas of the city of Hamburg and the neighbourhood of it are taken. The tool offers three possibilities to do area segmentation. The first one generates new territories if no distriction was done before. Within the test scenario a given number of territories are created by allocating the zip-code areas to districts. The number of zip-codes areas within every district should be as balanced as possible compared to the others. After the allocation is done locations are set into the middle of the territory to show the position of new company sites. This approach is similar to the Greenfield analyses which will be explained later. The second functionality uses an distribution of areas that already exists. The containing zip-code areas in every superior area will be rearranged if it is necessary to get a balanced area segmentation concerning the number of areas within every district. Existing locations within that dataset are not heeded during the rearrangement. The third option uses just existing locations. With the help of these sites the zip-code areas are allocated in this way that every location is placed within one district. Additionally every district should have the same number of zip-code areas again. These two provided functions are both from the field of the optimization of area segmentation. In reality often additional data are used during the rearrangement for instance the number of households or the purchasing power. These information are used during the area segmentation and are often the parameter which should be balanced. Such a use case was not possible to test because such data are not available for the chosen area. Nevertheless an analysis concerning the number of postcode

areas can be done so that an inspection of the results is possible. It is recognizable that the resulting territories are mostly coherent but there is no need to get an absolutely contiguous district. Especially at borders to other regions some postcode areas are located inside of the neighboured region. Consequently it is not connected directly to the superior district which it belongs to. Thus the conclusion can be achieved that no checking of coherence will be done. By that reason the islands of the north sea do not lead to some problems during the allocation.

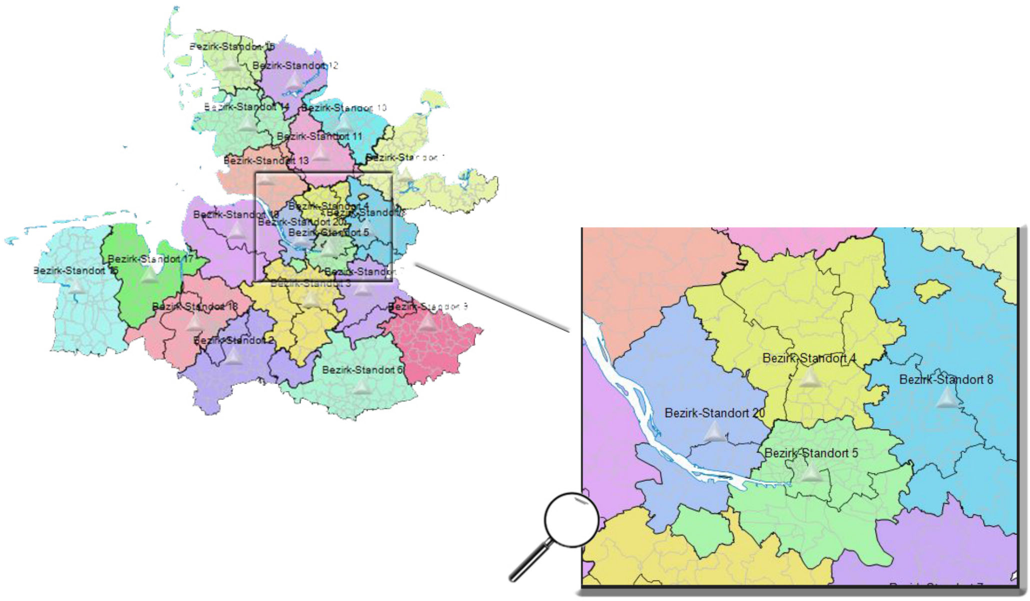


Figure 7: *Result of creating 20 areas with Easy Map District Manager concerning a balanced number of zip-codes areas for each territory. The clipping shows incoherent territories.*

After doing the area segmentation the locations are placed into the middle of the appropriate territory. In this case no other placement can be chosen. Just if an optimization of company sites should be done the locations can be placed into the weighted centroid of the district, too. Furthermore it is recognizable that no new locations can be added to the area where already a certain number of sites exist. But this will be necessary if a new company site needs to be opened. This fact will be described later on with the help of the term Whitespot analyses. Nevertheless the Easy Map Manager offers a lot of possibilities to do some Geomarketing analyses.

2.3 SIM Tool

SIM Tool is a tool for performing analyses of location information and managing existing locations. It was build by cooperation of microm and digital data services GmbH. The tool provides functionality for its users concerning topics that deal with locations of companies. Therefore simulations can be done to consider effects of changes of these sites. Additionally consequences of a movement of a location or a creation of a new one by competitors can be analysed. By specializing topics that deal with locations consequently Greenfield and Whitespot analyses are included as well. Additionally calculations can be done determining the portion of potential at the market.
The SIM Tool is compared to the outcome of that thesis in chapter 8 "Evaluation". Therefore the results of a Greenfield analysis in Hamburg will be considered. For analysing the outcome of area segmentation processes using SIM a calculation is done creating 20 new locations in Hamburg. Therefore zip-code areas are used to be clustered into territories. The result of the calculation is shown in the following figure.
The figure shows that SIM Tool do not check the contiguity of the territories. That is why some zip-code areas are located aside of the territory like it is recognizable at the clipping. Additionally it may happen that the territories are not well balanced. The analyses of the tool show, that no threshold value is used. Consequently the balance of the territories will not be satisfied. Furthermore no rearrangement of zip-code areas is done in order to get a better balance. Nevertheless the results of SIM convince a lot of Geomarketing experts so that they are often used for Geomarketing analyses.

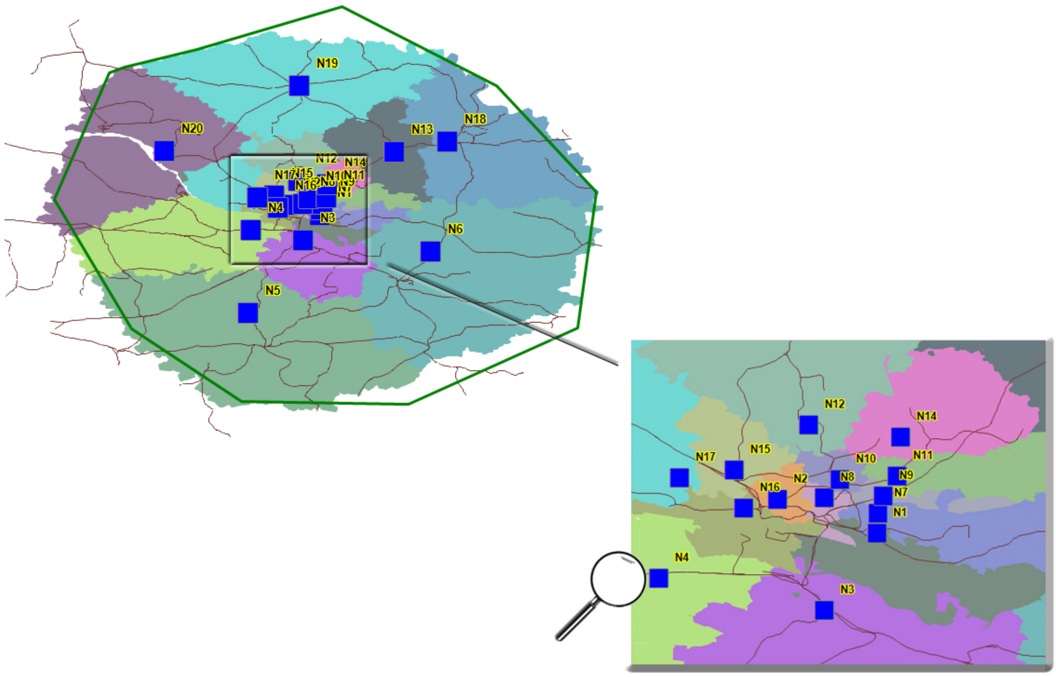


Figure 8: *Results of a Greenfield analysis in SIM Tool creating 20 new locations in Hamburg. The clipping shows incoherent territories.*

3 Fundamentals of area segmentation

Already for several decades people were doing researches in the field of area segmentation approaches. The first important model for area alignment was implemented by Hess et al. [HWS65] in 1965 whose solving a center-seeking political districting problem. Since then a lot of additional researches were done to acquire a lot of more possible approaches and to improve existing ones. The origin which created the need for area segmentation deals especially with marketing aspects or with political distriction. Per example democratic countries need to create constituencies for every election that will be done. Therefore the state or federal state needs to be divided into smaller sub-parts. Consequently an area segmentation approach is necessary. Besides that example area design approaches are often used for sales districting. Within these two special fields different researches exists. But it is recognizable that just a few approaches exist dealing with other aspects from the field of Geomarketing like Whitespot analyses. Nevertheless with reference to these examples it can be deduced that area segmentation is an aggregation of small geographic units into larger clusters in such a way that the latter are acceptable according to one ore more relevant planning criteria [KNS05, ZS83]. The smaller geographic areas are often called basic areas, the clustered units are mentioned as territories. Dependent on the context the relevant planning criteria may change. Per example if an economical context is used the number of costumers or the workload may be balanced. During the distriction several restrictions like compactness or contiguity needs to be satisfied. Such conditions will be explained in section 4.1 "Planning criterias" in more detail.

Considering the researches of the last decades it is recognizable that the most of the acquired approaches are optimization models. Three types of models can be identified: (i) location-allocation approaches, (ii) set-partitioning approaches and (iii) heuristic methods. The location-allocation technique uses two steps to achieve a territory alignment. Therefore no centre points are given in the beginning which will be used for the creation of the clustered territories. Consequently in first step the centres of the territories need to be chosen. This stage is called location phase. Within the second step, the allocation phase, the small geographic units, called basic areas, are assigned to these centres. Both steps are iteratively performed until a satisfactory result is obtained [KNS05]. During the location-allocation stages it is tried to balance a relevant planning criteria. In some cases no centres need to be determined, if there already exist ones. Per example this is the case if a sales districting should be done where already salesman exist whose homes should be the centres of the territories. In such a case just the allocation phase needs to be done. The location-allocation approach was used by Hess and Samuels [HS71], Zoltners and Sinha [ZS83], George et. al [GLW97] and Schröder [Sch01] for instance.

Second there exist so called set-partitioning approaches. These methods are based on the process of generating cantons as a subdivision of all geographical units. These cantons are considered as aspirants to achieve a satisfying area segmentation. After this step is done

a partition of the overall content should be done using such cantons to get a well balanced result [Sch01]. Consequently two steps are necessary again as it was also done in the location-allocation approach. Both steps are performed consecutively or simultaneously as it was done in Garfinkel and Nemhauser [GN70], Nygreen [Nyg88] and Mehrotra et al. [MJN98] for example.
As a third type of optimization models heuristic approaches exist. Contrary to the other two model types this one considers no processes of mathematical programming during the alignment. Such approaches were used by Mehrotra et al. [MJN98], Deckro [Dec77] and Bodin [Bod73] for example.
All three types of models having advantages and disadvantages which make them more or less useful to implement them for Geomarketing strategies. These will be explained in more detail in section 4 "Selecting approaches for implementation".

3.1 Notions and criteria

Each area segmentation process is a subject of several parameters that should be considered. In the case of political restriction one parameter may be a balanced distribution of the population in every created territory. Besides a balanced criteria there exist some more parameters and conditions which can be borne in mind. In the following section the most typical parameters will be explained in more detail. However at first some important notions are mentioned which are correlated with every territory alignment approach.

3.1.1 Basic areas

Every area segmentation problem consists of a set V of areas to which the alignment should be done. These areas are geographical objects in the plane consequently there are maybe points (e.g. addresses of costumers), lines (e.g. street-sections) or polygons (e.g. districts). These areas are called *basic areas*. Let B denotes a basic area then:

$$B_1 \cup B_2 \cup ... \cup B_n = V$$

In the case of that thesis all implementations will be done to zip-code areas which will be represented by polygons. Further information can be found in section 5 "Implementation of area segmentation approaches". Each basic area is linked with one or more quantifiable attributes which should be considered during the distriction. That attributes may be the number of households, the workload or the purchasing power.

Figure 9: *Zip-code areas from Dresden as basic areas which may be used in an area segmentation process*

3.1.2 Number of territories

The number of territories defines how much higher ranking areas will be created using the basic areas. The number of territories will be given by the user in case of that thesis. Within some approaches like it was done in Kalcsics et al. [KNS05] the number of territories was handled as a planning parameter. Let B denotes the basic areas again and T denotes a territory then:

$$B_1 \cup ... \cup B_n = T_1 \text{ and } B_{n+1} \cup ... \cup B_m = T_2 \text{ etc.}$$

$$T_1 \cup T_2 \cup ... \cup T_n = V$$

3.1.3 Territory centers

Usually a created territory is associated with a territory centre. If the area segmentation was done without predefined centres, they are often located afterwards into the geographical centre of the territory. In other cases centres already exist. These will be the origin from that the alignment will start. Let T denotes the territories again and Z denotes a territory centre then

$$Z_1 \in T_1 \ , \ Z_2 \in T_2 \ , \ ... \ , \ Z_n \in T_n$$

3.1.4 Unique assignment of basic areas

A determined condition during the application of an area segmentation is the unique assignment of basic areas. This means that every basic areas is allocated to exactly one territory. Consequently no basic area exists which is not assigned to any territory. At the same time no territory exists which shares one basic area with another territory. Let B denotes a basic area and T denotes a territory consisting of several basic areas then:

$$B_1 \cup ... \cup B_n = T_1 \text{ and } B_{n+1} \cup ... \cup B_m = T_2 \text{ etc.}$$

$$T_i \cap T_j = \emptyset, i \neq j$$

3.1.5 Additonal Planning criterias

Area segmentation processes are done to group small geographical units into larger clusters. Thereby often at least one planning criteria will be considered during the alignment. The planning criteria depends off the context for which the territory distriction will be done, consequently there exist several different criteria. The most common will be explained in that section.

3.1.5.1 Balance

In the most of the cases a activity measure value should be balanced during the area segmentation process. Consequently the created cluster needs to be created in such a way that the sum of the activity measure of the containing basic areas is similar comparing to the other territories. Let B denotes the basic areas, then the total activity measure of one territory can be formulated as:

$$w(T_i) = \sum_{B \in T_i} w_B$$

Considering results of area segmentation processes a perfectly balance of the activity measurement cannot be accomplished. This is caused by the discrete structure of the problem and the unique assignment assumption [KNS05]. The activity measure value depends on the context for which the planning distriction should be done. Either it may be economical origin like purchasing power and working load or it comes from demographic origin like the number of households and the number of population. Furthermore several activity measure values can be considered at the same time too. Instead of an activity measure it also may be possible to regard just the size of the area. Per example every territory should contain a similar number of zip-code areas.
Two examples of different balanced territories are shown in Figure 10. The territories in

Figure a) are more balanced considering an activity measure then the ones in Figure b). The area segmentation process is done to zip-code areas from Dresden and its neighbourhood. The used zip-code areas are more detailed as the common ones which contain 5 numbers, for more information about the used data see section 5 "Implementation of area segmentation approaches". In this case the used activity measure is the number of households.

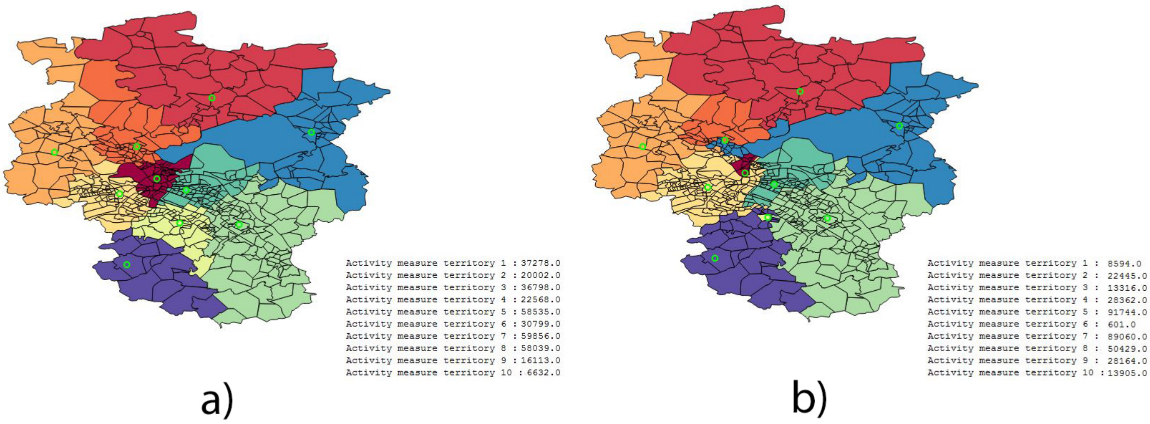


Figure 10: *Comparison of different results concerning balance of activity measure. a) Better balancing of activity measure. b) Poor balancing of activity measure*

3.1.5.2 Contiguity

Often a constraint of the created territories is the contiguity of them. This means that every basic area that is contained in a territory is directly neighboured to one other basic area of the territory set. Therefore explicit neighbourhood information of the basic areas are required. For satisfying the condition neighbourhood graphs may be used in the algorithm. In Figure 11 coherent territories are confronted with not coherent territories.

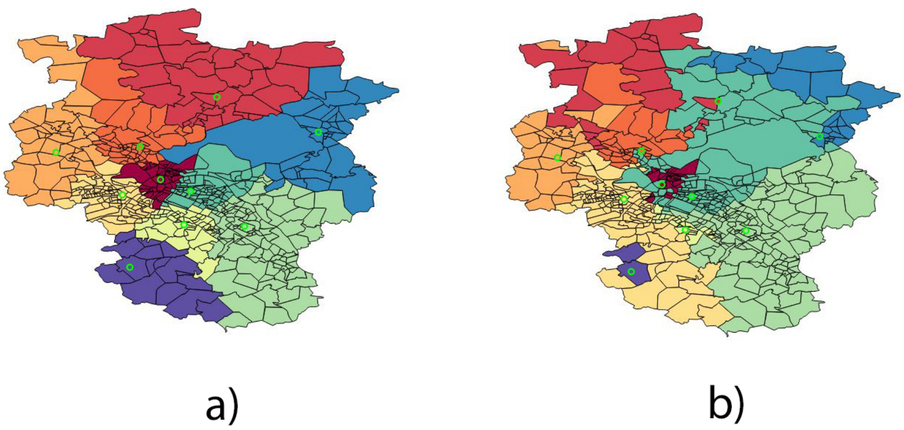
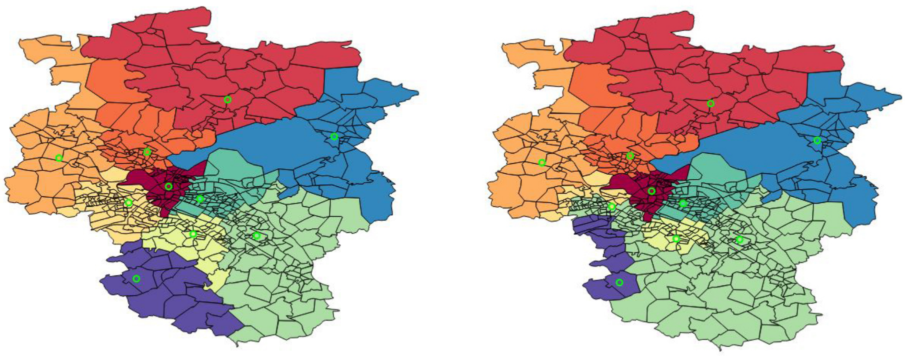


Figure 11: *Comparison of different results concerning contiguity a) Coherent territories. b) No full coherence of territories.*

3.1.5.3 Compactness

In economical and demographical context the resulting territories often need to be compact. In such cases compactness describes a "fine" and "meaningful" shape of the territory boundary, per example the shape ought to be round or quadratic. Young defined a compact figure as "homogeneous and in a limited space located figure which is not scattered over a spacious area" [You88]. In case of an economic context a round shape of the territories is defined as compact to minimizes the access route starting from the territory centre per example. The compactness can be calculated with the help of different compactness measures like the compactness measure of Cox or the one of Harris [Kö12]. Additionally the weight of the euclidean distance can be used to define the compactness rate. Figure 12 illustrates two examples with different compactness of the territories.



a) b)

Figure 12: *Comparison of different results concerning compactness a) More compact territories. b) More scattered territories.*

3.2 Use cases

3.2.1 Political Districting

Within determined time intervals elections are done within a country, federal state etc. to vote for persons who wants to be the representatives of it. Therefore the area have to be divided into sub-parts, so called constituencies. Every constituencies nominees one candidate who will be elected directly into the parliament, this districting problem is called the "one man-one vote" problem. A democratic election is based on the same weight of every voting that is why some restrictions have to be followed during the political distriction. In Germany these conditions are set down into the Federal Electoral Law §3 Art. 1. It determines that the creation of constituencies should be done in this way that the number of constituencies within the federal state should agree with the part of the population [DB]. This means that every constituency should hold a similar number of voters compared to other constituencies. The number of inhabitants of Germany is used as stipulation to satisfied that condition. During the distraction the boundaries of townships, districts and cities should be preserved as much as possible. Considering different uses cases of political distriction three essential characteristics of districts can be defined:

1. The created constituencies should have nearly equal populations in order to respect the principle.
2. The created constituencies should be coherent.
3. The created constituencies should be geographically compact.

Before an election can be carried out the existing constituencies have do be proofed and adapted if it is necessary because local alteration of the population can be recognized over the time. A commission will do this in front of every voting. If an adaptation of the constituencies is necessary an area segmentation approach needs to be applied. Figure 13 illustrates the political distraction in 2013 for the parliamentary elections in Germany.

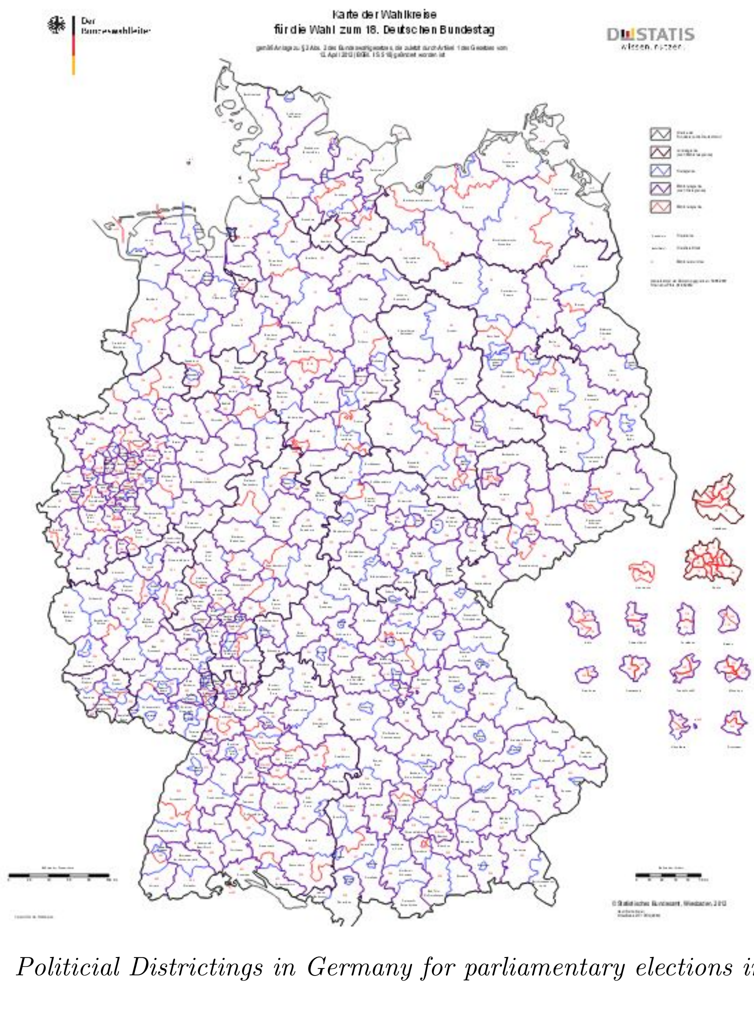


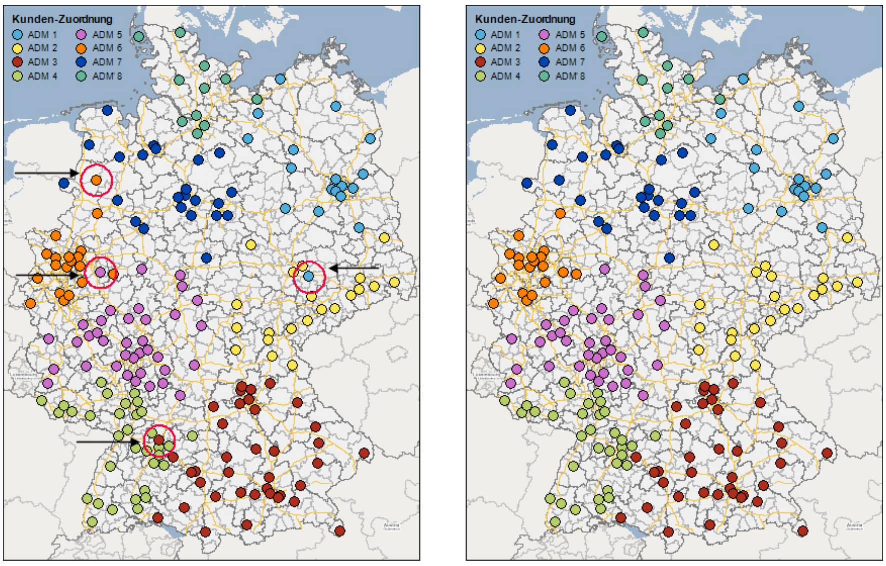
Figure 13: *Politicial Districtings in Germany for parliamentary elections in 2013 [DB]*

3.2.2 Sales Districting and optimization

One of the most common tasks of marketing organization is the arrangement of areas of operation for the branch offices as well as the organisation of the costumer care within these areas of operation. In every company which deals with branch offices this planning task is important and needs to be done periodically. During the planning process several territories are created which contain costumers. For each territory one salesman is responsive for. Historically the sales districting was done with the help of a paper map, some pins, strings and adhesive labels. Until today some companies are doing it this way. The fixedness of that approach is obvious, additionally it is really fault-prone during a new alignment or a rearrangement of the sales distriction.

Figure 14: *Historical sales distriction with map, pins and strings [Tap07]*

More and more companies recognize the advantages of software tools which offer funcionality for sales distraction. During the alignment and optimization several criteria may be kept in mind. It is necessary to know how much territories should be created (dependent on the number of salesman) per example. Furthermore the territories need to be compact and coherent to minimize the travelling time for each salesman. In addition to cost and time saving by having well formed sale territories a lot of more advantages are given by that approach. Per example an unequal workload or territory potential compared from one salesman to another can lead to low morale, poor performance, a high turnover rate, and an inability to assess the productivity of individual territories or districts [HS71]. Additionally some costumers may be unattended by too much workload of a salesman or by no definite allocation to one territory. With the help of visualising existing costumer data an incorrect assignment of costumers to a salesman can be demonstrated. These ones may be owing to a historically growth because every salesman amplifies its stock of costumers. By a reallocation of that costumer data the territories can be redefined so that the work load and the travelling time can be optimized.



a) b)

Figure 15: *Redefinition of affiliation of costumers. a) Faulty allocation of costumers. b) Redefined assignment. [ibm]*

Additionally to workload, travelling time etc. the potential value of every territory should be balanced. As it is shown in the following figure this is not the case in Figure a). Consequently an area optimization needs to be done to balance the potential value of both areas.

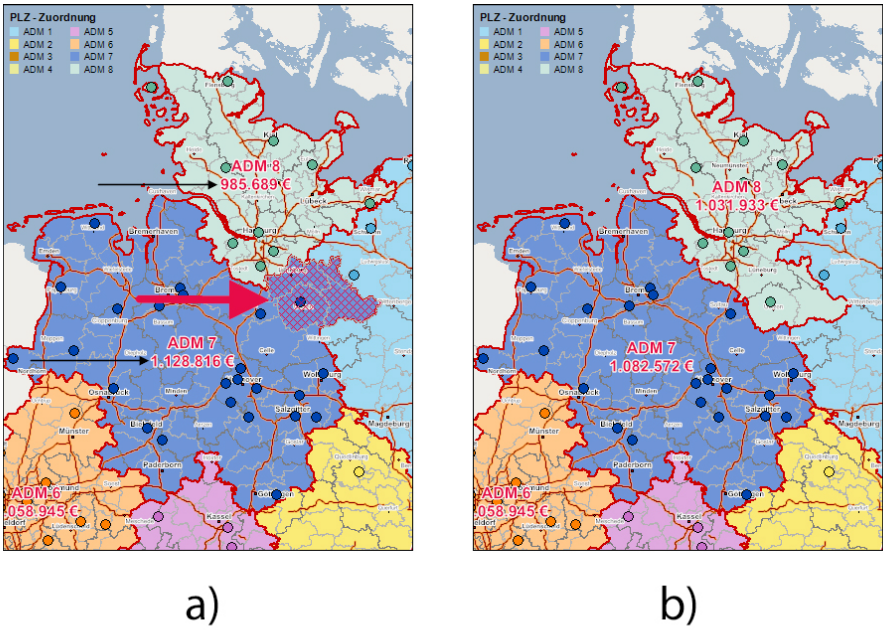


Figure 16: *Rearrangement of basic areas to get a blanced result. a) Existing situation of unbalanced territories. b) Territories after optimization. [ibm]*

3.2.3 Whitespot analysis

The most of the companies aspire for expansion to raise their profit. For this step it is necessary to know where unexploited potentials of markets exist. Therefore the so called white spots of a map need to be determined - this process is called whitespotanalysis. Every white spot contain costumers which may not be bonded to a competing company so that they receive the potential for more profit of a company. The following figure shows an example of the natural gas network in Germany. The white spots are communities which are not connected to the integrated network. Consequently these areas are interesting for other thermal care systems like oil or coal.

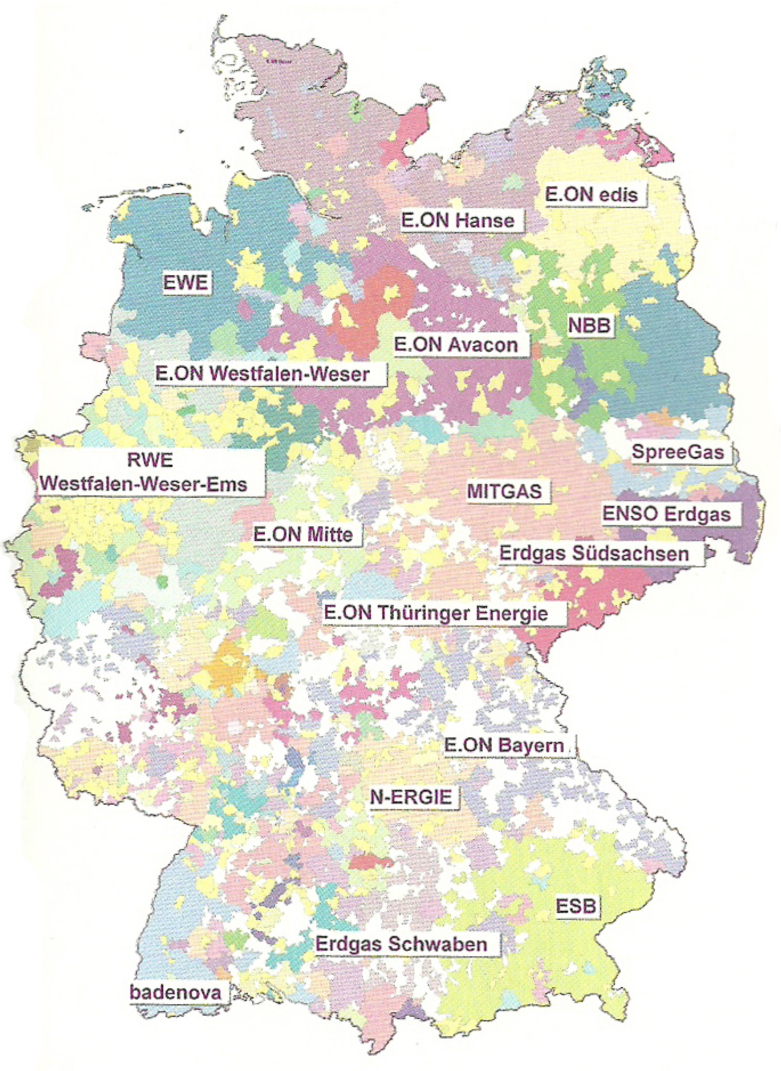


Figure 17: *Map of natural gas network in Germany. The white spots are communities that are not connected to the integrated network. [Tap07]*

4 Selecting approaches for implementation

4.1 Requirements of approaches considering Geomarketing analyses

For the application of area segmentation approaches to Geomarketing strategies some dedicated requirements are need to be kept in mind to get an useful result. Possible planning criteria where already explained in detail in section "Additonally Planning criterias". It was mentioned that contiguity, balance and compactness may play a huge role during the distriction. Considering Geomarketing aspects all three planning criteria need to be satisfied. Taking the example of sales territories owing a certain number of salesman shows that a not balanced alignment yield to an imbalance concerning workload of the employees. This could lead to disaffection. Additionally the territories need to be compact to make the sale process more effective by minimizing travelling time.
The area segmentation in this case study will be done to already existing territory centres consequently a determination of new centres in the beginning will be not necessary.
All these requirements and conditions needs to be kept in mind during the comparison of model types and the selection of promising algorithms for the implementation.

4.2 Comparison of model types

In section 3 "Fundamentals of area segmentation" it was mentioned that there exist three different types of models which where applied in history for distriction processes. The first model type are location-allocation approaches and are the ones that are mostly used. Within the location-allocation process in the location phase territory centres are chosen. In the second step the basic areas are assigned to these centres. Both steps are done iteratively until a satisfying result of the area segmentation is achieved. For selecting approaches for the implementation within that master thesis that method will be analysed in more detail to determine whether it will be used for that thesis or not. The first model of the location-allocation approach was developed by Hess et al. [HWS65] in 1965 to solve a political distriction problem. Due to application of the model to sales districting the model of Hess et al. [HWS65] was enhanced by Hess and Samuels [HS71] and has established known as "GEOLINE" model. After the implementation they admit that their "GEOLINE" approach "does not provide optimal sales territories" [HS71]. Additionally the gained solution may be not well balanced and coherent. Consequently the practical use of this model is fairly limited. That is why Fleischmann and Paraschis [FP98] developed a modified solution of the "GEOLINE" model, but their approach demonstrates the same disadvantages like the one of Hess and Stuart [HS71]. The difficulty of the solution of the capacitative transportation problem is the assignment of portions of basic areas to more than one territory centre to satisfy the balancing constraint. Consequently these so called split areas require a more detailed consideration. Hess and Stuart [HS71] tried to achieve a well balanced solution using a so called AssignMAX approach.

Within the alignment they had district the split areas to the territory centres which "own" the largest share of the split area [HS71, KNS05]. But Fleischmann and Paraschis [FP98] had proofed that this approach leads to very poor results for their application. Consequently they tried to implement an improved solution of the alignment of split areas but the solution could not resolve all splits automatically thus it was necessary to do some manual post processing [FP98, KNS05]. Due to these problem different improvements were developed to find a good split solution. Schröder [Sch01] tried to find an optimal split resolution using tree decomposition. With the help of that approach the best area segmentation in the system of the equations can be determined. Contrary to solving a transportation and split problem Zoltners and Sinha [ZS83] implemented an approach using sub-gradients. The advantage of that method is the calculation of several possible area segmentation solutions. From this amount of solutions the best one can be chosen. But comparison to the approach implemented by Schröder [Sch01] show that the calculation time using the approach of Zoltners and Sinha [ZS83] is higher then the one using Schröders algorithm [Sch01]. Additionally Zoltners and Sinha [ZS83] approach needs well distributed territory centres to achieve a good solution of the calculations. Consequently the algorithm indeed provides coherent areas, but the resulting territories may not be well balanced. Considering all mentioned approaches it can be recognized that no one yields to optimal results considering the area segmentation process. Either the created territories are not coherent or they are not well balanced. Additionally due to the complexity of the split resolution the calculation time is still to time consuming for solving large scale problems. [KNS05]. Furthermore the location-allocation approach is not owing to the linear terms of the equations that need to be answered. Consequently the application of measures of compactness will be constrained considerably. Considering all these problems location-allocation approaches seem to be inapplicable for the application to Geomarketing strategies. Consequently such approaches will not be considered in that thesis anymore.

The second type of model is called set-partitioning approach and was implemented by Mehrotra et al. [MJN98] for example. During the set-partitioning method subdivision of all geographical units will be created. These ones will be used to get a balanced result. Compared to location-allocation methods a major advantage of set-partitioning is the higher flexibility concerning a satisfying result of the area segmentation. In contrary to only limited use of criteria in the location-allocation methods in this approach any criterion can be applied during the generation of candidate districts [KNS05]. Nevertheless at the same time this advantage is a disadvantage too because the huge flexibility causes a raising combinatoric complexity. That is why the set-partitioning approach can be only used for smaller problems. It have not been used with more than 100 basic areas [KNS05]. Compared to the location-allocation approaches this method is more ineffective, cumbersome and computationally unattractive [ZS83]. Considering that statement it is obvious that the set-partitioning approach can not be applied to Geomarketing strategies to achieve satisfying results. Additionally Geomarketing analysis are mostly done to huge area segmentation problems containing a lot of basic areas.

Consequently the set-partitioning method can be seen as unusable for that application.
The third type of models are heuristic approaches. These ones do not need any solver for linear problems like set-partitioning and location-allocation approaches do. Instead, just some mathematical programming is necessary. The main advantage of heuristic methods is the huge flexibility concerning the integration and observance of one or more criteria. At the same time a forecast of the quality of the created territories may be difficulty previously. Normally the quality will be measured afterwards comparing different solutions. Consequently just a relative rating of the quality is possible [Sch01]. Nevertheless compared to the disadvantages of location-allocation and set-partitioning methods heuristic approaches seems to be the most promising ones to apply for Geomarketing strategies. By comparing different heuristics it will be tried to find an algorithm with a high quality of alignment results. To make the algorithm more dependable different heuristics will be combined. Additionally some completely new heuristic approaches will be implemented. In Table 2 an overview of the advantages and disadvantages of all three types can be found.

	location-allocation	set-partitioning	heuristics
short explanation	first step: chosing centres, second step: assigning basic areas to the centres	using subdivisions and partition	mathematical solution of problems
advantages	• chooses best solution of several calculations	• high flexibility	• huge flexibility • do not need a solver • easy to implement • usable for one or several criteria
disadvantages	• territories may be not well balanced or not coherent • calculation is time still too high caused by split areas • just one criteria applicable	• useable just for small problems	• difficult to determine a forecast about the quality of the result previously

Table 2: *Comparison of different types of models applied for solving area segmentation problems*

4.3 Heuristic approaches

During the last decades several heuristic algorithms were implemented to solve area segmentation problems. The common ones will be considered in that section to choose the approaches that will be used within this thesis.
Mehrotra et al. [MJN98] for example developed an algorithm called **Eat-up**. During the Eat-up approach one territory after the other is extended at its boundary by adding yet unassigned basic areas to the territory successively. This will be done until the territory satisfies the criteria that need to be balanced. The algorithm was implemented for political distriction with the goal "to develop a districting method that provides population equality and contiguous and compact districts while retaining jurisdictional boundaries of counties or other political subunits insofar as possible" [MJN98]. Within their case study of political distriction in South Carolina they conclude that their implemented algorithm is "an effec-

tive way of generating high quality districting plans" [MJN98]. Consequently the Eat-up approach may be one of the potential methods used in that master thesis.
Deckro [Dec77] implemented an algorithm called **Clustering**. That approach treats each basic area initially as a single district. After creating a ranking of neighboured basic areas pairs of districts are merged together iteratively so that a new bigger territory will be created. During the creation of the districts a particular criterion is considered and needs to be satisfied within a range of acceptable variation [Dec77]. The districts are merged together until the number of prescribed territories is reached. It was not possible to find further examinations of the usability of that approach. Although it seems to be promising it is not applicable for the area segmentation processes that needs to be done within this thesis. The reason for this is that the distinction will be done using existing territories centres. But the clustering algorithm do not use any centres.
The Multi-kernel growth approach was used for example by Bodin [Bod73] in 1977. This method is made up of two steps. In the first step a certain number of basic areas are determined as centres of the territories that should be created. After this step to each centre neighbouring areas are successively added. The neighboured areas are added in order of decreasing distance to the centre. The alignment of basic areas is done until the desired territory size is reached. The second step of this approach is similar to the **AllocMinDist** method which was implemented by Kalcsics et al. [KMNG02] where the basic areas are allocated to the closest territory centre like it is done in the Multi-kernel growth approach. They conclude that the AllocMinDist algorithm leads to "disjoint, compact and often connected, however, usually not well balanced territories as the balance criterion is completely neglected when deciding about the allocation" [KMNG02]. Nevertheless they have mentioned that "the attractiveness of this method [...] lies in its simplicity and computational speed" [KMNG02]. Consequently these results are also adaptable to the second step of the multi-kernel growth algorithm. The first step can be ignored because in the application of that thesis territory centres are already given. Owing to the mentioned advantages of the AllocMinDist algorithm this one may be useful during combining different heuristic approaches. That is why it will be tested by implementation. Additionally it will be proofed whether the same results of the AllocMinDist algorithm are achieved like in the application of Kalcsics et al. [KMNG02].
Another developed approach for optimization is the so called **local search**. Within that method the basic areas of neighbouring territories are shifted to minimizing a weighted additive function of different planning criteria [KNS05]. Consequently an area segmentation needs to be done at first so that the local search can be applied to the solution. With the focus on this local search may be a promising method for combining different heuristics. Per example at first the allocMinDist algorithm may be applied, accordingly the local search approach.
Another approach that needs to be mentioned is the algorithm which is implemented within the Lizard Library of KIT (see section 2.1 "KIT - Institute of Operations Research: discrete optimization and logistic" for more information). The **Recursive Partitioning Algorithm** divides a problem into smaller sub problems until each sub problem satisfies the considered

criteria. A similar approach was implemented by Forrest [For64]. Problems of the Recursive Partitioning Algorithm were already mentioned in section 2.1 so that this approach will not be used within this thesis. Although the created territories seem to be well balanced, they not need to be coherent after the calculation. Additionally the algorithm is based on no territory centres, but the Geomarketing application which will be considered first provides existing centres.

Considering all mentioned approaches it is recognizable that just some methods seem to be promising while other do not. In conclusion three approaches will be used during the comparison. These approaches are Eat-up, AllocMinDist / Multi-kernel growth and local search. Additionally some further algorithms are developed. Related to the AllocMinDist approach the same will be done considering only the given criteria that should be balanced. This approach is called AllocCrit in the following. In addition to this several approaches are implemented combining different methods. The first one is appropriated to the AllocMinDist algorithm which will be combined with the consideration of the activity measure of the basic areas. Therefore the basic areas are assigned to every territory centre iteratively considering the activity measure of the centres. That means that the centre with the smallest activity measure gets the nearest basic area that is not assigned yet. These steps will be done for each territory centre until all basic areas are allocated. This approach will be called SmallestCrit-GetsNearest. An improvement of that approach is called SmallestCritGetsTrueNearest which considers neighbouring relationships in more detail. The procedure of the algorithm is similar to SmallestCritGetsNearest but the nearest neighboured basic area is just allocated if it is really the nearest one to this territory centre. If it is not the case another basic area will be taken.

Researches during the last decades show that algorithms that raise the size of the territories from centre may yield to bad results concerning the coherence of the territories. That is why algorithms were implemented which are using a starting point. The starting point of alignment lies at the boundary of the observation area. Consequently such a heuristic is implemented within that master thesis too. But this one is combined with the SmallestCritGetsNearest algorithm to yield to better results. Because of the combination of two heuristics the algorithm contains two steps. In the first one all outer territories which lie on the border are allocated to territory centres if they are nearest to one centre. To consider also territories which may be allocated to different territory centres, similar to split areas, a coefficient is used, which should be satisfied for allocation. If the assignment of outer territory centres is done the second step will be initialized which is the SmallestCritGetsNearest approach. The whole algorithm will be called OutsideSmallestCritGetsNearest. The next chosen algorithm is a combination of the Eat-up and the AllocMinDist approaches. Within the Eat-up algorithm each territory centre gets basic areas until the termination criterion build up from the planning criteria is reached. This allocation determines no relationship which basic area will be chosen for the assignment. Within the improved algorithm just the basic areas are allocated which are nearest to the territory centre. The algorithm will be called EatUpMinDist. Furthermore to the mentioned algorithms a combination of AllocMinDist and local search will be implemented too. The algorithm contains two steps similar to the OutsideSmallestCritGetsNearest approach. At first the AllocMinDist processes will be applied. Afterwards the assigned basic areas are rearranged again using local search to create a balanced resolution. Consequently this algorithm is called AllocMinDistLocalSearch All algorithms are implemented with the aim to find an usable calculation procedure for area segmentation processes for the implementation to Geomarketing strategies. All approaches will be compared to find the most promising one. For summarizing the used algorithm they will be shown in Table 3.

	Short explanation
AllocCrit	Assigns basic areas to each territory centres dependent on the activity measure of the centre achieving by the sum of the activity measures of the assigned basic areas.
AllocMinDist	Assigns the basic areas to the territory centre which is closest.
Eat-up	Extends one territory centre after the other at its boundary by adding yet unassigned basic areas to the territory successively.
SmallestCritGetsNearest	Combination of AllocCrit and AllocMinDist: Assigns the closest yet not assigned basic areas to each territory centre dependent on the activity measure of the centre.
SmallestCritGetsTrueNearest	Improvement of SmallestCritGetsNearest: Process is similar to SmallestCritGetsNearest but the allocation of the closest basic area will be done only if it is the closest territory centre at all.
OutsideSmallestCritGetsNearest	Assigns first basic areas at the boundary of the area under investigation. Accordingly SmallestCritGetsNearest is applied.
EatupMinDist	Combination of Eat-up and AllocMinDist: During the Eat-up approach just basic areas are allocated which are close to the considered territory centre.
AllocMinDistLocalSearch	Combination of AllocMinDist and local search: At first AllocMinDist will be applied. Afterwards a rearrangement is done using local search to achieve a well balanced solution.

Table 3: *Overview about selected heuristic approaches for implementation*

5 Implementation of area segmentation approaches

The comparison of different optimization models in the section before shows that heuristic approaches seems to be the most promising methods for the application of that master thesis. That is why 8 different heuristic approaches were be chosen which will be implemented. The results of all algorithm will be compared to find the most usable algorithm of that amount.
Section 3.1 "Notions and criteria" shows the fundamentals and necessary notions to apply to area segmentation. They will be discussed in the following chapter for visualizing the used data of the application.
Basic areas are geographical objects in plane which represent the areas that should be assigned to territory centres. In the case study basic areas are zip-code areas. There exist two different types that will be used:

1. common zip-code areas with 5 numbers, for instance: 01159, 48143. In the following these areas are mentioned as zip-code5. The used investigation area contains 37 zip-code5 areas.

2. subdivision of zip-code5 areas which is done by the microm to make more detailed analyses possible. These areas contain 8 numbers, for example the zip-code5 area of 01159 consists of 25 zip-code8 areas e.g. 0115903, 0115904, 0115912. In the following these areas are mentioned as zip-code8. The used investigation area contains 564 zip-code8 areas.

Each zip-code area is defined by coordinates to represent the shape and orientation in space. Additionally each area is linked with the activity measure. The activity measure will be household numbers. During the calculation just one measurement will be balanced. The calculations are applied to zip-code areas of the city of Dresden, Saxony and surrounding regions.

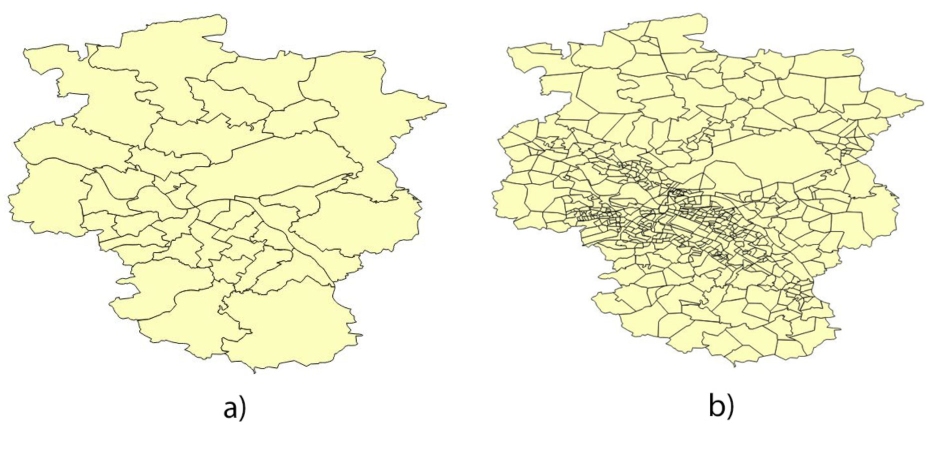


Figure 18: *Zip-code areas from Dresden and surrounding regions as basic areas which are used within the area segmentation. a): zip-code5 areas, b): zip-code8 areas*

The territory centres are given in that case study. The centres are a point data set provided by microm. Each centre represents a branch of the bank "Sparkasse". Consequently the number of territories is predefined. The locations of Sparkasse are used to make the area segmentation process more realistic so that a better analysis of the results may be possible. For the calculations 10 locations of Sparkasse are used, that are well destributed on the investigation area. Caused by the distribution it is recognizable that the comparison of the approaches is first done to an optimal starting situation. This is done to rate the quality of the algorithm first under optimal situations. Later on further researches are necessary to determine the real quality of the chosen algorithm in the end.
The area segmentation should achieve a well balanced result with coherent and compact territories. To check the compactness of each territory the compactness measure of Cox will be used. The decision to take this measure is based on a research by Köhler [Kö12]. This ones shows that the Cox compactness measure is the best choice to get an satisfying ratio of calculation duration and quality of the calculated compactness. That result was confirmed by Butsch and Ludwig [BL14], too. The compactness measure of Cox is the ratio of the area of a territory compared to the area of a circle which has the same circumference. The resulting value is

$$cp(D_j) \leq 1$$

The nearer the value of c_p to 1, the better the compactness of the territory is.

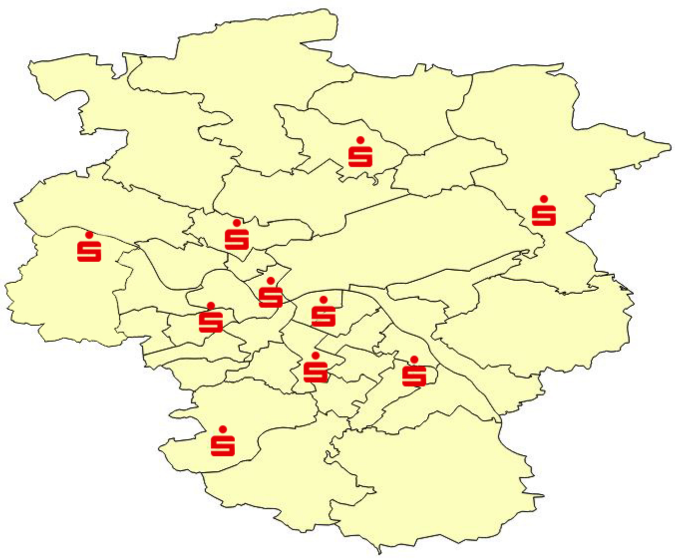


Figure 19: *Used territory centres of Sparkasse showed on zip-code5 areas from Dresden*

For doing the calculations some software package are necessary. The implementation is done in JAVA using Eclipse. The basic areas and appropriate planning criteria are stored within a

PostgreSQL/PostGIS database. The visualization of the areas is done using Quantum GIS. Table 4 summarizes the used data.

	Detailed information
Basic areas	37 zip-code5 areas and 564 zip-code8 areas
Investigation area	Dresden, Saxony and surrounding regions
Territory centres	Locations of Sparkasse provided by microm
Activity Measure	Household numbers
Planning criteria	Balance, contiguity, compactness

Table 4: *Summary of used data within the implementation*

5.1 AllocCrit

The AllocCrit algorithm assigns basic areas to each territory centre dependent on the sum of activity measure which each centre contains. Remember the formulae

$$w(T_i) = \sum_{B \in T_i} w_B$$

to achieve a balance of the territories for each one the activity measure of the contained basic areas in that territory are summed. Consequently the main focus of AllocCrit lies on a balanced resolution. Doing so the positions of the basic areas are not considered. The calculation is applied on zip-code5 and zip-code8. The procedure and the results of the AllocCrit area segmentation are shown in the following figures and in Table 5.

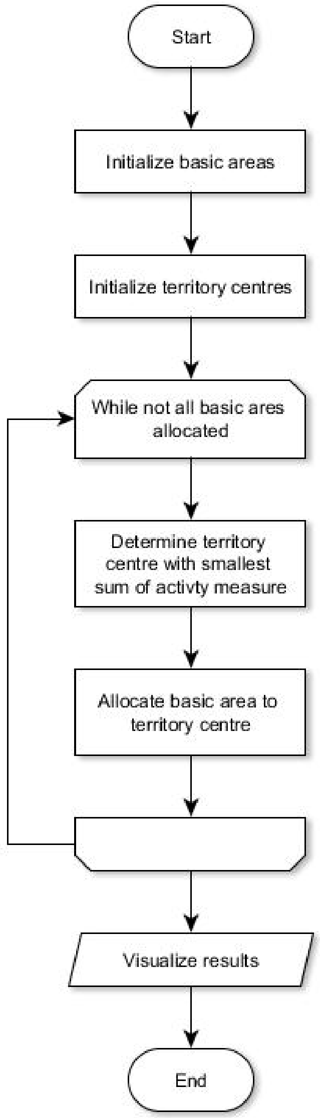


Figure 20: *Workflow of AllocCrit algorithm*

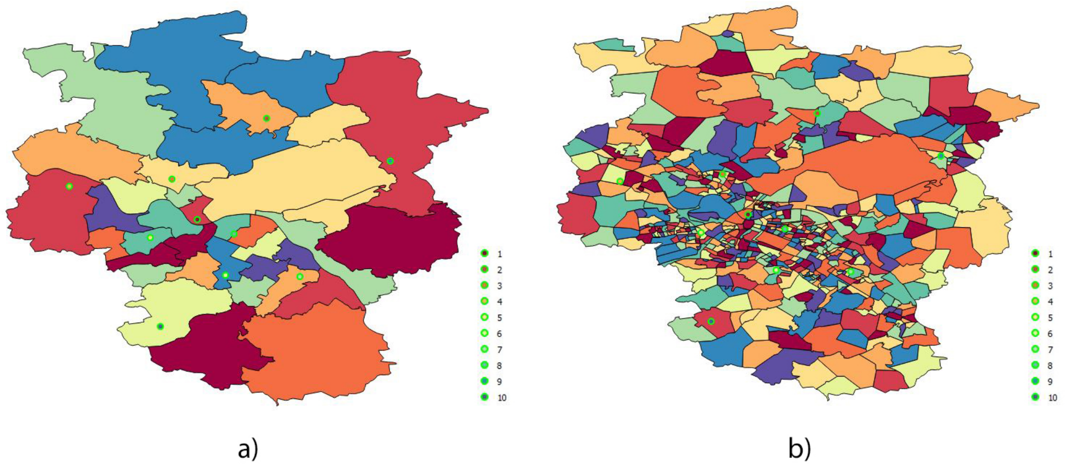


Figure 21: *Result of area segmentation process using AllocCrit. The territory centres are visualized as circles. a) zip-code5 areas. b) zip-code8 area.*

	number of basic areas		sum of activity measure		compactness	
	zip-code5	zip-code8	zip-code5	zip-code8	zip-code5	zip-code8
Territory 1	4	56	30856	34716	0.11054	0.01424
Territory 2	4	53	35985	34605	0.08989	0.01652
Territory 3	3	57	41154	34608	0.17331	0.01909
Territory 4	4	59	39397	34572	0.08514	0.01366
Territory 5	4	60	35316	34628	0.21168	0.01479
Territory 6	3	54	33055	34608	0.13450	0.01724
Territory 7	5	55	30980	34578	0.06247	0.01654
Territory 8	3	57	36382	35074	0.24263	0.01345
Territory 9	4	56	31466	34657	0.11085	0.01490
Territory 10	3	57	32029	34574	0.13489	0.01308

Table 5: *Results of area segmentation using AllocCrit*

The results of the AllocCrit algorithm shows a patchwork rug of territories. Although the activity measure of the territories is balanced they are not contiguous and compact. The reason for the result is the violation of orientation of the basic areas. They are just considered like the order in database is. If the order within the database will be change, also the results of the territories will change. But never a compact and contiguous shape of the territories will be achieved. To obtain such regions a consideration of the orientation needs to be done. Caused by the order of basic areas within the database the solution is not reproducible all the time. As soon as the order changes, the result changes, too. Consequently the approach is not usable for Geomarketing strategies. Nevertheless a well balanced resolution is recognizable. Considering zip-code5 areas the difference between the sum of activities is larger than in the case of zip-code8 areas. By the finer subdivision of regions the basic areas in zip-code8 can be assigned better in such a way that the balance is ideal. However, aberration are possible all the same because the balance is not linked to a threshold. Instead just an allocation is done without knowing the activity measure of the following basic area. Consequently if the activity measure of e.g. the last basic area is too huge compared to others the sum of the territory to which it will be allocated may be much larger than the sums of the activity measure within the other territories.

5.2 AllocMinDist

The AllocMinDist algorithm is based on the distances between the territory centres and the basic areas. Dependent on the distance each basic area will be assigned to that territory centre which is closest to it. Within previous researches different approaches are determined how the calculation of distances can be done. One possibility is to use euclidean distances, which is the linear distance between territory centre and basic areas. On the other hand quadratic euclidean distances are often used in researches for example considering Fleischmann and Paraschis [FP98], George et al. [GLW97] and Hess et al. [HWS65]. Although quadratic euclidean distances are used more often than euclidean distances in this case study the second one will be used for calculations. The reason is provided by Marlin [Mar81]. He concludes that the "success of squared Euclidean distances depends on the ability to redefine territory centres"[Mar81]. But in this case study the centres are fixed, so that an application of quadratic euclidean distances is not possible. That is why quadratic euclidean distances are used mostly for political districting problems whereas euclidean distances are used for sales territory design problems. The case study of that master thesis is similar to sales territory design problems, consequently euclidean distance can be used here, too.
The workflow diagram and the results of the AllocMinDist algorithm are shown in the following figures and in Table 6.

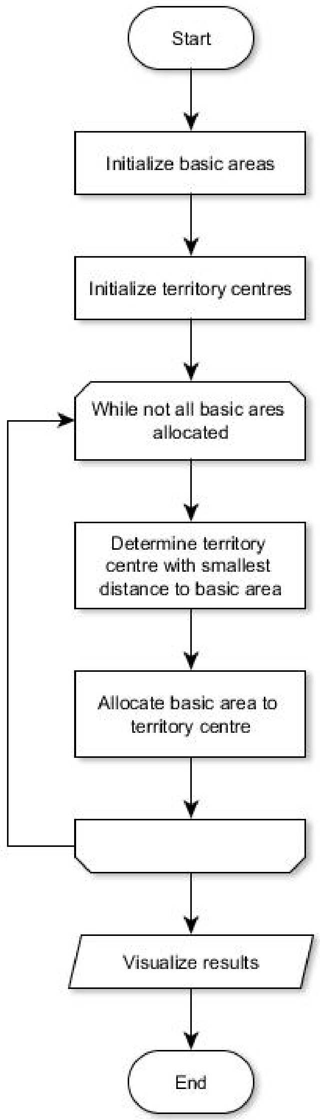


Figure 22: *Workflow of AllocMinDist algorithm*

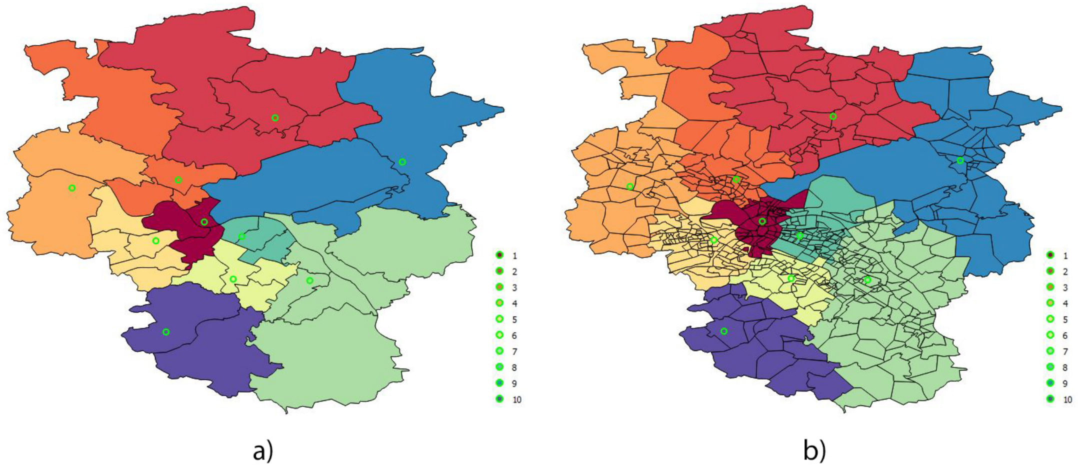


Figure 23: *Result of area segmentation process using AllocMinDist. The territory centres are visualized as circles. a) zip-code5 areas. b) zip-code8 areas.*

	number of basic areas		sum of activity measure		compactness	
	zip-code5	zip-code8	zip-code5	zip-code8	zip-code5	zip-code8
Territory 1	3	53	29587	37112	0.30051	0.24085
Territory 2	5	33	23964	20570	0.27904	0.27857
Territory 3	4	49	33033	37072	0.12038	0.21838
Territory 4	2	43	23390	23700	0.27541	0.15193
Territory 5	5	86	51925	56054	0.30495	0.27057
Territory 6	4	48	37216	30058	0.29541	0.34401
Territory 7	6	123	56230	65734	0.20121	0.26552
Territory 8	3	75	49971	52067	0.45405	0.37194
Territory 9	3	38	33964	17274	0.21720	0.18477
Territory 10	2	16	7340	6979	0.29587	0.29500

Table 6: *Results of area segmentation using AllocMinDist*

The accomplishment of the AllocMinDist algorithm yields to contiguous and at first sight compactness territories, but they are not well balanced. By considering just the distance the probability of creating contiguous territories raises up a lot. But nevertheless the coherency is not compulsive. In some case, especially if a basic area has a unusual shape, incoherent territories may be created. The compactness for that case study is quite well although no "round" territories are formed. The reason for this is shape and adjustment of the basic areas. Consequently the calculated compactness value is not quite good. However, showing these results to persons how planned sale territories in the past, they confirmed a good assignment of the territories. But the sum of activity measure is not balanced because the activities were not considered during the calculation. The range of the different sums is too huge. Consequently the difference of the number of basic areas containing each territory is quite high, too. Caused by these two reasons an application of that algorithm to Geomarketing analyses is not promising if no adaptation will be done.

5.3 Eat-up

The Eat-up algorithm extends territory centre one after another by adding yet unassigned basic areas to the territory successively. Consequently one territory centre is taken and basic areas are aligned to it until a threshold value is reached. The threshold value is calculated by creating the sum of all activity measures and dividing it by the number of territories that will be formed. Let t donates the threshold and N the number of territories, this calculation can be formulated as

$$t = \frac{\sum_{B \in V} w_B}{N}$$

As soon as the threshold is reached, the next territory centre is taken, if not all centres were already used.
The workflow of the Eat-up algorithm and the results will be shown in the following figures and in Table 7.

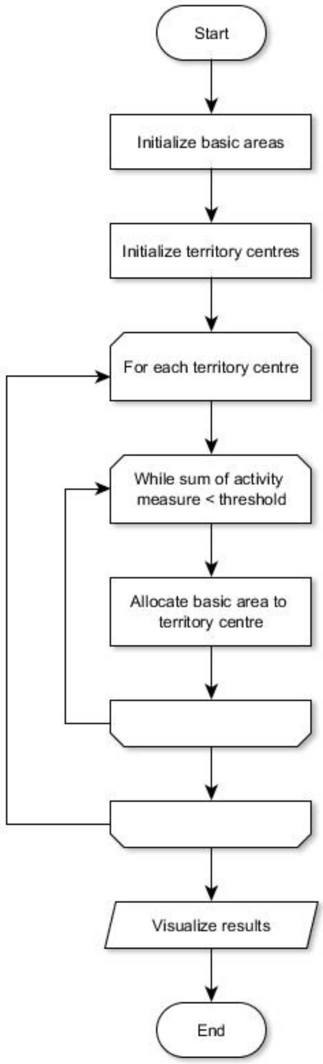


Figure 24: *Workflow of Eat-up algorithm*

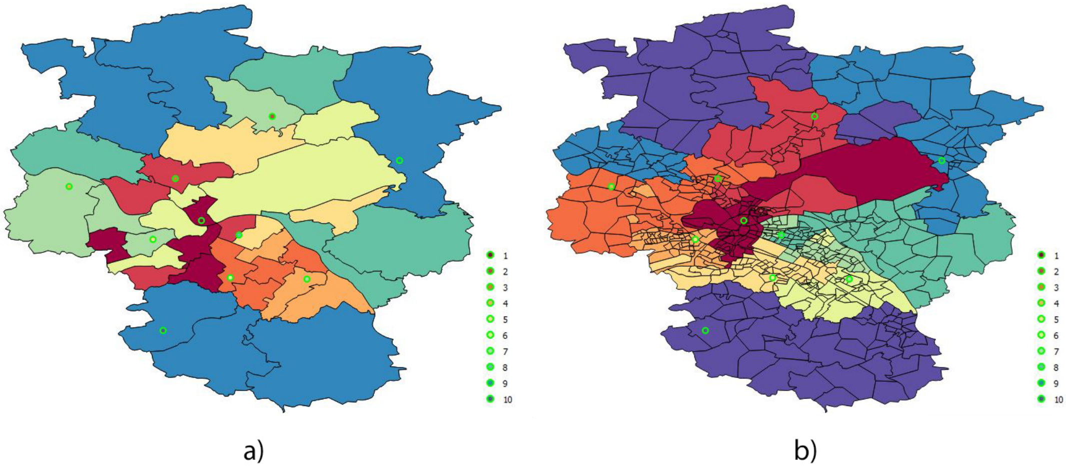


Figure 25: *Result of area segmentation process using Eat-up. The territory centres are visualized as circles. a) zip-code5 areas. b) zip-code8 areas.*

	number of basic areas		sum of activity measure		compactness	
	zip-code5	zip-code8	zip-code5	zip-code8	zip-code5	zip-code8
Territory 1	4	51	41726	35656	0.12937	0.15024
Territory 2	4	46	39196	34778	0.12697	0.14228
Territory 3	4	53	40119	35281	0.22537	0.15231
Territory 4	3	61	37972	34882	0.18226	0.13125
Territory 5	3	51	34748	34922	0.11567	0.14501
Territory 6	5	58	43818	35240	0.14354	0.20280
Territory 7	4	53	35620	35459	0.16096	0.18325
Territory 8	4	56	35143	35113	0.11095	0.20593
Territory 9	6	59	38278	35266	0.08428	0.11736
Territory 10	0	76	0	30023	null	0.11216

Table 7: *Results of area segmentation using Eat-up*

The outcome of Eat-up shows two different results. Looking at the resolution of zip-code5 areas it is obvious that the algorithm leads to a bad result concerning balance, compactness and contiguity. While using zip-code8 areas the resolution seems to be betters concerning the distribution of the basic areas. This causes the impression of using a promising approach. Looking in more detail to the results makes an exact evaluation possible. Concerning the coherence of the territories in both use cases (zip-code5 and zip-code8 areas) shows that it is not given always. Instead of this the territories may be disconnected. This is caused by ignoring any neighbouring relationships and distances. The basic areas are just allocated to a territory centre depending on their order within the database. If the order will change, also the resolution changes so that the result may not be comprehensible anymore. In worst case a patchwork rug of basic areas belonging to one territory centre may be created. In the case of the case study often basic areas are neighboured in database which are neighboured in reality, too. That is why the territories are coherent partly. Besides the problem of the contiguity and consequently the compactness also the balance may be not satisfying. One problem is shown in the calculation using zip-code5 areas. Territory 10 does not have any allocated basic areas. This phenomena is caused by assigning basic areas to the previous territory centres until the threshold is reached. Doing so the threshold value will be exceeded all the time so that the termination criterion is reached. Caused by the rough classification of zip-code areas all basic areas will be already assigned to a centre although not all territory centres are considered yet. In case of zip-code8 areas this phenomena does not occur because the basic areas are classified finer. Considering the two main problems of no coherence and the dangerous of no satisfying balance the conclusion of non-applicability is obvious if no adaptations are done to the existing algorithm.

5.4 SmallestCritGetsNearest

The SmallestCritGetsNearest algorithm combines the approaches of AllocCrit and Alloc-MinDist. Consequently it allocates the closest basic area to the territory centre that owns the smallest sum of activity measures. To do so at first all sums are compared to determine the centre with the smallest value. After this step the comparison of distances to yet not assigned basic areas can be done. The basic area which is closest will be allocated afterwards. The segmentation is done until all basic areas are allotted. By combing two heuristics an approach that considers balance as well as compactness and coherence is requested to find. The workflow of the algorithm and the results are shown in the following figures and in Table 8.

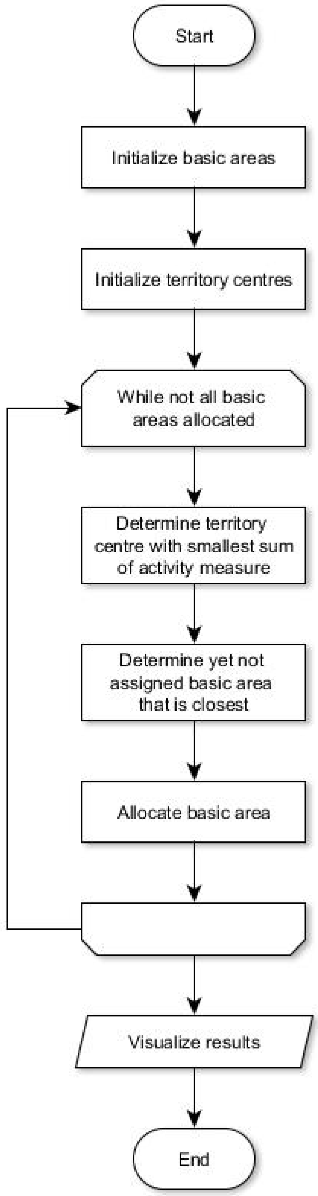


Figure 26: *Workflow of SmallestCritGetsNearest algorithm*

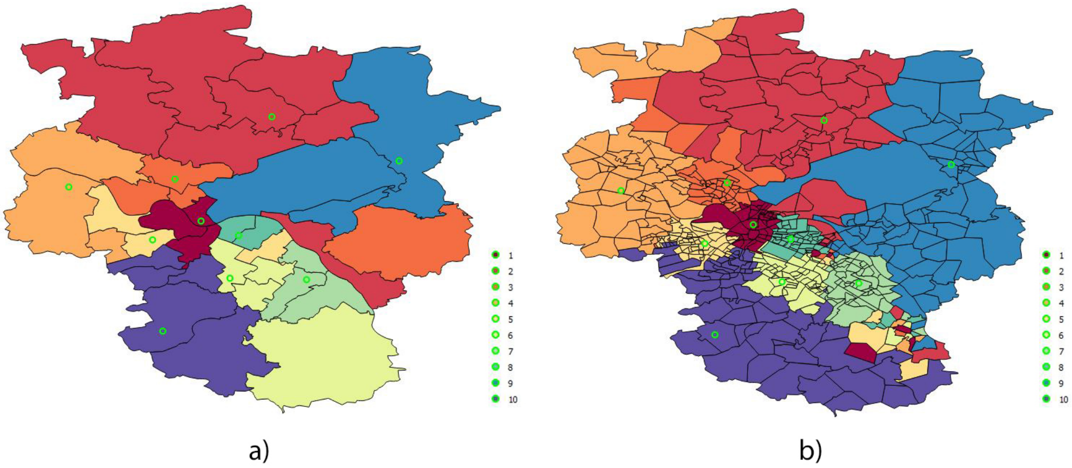


Figure 27: *Result of area segmentation process using SmallestCritGetsNearest. The territory centres are visualized as circles. a) zip-code5 areas. b) zip-code8 areas.*

	number of basic areas		sum of activity measure		compactness	
	zip-code5	zip-code8	zip-code5	zip-code8	zip-code5	zip-code8
Territory 1	3	51	29587	34548	0.30051	0.10532
Territory 2	7	50	33432	34645	0.15803	0.14166
Territory 3	4	48	35890	34716	0.15647	0.09506
Territory 4	3	59	33794	34830	0.21891	0.13064
Territory 5	3	59	45353	34549	0.14055	0.07482
Territory 6	4	55	40275	34974	0.17941	0.15727
Territory 7	3	56	31945	34477	0.28958	0.30115
Territory 8	2	51	32140	34469	0.61214	0.14139
Territory 9	3	71	33964	34753	0.21720	0.15885
Territory 10	5	64	30240	34659	0.23357	0.12474

Table 8: *Results of area segmentation using SmallestCritGets-Nearest*

The results of SmallestCritGetsNearest algorithm shows that the territories may be not compulsively contiguity. In both cases incoherence exists. Although the closest basic area will be allocated to a territory centre the assignments that are already done are not considered. If there are basic areas of another territory in between the closest basic areas which is not assigned yet will be chosen, also if it is far away from the territory which will get it. Consequently an examination of contiguity is necessary if coherent territories need to be achieved. Thus the compactness of the territories can be neglected, too. Nevertheless the balance is quite good. But there exist restrictions, too. A huge discrepancy may occur during the last allocations because the activity measure is not considered previously. After the last assignments the sum of activity may be too huge of the territories which got the last basic areas. This danger existing as not as much in case of zip-code8 areas caused by the finer partitioning of the basic areas.

5.5 SmallestCritGetsTrueNearest

The SmallestCritGetsTrueNearest algorithm is an improvement of SmallestCritGetsNearest. The main part of the algorithm is similar to SmallestCritGetsNearest. Consequently the first procedures are the same. After initializing basic areas and territory centres the basic areas will be assigned to the centres. Therefore in each step the territory centre with the smallest sum of activity measure is determined. Afterwards the not yet allocated basic area which is closest to the centre will be detected. Contrary to SmallestCritGetsNearest this one is not allocated immediately. Instead it is checked whether the basic area is located closer to another territory centre. If so, a basic area will chosen which is already allocated but closer to the territory centre than the not assigned one that was taken first. Consequently the basic area is removed from the territory it was allocated previous and it is added to the territory with smallest sum of activity measure. According to this the territory with smallest sum will be determined again and the processes is starting afresh. These steps are done until all basic areas are allocated.

The workflow of the algorithm and the results are shown in the following figures and in Table 9.

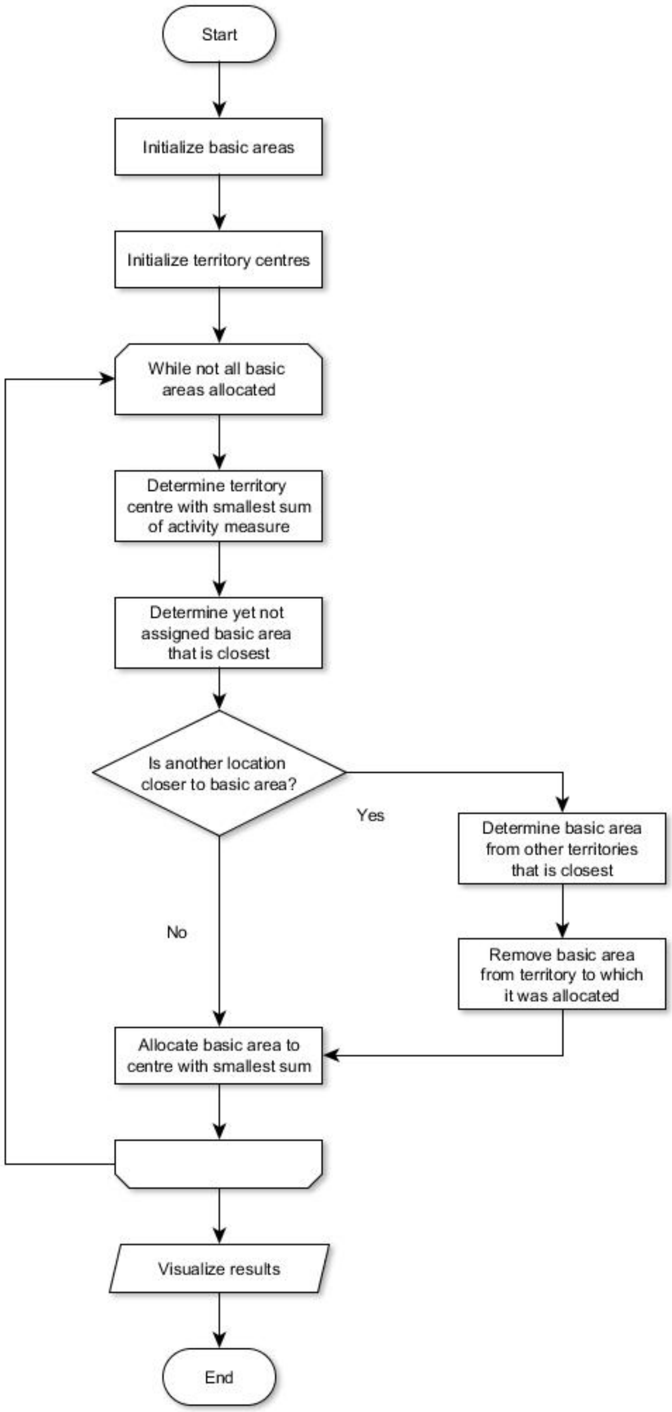


Figure 28: *Workflow of SmallestCritGets-TrueNearest algorithm*

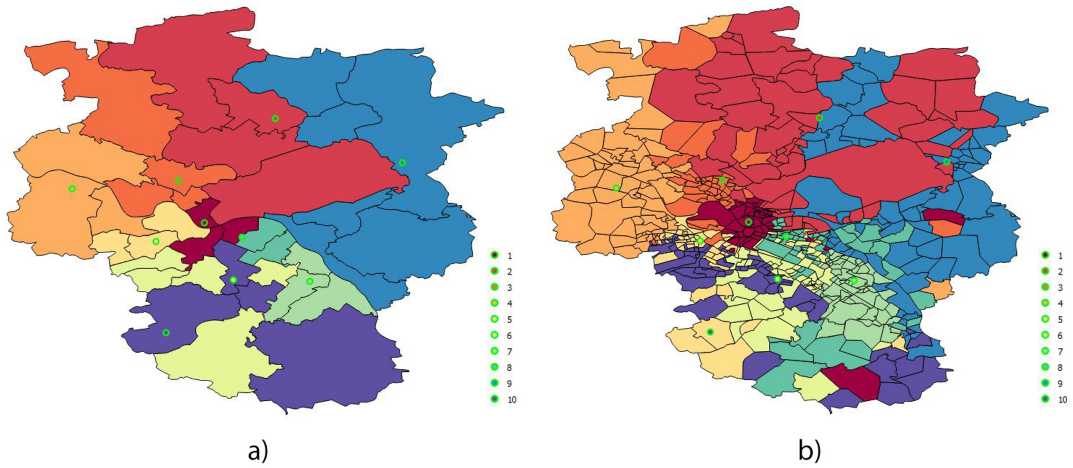


Figure 29: *Result of area segmentation process using SmallestCritGets-TrueNearest. The territory centres are visualized as circles. a) zip-code5 areas. b) zip-code8 areas.*

	number of basic areas		sum of activity measure		compactness	
	zip-code5	zip-code8	zip-code5	zip-code8	zip-code5	zip-code8
Territory 1	3	51	35771	34712	0.18579	0.11388
Territory 2	4	53	34267	34954	0.19696	0.08049
Territory 3	4	48	33033	34508	0.12038	0.05897
Territory 4	3	59	35046	34771	0.24743	0.10476
Territory 5	3	54	33436	34552	0.29346	0.04600
Territory 6	5	59	33542	34584	0.11221	0.04356
Territory 7	3	58	31945	34638	0.28958	0.27358
Territory 8	2	57	36621	34719	0.42244	0.03948
Territory 9	6	67	35986	34314	0.18329	0.03754
Territory 10	4	58	36973	34868	0.12917	0.04157

Table 9: *Results of area segmentation using SmallestCritGets-TrueNearest*

The results of SmallestCritGetsTrueNearest shows that the hoped-for success of the improvement does not occur. The created territories are not coherent. Instead of this a patchwork rug of territories is recognizable. Although the sums of criteria of each territory are well balanced the algorithm is not usable caused by the not contiguous and compact territories. The balance is just given conditionally because no check of it will be done at the end. Consequently if the basic areas that are allocated last own a huge activity measure the discrepancy of the sums may be large, too. Another problem occurs by trying to allocate to the nearest territory centre. It is shown in the workflow diagram that an examination is done whether the basic area is closest to the territory centre or not. If the check ends up with the conclusion that another territory centre is the closest one, a change of the basic areas is done. But this change leads to problems. In some cases an endless rearrangement of basic areas occurs because in one step a basic area is taken from one territory and in the next step the same area is returned to the territory. This process loops infinite. Consequently a workaround is needed to solve that problem. In this algorithm it was declined to change the same basic area again. Other approaches were tested too, but either the contiguity or the compactness are affected negatively. Consequently this approach needs a necessary improvement so that an advisable area segmentation can be done.

5.6 OutsideSmallestCritGetsNearest

The OutsideSmallestCritGetsNearest algorithm is as well as SmallestCritGetsTrueNearest an extension of SmallestCritGetsNearest. The developed algorithm consists of two different steps. First all basic areas are checked whether they are allocated within a specific distance to one territory. This means it is checked whether the distance of one territory centre and the basic area is small enough so that it will be allocated clearly to that territory. During the comparison of the distances of basic area to all territory centre a threshold value is used which can be adapted individually. Per example if 30% is used as threshold value all basic areas are allocated to territory centres which are 30% closer to one centre compared to the others. After doing the pre-allocation of some basic areas the second step starts which is similar to the approach of SmallestCritGetsNearest. Consequently all basic areas that are not allocated yet will be assigned iteratively to the territory centres dependent on the sum of activity measures of the centres. This will be done until all basic areas are distributed.
The workflow of the algorithm and the results are shown in the following figures and in Table 10.

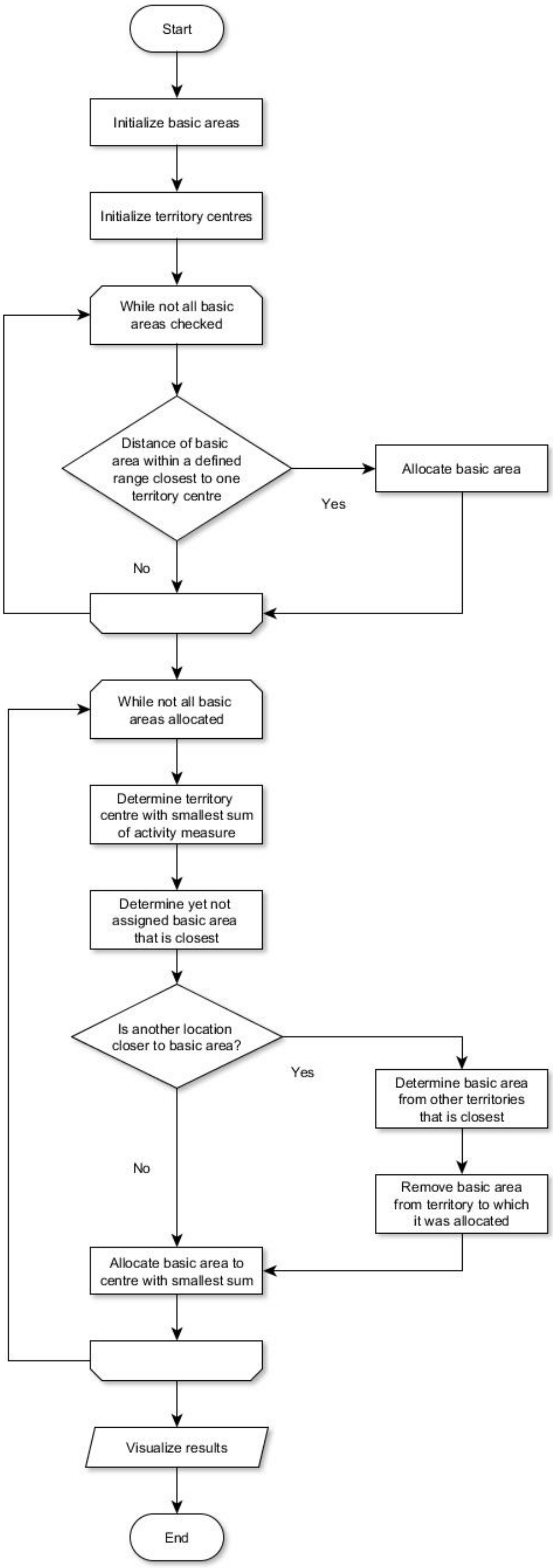


Figure 30: *Workflow of OutsideSmal-lestCritGetsNearest algorithm*

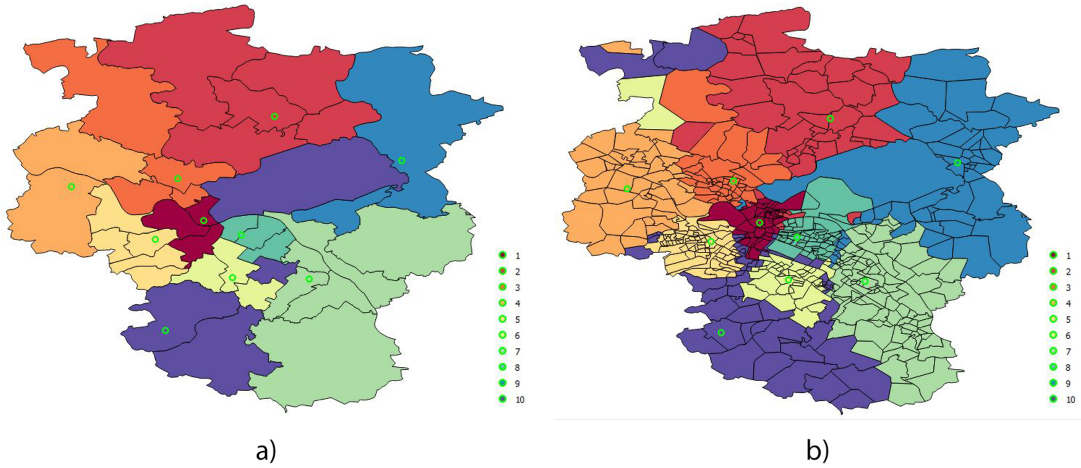


Figure 31: *Result of area segmentation process using OutsideSmallestCritGetsNearest. The territory centres are visualized as circles. a) zip-code5 areas. b) zip-code8 areas.*

	number of basic areas		sum of activity measure		compactness	
	zip-code5	zip-code8	zip-code5	zip-code8	zip-code5	zip-code8
Territory 1	3	44	29587	32979	0.30051	0.18142
Territory 2	5	43	23964	27017	0.27904	0.15807
Territory 3	4	42	33033	32103	0.12038	0.22956
Territory 4	2	43	23390	26360	0.27541	0.26590
Territory 5	5	74	51925	47419	0.30495	0.34101
Territory 6	3	42	28647	26189	0.25768	0.13975
Territory 7	6	102	56230	54735	0.20121	0.29949
Territory 8	3	64	49971	46342	0.45405	0.32603
Territory 9	2	55	16706	27160	0.15201	0.14998
Territory 10	4	55	33167	26316	0.13556	0.06752

Table 10: *Results of area segmentation using OutsideSmallestCritGetsNearest*

The results of OutsideSmallestCritGetsNearest shows that the created territories are not persuasive concerning balance, contiguity and compactness. It is obvious that the territories are not coherent in all cases. Depending on the threshold value used during the allocation of the basic areas which are within a determined range of distance to one territory centre the contiguity may be better or worse. But if the threshold value is changed, this affects the balance too. The smaller the threshold value, the worse the balance will be. The calculations were done to a threshold value of 30%. During the assignment of basic areas in first step the sum of activity measure of each territory will be not considered. Only the second step does not ignore the activity measures. Consequently the balance is not quite well. Although the result is not completely persuasive in this case, the algorithm may have potential to be an useful heuristic. But therefore several adaptations need to be done.

5.7 EatUpMinDist

The EatupMinDist algorithm is an improvement of Eat-up. It combines the approaches of Eat-up and AllocMinDist. The main processes are similar to Eat-up: The algorithm extends one territory centre after another by adding yet unassigned basic areas to the territory successively. But contrary to Eat-up the next basic area in database will not be chosen. Instead of this the distance of basic areas to territory centres will be considered. Consequently the order of allocated basic area depends not on the sequence in database anymore. Similar to Eat-up a termination criterion is used to stop the allocation of basic areas to one territory centre. That threshold is calculated by creating the sum of activity measures of all basic areas and dividing that sum by the number of territories that should be created.
The workflow of the algorithm and the results are shown in the following figures as well as in Table 11.

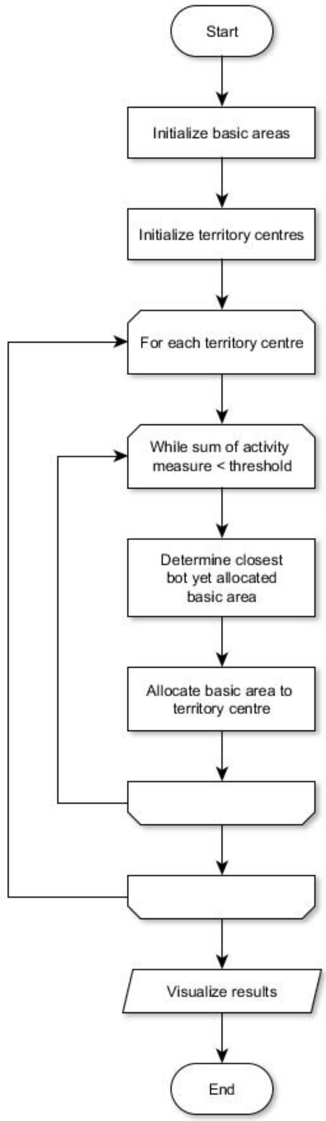


Figure 32: *Workflow of EatUpMinDist algorithm*

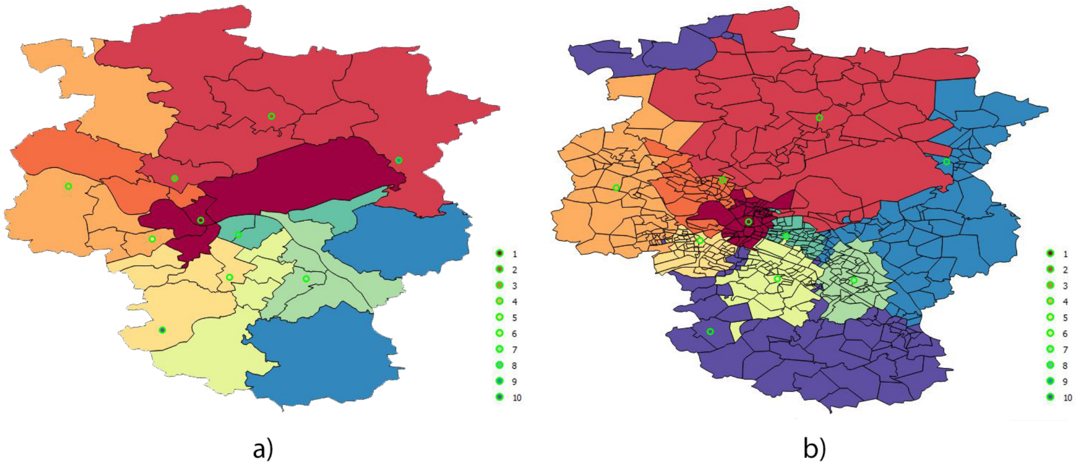


Figure 33: *Result of area segmentation process using EatUpMinDist. The territory centres are visualized as circles. a) zip-code5 areas. b) zip-code8 areas.*

	number of basic areas		sum of activity measure		compactness	
	zip-code5	zip-code8	zip-code5	zip-code8	zip-code5	zip-code8
Territory 1	4	48	46845	35216	0.23714	0.23448
Territory 2	7	54	45974	35839	0.16812	0.23678
Territory 3	3	46	36416	35666	0.24473	0.15943
Territory 4	5	60	47245	35085	0.13417	0.20270
Territory 5	5	57	40212	34841	0.24963	0.38059
Territory 6	4	58	36174	34784	0.16139	0.24416
Territory 7	4	55	37635	34796	0.19808	0.33026
Territory 8	3	50	37524	35716	0.19377	0.27660
Territory 9	2	71	18595	35199	0.17287	0.12722
Territory 10	0	65	null	29478	null	0.08759

Table 11: *Results of area segmentation using EatUpMinDis*

The results of EatupMinDist shows similar problems that occur during the application of Eat-up algorithm. The created territories are not well balanced and not coherent in all cases. By considering the distances from a territory centre to basic areas the allocated areas are more closely to the centre. At the same time some territory centres lying outside of their territories because the basic areas which are closest to them were already allocated to other territories. Additionally some territories are not coherent because at the end just basic areas may exist which are located far away of the centre. Besides these problems the balance is not given all the time. The algorithm just stops the allocation as soon as the threshold is exceeded, consequently too much basic areas are allocated so that some territories may get no basic areas assigned, see case study of zip-code5 areas. Territory 10 does not own any basic areas because already all areas are allocated previously. Caused by a finer subdivision of areas using zip-code8 the problem does not occur in this case. Concluding it is recognizable that this approach is not usable for Geomarketing strategies using fixed centres.

5.8 AllocMinDistLocalSearch

The AllocMinDistLocalSearch algorithm is a combination using AllocMinDist and LocalSearch approaches. Consequently the process consists of two steps. In first step basic areas are allocated by their distances to the closest territory centres. That approach is similar to AllocMinDist. Afterwards the allocated basic areas are rearranged again in order to get a balanced distribution. Therefore it will be checked whether the balance is satisfied. During the balance check a threshold value is used. It determines the range of allowed aberrancy comparing all territory activity measures. If the examination concludes that the territories are not well balanced the rearrangement starts. Doing so at first the territory with the highest sum of activity measure will be determined. Afterwards one of the containing basic areas is given to the neighboured territory that owns the smallest sum of activity measure. As soon as these steps are done the examination of the balance starts again.

The workflow of the algorithm and the results are shown in the following figures and in Table 12.

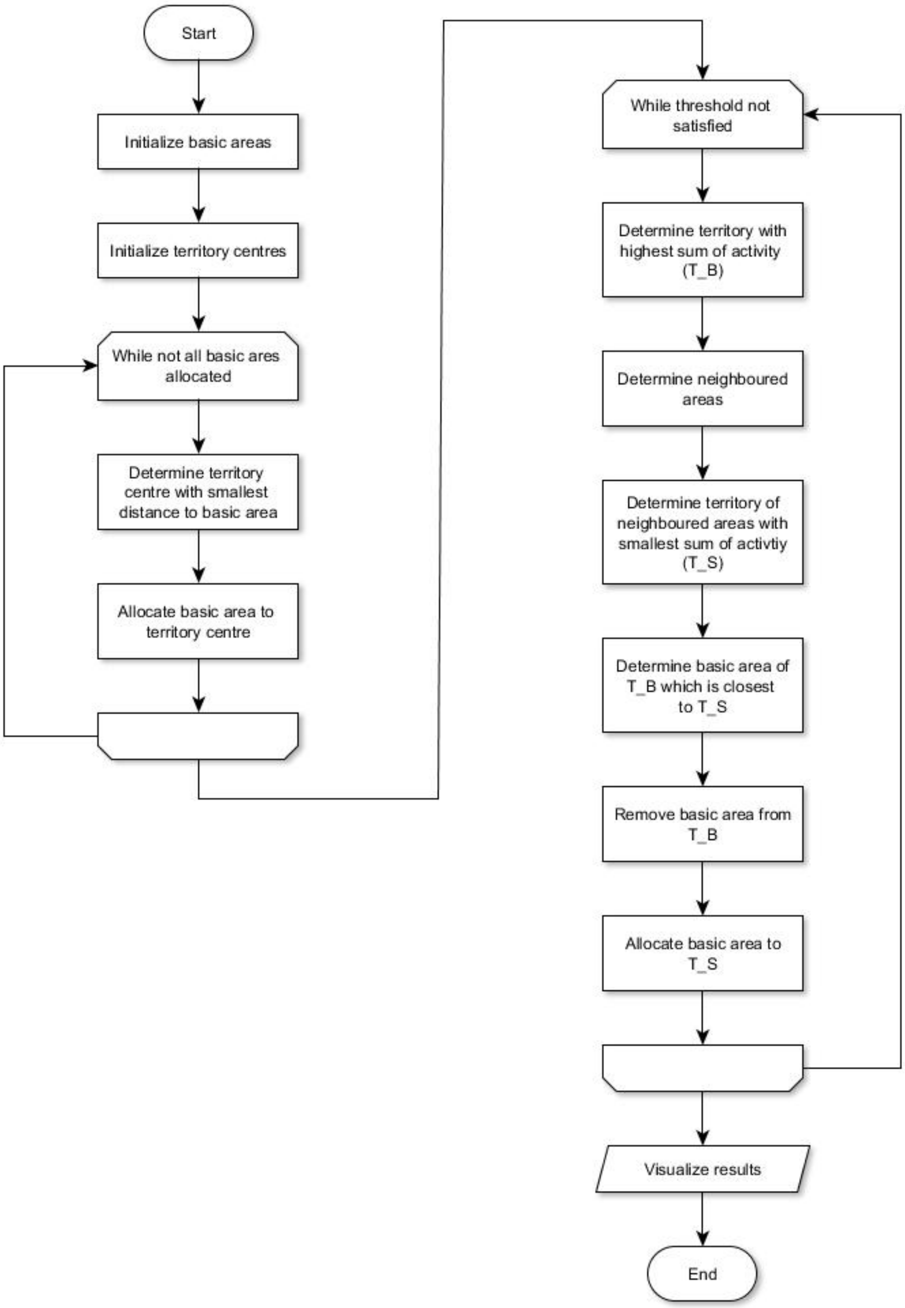
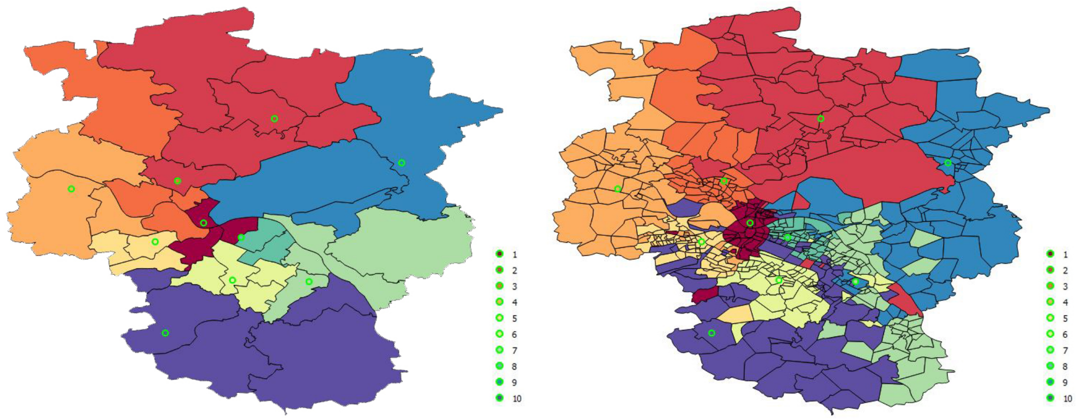


Figure 34: *Workflow of AllocMinDistLocalSearch algorithm*



a) b)

Figure 35: *Result of area segmentation process using AllocMinDistLocalSearch. The territory centres are visualized as circles. a) zip-code5 areas. b) zip-code8 areas.*

	number of basic areas		sum of activity measure		compactness	
	zip-code5	zip-code8	zip-code5	zip-code8	zip-code5	zip-code8
Territory 1	3	51	35771	35329	0.18579	0.18877
Territory 2	6	44	34652	27881	0.26533	0.15301
Territory 3	4	46	29511	35266	0.11334	0.14947
Territory 4	3	62	35046	35299	0.24743	0.08732
Territory 5	3	55	35068	35763	0.34382	0.06488
Territory 6	4	56	37216	35401	0.29540	0.12332
Territory 7	4	66	34219	35355	0.17911	0.06317
Territory 8	2	48	36621	35467	0.19377	0.09523
Territory 9	3	68	33964	35084	0.17287	0.06374
Territory 10	5	68	34552	35775	0.26951	0.04481

Table 12: *Results of area segmentation using AllocMinDistLocalSearch*

The results of AllocMinDistLocalSearch show well balanced territories, which are sometimes not full coherent. The not given contiguity is caused by the local search. Using AllocMinDist in first step creates mostly coherent territories, consequently at the beginning well formed areas were created. During the rearrangement to get balanced territories the coherence will not be considered thus basic areas may change in such a way that they are not linked anymore to their territory. Consequently an improvement will be necessary to satisfy the contiguity constraint. Just the balance is considered during the local search. The admissible range of differences between the sum of activity measure of each threshold is defined by a threshold value. In this case study the threshold value was set to 30%. The admissible range is just a guiding value to represent the maximal difference. Anyhow the resulting differences of sums of activity measures may be smaller than the threshold value. Although the implemented algorithm contains some problems concerning contiguity and compactness at the moment, it offers a great potential to be an promising approach. Nevertheless some improvements are necessary to use it for Geomarketing strategies effectively.

6 Comparison of implemented approaches

In order to compare the results of all introduced algorithms they have been implemented. The results and a summary of each approach is available in section 5 "Implementation of area segmentation approaches". These results are now used to define parameters that measure the quality of the implemented algorithms. Therefore different terms are defined. One term is called δ. This one defines the average difference to the balance value that should be reached. The balance value, denoted as B is calculated by the quotient of the sum of activity measures of all basic areas and the number of territories created during the calculation. In the case study the balance value is an activity measure of 34662. Consequently each territory needs to own such a number of basic areas so that this value is approximately reached. For defining δ the differences of the activity measures of the territories to the balance value are calculated. Afterwards the differences are added and divided by the number of territories. It can be formulated as

$$\delta = \frac{\sum_{T_i \in V} w(T_i) - B}{N}$$

Additionally to δ a second parameter θ will be defined. θ describes the difference to the desirabled number of territories and consequently evaluates the quality of the contiguity of the algorithms. Therefore for each territory the number of sub territories will be calculated. To achieve a satisfying coherency the number of sub territories needs to be one. But the algorithms show that this is not always the case. Consequently each territory can be evaluate concerning the sub territories. These territories can be summed up to determine a conclusion about the coherency. Caused by the fact that the desirabled number of territories needs to be the number of how much territories need to be created (in the case study these are 10 territories), the number of territories can be divided by the number of calculated sub territories to get a quality measure of compactness. Let n_T denotes the number of sub territories of each territory then it can be formulated as

$$\theta = \frac{N}{\sum_{T_i \in V} n_T}$$

Besides these two parameters additionally another notion, denoted as γ, will be used showing the average compactness measure for each algorithm. Therefore the compactness measure of each territory will be summed up. Afterwards they are divided by the number of territories that should be created. Let c_T denoting the compactness of the territories. Thus the calculation of γ can be formulated as

$$\gamma = \frac{\sum_{T_i \in V} c_T}{N}$$

All three parameters will be used for evaluating and comparing the algorithms. For differentiating the distinct sources of basic areas that were used, the parameters getting a footnote,

- 5 denotes that the calculation was done to zip-code5 areas.
- 8 denotes that the calculation was done to zip-code8 areas.

The results of all algorithms are shown in Table 13. Besides the determined parameters additionally the visualization of the results as well as the calculation time will be shown.

After summarizing the results the comparison can be done in more detail. Therefore a ranking of the qualities of the different parameters will be done. In section 3.1 "Notions and criteria" the conditions of a satisfying area segmentation process have been defined. It was determined that balance, contiguity and compactness are the three main conditions which should be considered. Consequently the quality of the algorithms can be defined by ranking the results concerning balance, contiguity and compactness. Therefore points between 1 and 10 will be given dependent on the quality of the results. The higher the value, the higher the quality. Afterwards the earned points will be summed up, thus the sums of all algorithms can be compared easily. Defining the ranges for every point the biggest achieved value of a criteria will be taken and divided by 10. Consequently 10 intervals are created. Dependent on the interval a value lies in, the points are given to it. The result of the ranking is shown in Table 14. The calculation time will be not ranked for two reasons. First the main focus is set to the three mentioned parameters. Consequently these are the most important and needs to be chosen to evaluate the algorithms. The second reason is caused by the limited comparability of the performance times. Each algorithm owns a different complexity. This one is influenced by the number of accesses to the databases for example. The database accesses are necessary to get data for solving the area segmentation problem and to perform spatial calculations. Taking AllocMinDistLocalSearch it can be recognized that neighbouring relationships are used to rearrange basic areas to get balanced territories. Consequently a lot of more accesses to the database are necessary to use this relationships. It is obviously that an algorithm without using such relationships is faster during the calculations but may yield to worse results.

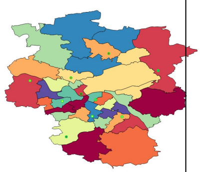
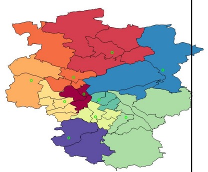
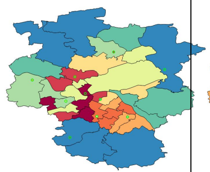
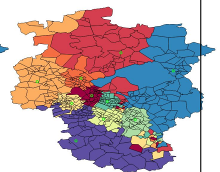
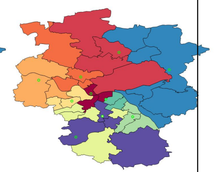
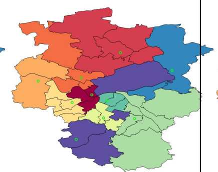
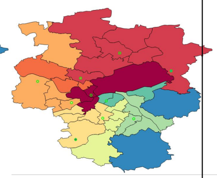
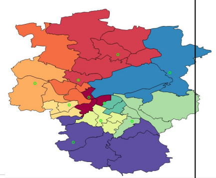
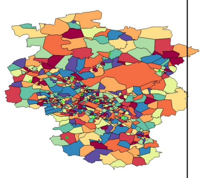
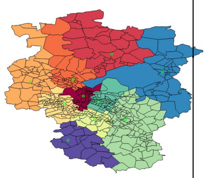
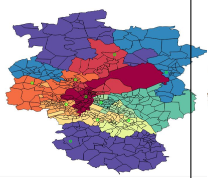
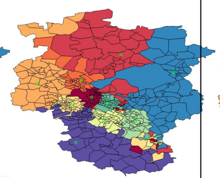
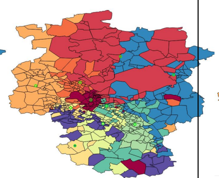
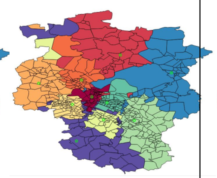
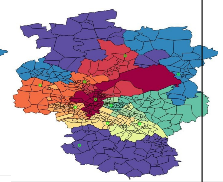
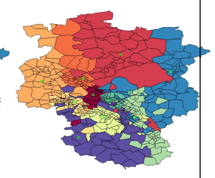
	AllocCrit	AllocMinDist	Eat-up	SmallestCrit GetsNearest	SmallestCrit GetsTrue-Nearest	Outside SmallestCrit-GetsNearest	EatUp MinDist	AllocMin DistLo-calSearch
zip-code5								
zip-code8								
δ_5	2984.8	11338.8	6932.4	3506.4	1417.8	10.828	10145.8	1282.4
δ_8	93.2	14945.8	927.8	125	142.8	8902	1036.8	1356.2
θ_5	0.3448	1	0.4762	0.7143	0.7692	0.8333	0.8333	0.9091
θ_8	0.0379	0.9091	0.5	0.2439	0.2222	0.4	0.5556	0.2381
γ_5	0.13559	0.12793	0.12794	0.25064	0.21807	0.24808	0.17599	0.22664
γ_8	0.01535	0.15426	0.15426	0.14309	0.08398	0.21587	0.22798	0.10337
t_5	0.421 s	2.707 s	0.501 s	0.858 s	0.779 s	0.668 s	1.023 s	3.525 s
t_8	0.506 s	3.979 s	0.583 s	3.401 s	3.107 s	2.669 s	3.965 s	129.920 s

Table 13: *Overview of results of area segmentation using different heuristics.*

	AllocCrit	AllocMinDist	Eat-up	SmallestCrit GetsNearest	SmallestCrit GetsTrue-Nearest	Outside SmallestCrit-GetsNearest	EatUp MinDist	AllocMin DistLo-calSearch
δ_5	8	1	4	7	9	1	2	9
δ_8	10	1	10	10	10	5	10	10
θ_5	4	10	5	8	8	9	9	10
θ_8	1	10	6	3	3	5	6	3
γ_5	2	2	2	3	3	3	2	3
γ_8	1	2	2	2	1	3	3	2
sum	**26**	**26**	**29**	**33**	**34**	**26**	**32**	**37**

Table 14: *Ranking of different heuristics for area segmentation*

The ranking of the algorithms shows abilities and weaknesses of the approaches. By calculating the sum of the given points it can be determined that AllocMinDistLocalSearch is the most promising algorithm of the amount of implemented approaches. Although the algorithm has some weaknesses yet it associates the best results concerning balance, contiguity and compactness. In the following all algorithm will be compared in more detail to explain the solutions. The smallest sum of points owned AllocCrit, AllocMinDist and OutsideSmallestCritGetsNearest. The problem by using AllocCrit is the missing consideration of coherence and compact territories. Just the activity measures are kept in mind consequently a patchwork rug of territories is created, thus the algorithm is not usable. The same problem can be determined using AllocMinDist but contrary to AllocCrit just contiguity and compactness are kept in mind. Hence the balance is completely ignored so that the resulting territories are mostly coherent but not well balanced. Thus improvement or combinations of different heuristics are necessary to make this approach usable.
OutsideSmallestCritGetsNearest was used as an improvement of SmallestCritGetsNearest but owns a lot of weaknesses. Dependent on the used threshold value defining the basic areas which are allocated first by considering their distances to the territories centre the result of the algorithm will be more or less balanced. If the threshold is very huge no basic areas will be allocated first so that the algorithm is working equally to SmallestCritGetsNearest. If a small threshold value is taken a lot of basic areas will be assigned in first step. This process causes a great difference concerning the sum of activity measure in each territory. Consequently the tried improvement failed. Thus SmallestCritGetsNearest is more applicable than OutsideSmallestCritGetsNearest. That result can be confirmed showing to the ranking. A little bit more successful than the three mentioned algorithm is the Eat-up approach. Nevertheless that algorithm owns some weaknesses, too. The greatest problem is that it will not be satisfied that even enough territories will be created. By aborting the allocation of basic areas to one territory using a threshold value, for each territories too many basic areas are assigned. Consequently it may be happen that no basic areas can be assigned to the lasts territory centres because all ones were already allocated (see case study using zip-code5 areas).Thus the algorithm can be not used within area segmentation processes using fixed centres. The same problem occurs during the application of EatupMinDist. Within that approach an improvement were implemented to get more coherent territories compared to the ones which were achieved using Eat-up. This target is now satisfied. Nevertheless the creation of the predefined number of territories is not given again. Consequently like Eat-up this algorithm is not usable using fixed centres, too. The results where much better using zip-code8 areas but the algorithm should be more robust in order to fit real world use cases for Geomarketing. But for some data no promising result will be achieved, hence the algorithm is classified to be not applicable. Better results are achieved by using SmallestCritGetsNearest and SmallestCritGetsTrueNearest. Both algorithms almost got the same ranking. There exists just some small differences concerning the given points of balance and compactness. Although both algorithm got similar points within the ranking SmallestCritGetsTrueNearest

own a lot of more problems than SmallestCritGetsNearest. During the application of SmallestCritGetsTrueNearest the main problem occurs during the allocation of the basic area that is the closest one. Remembering the algorithm the territory centre with smallest sum of activity measure was determined. Afterwards the nearest basic area which is not allocated yet is chosen. Before allocation it is checked whether another territory centre is located nearer to that basic area. If this is the case the territory centre with smallest activity measure gets the nearest basic area which is allocated in another territory. This change affects an invinite loop because always sane basic area are redistributed circular. A workaround was needed to solve that problem. But the chosen algorithm influences the resolution so much that it is not usable for Geomarketing analyses. Until yet no solution is found which is working dependably. Consequently the SmallestCritGetsTrueNearest approach is not recommended for area segmentation processes yet. If a dependable solution for the infinite loops will be found the algorithm become maybe more promising. Especially the consideration of the distances of each basic areas in more detail is a great advantage to the processes being used in SmallestCritGetsNearest. Because the main problem of SmallestCritGetsNearest is that some basic areas are far away from their territory because just there areas were not allocated yet. Consequently a useful improvement is necessary to make the algorithm applicable for Geomarketing strategies.
The ranking shows that the most promising algorithm is AllocMinDistLocalSearch. Compared to all other implemented algorithm it delivers the best results. The greatest advantages compared to the others are the two steps which are used in AllocMinDistLocalSearch. With the help of these steps it is tried to find a satisfying solution of balanced and compact territories. Other algorithm like AllocCrit and AllocMinDist considered just one of the parameters whereas AllocMinDistLocalSearch consider both. Nevertheless it is recognizable that there exist some weaknesses concerning compactness and contiguity yet. Consequently improvements are necessary to make it fully usable for the desired analyses.

6.1 Requirements from the field of Geomarketing

Within the chapters before several times requirements from the field of Geomarketing were mentioned. Consequently these will be considered in the following with the regard to the most promising algorithm. The approach with best results is AllocMinDistLocalSearch.
The aim during the application of area segmentation process is to define territories which contain a balanced value of an activity measure like household numbers, purchasing power or constituents. Consequently one requirement is the balance of the given activity measure. AllocMinDistLocalSearch uses the local search approach to achieve a balanced result. Within the case study the average difference to the desired activity measure value of each territory was 1282.4 in case of zip-code5 areas and 1356.2 in case of zip-code8 areas. These results show that the territories are well balanced. It is important to check whether such well balanced results will be achieved also during the calculation of other examples to get a convincing evaluation. But at the moment the results satisfying the balance constraint.

The second requirement from the field of Geomarketing is contiguity. Considering sales distriction for instance, it is obvious that each sales man needs to have a coherent territory for which he is responsible for. In the most of the cases completely coherent areas are required, too. The results of AllocMinDistLocalSearch shows that satisfying this constraint is not given all the time yet. Concerning the calculation using zip-code5 areas the created territories are not contiguous. The same result is identifiable in the case of using zip-code8 areas. This is caused by the missing consideration of coherence during the local search. Consequently an enhancement of the existing algorithm is necessary which includes also the check of contiguity. This one needs to be implemented to get an applicable algorithm for Geomarketing strategies.

Third, compactness is another requirement. Taking again the example of sales distriction compact territories need to be created to minimize travelling times and thus work load. A lot of other approaches also require compact territories consequently it will be determined as one necessary constraint. Concerning the results of AllocMinDistLocalSearch the average compactness is better than the one other algorithms had achieved. Nevertheless the values are just 0.22664 (zip-code5) and 0.10337 (zip-code8). For evaluating the compactness the measure of Cox is used. This one determines that a value between 0 and 1 can be achieved. The nearer the calculated compactness value to 1, the better the compactness is. Both values are nearer to 0 than to 1. Consequently the areas own no good compactness. This fact has two reasons. One problem is the incoherence of some territories because this one takes a huge influence to the compactness, too. Another problem is the shape of the investigation areas. Considering the shape and the containing basic areas it will be not possible to create 10 round territories consequently no compactness measure of 1 can be reached. Nevertheless the compactness values are not satisfying at the moment. Consequently an approach needs to be implemented so that the compactness will be considered much more.

Another not yet mentioned requirement for area segmentation process is the location of the territory centre within the territory that belongs to the centre. Some of the implemented algorithms do not provide this requirement, see EatupMinDist for example. Also during the application of AllocMinDistLocalSearch such problem may occur. Consequently a constraint need to be implemented which satisfies that the territory centre is located within the territory.

6.2 Conclusion

The comparison of the implemented approaches shows that some of the algorithms are not usable but some of them are. A ranking of different parameters for every use case was done to make the results comparable. Therefore three parameters were defined to summarize the calculation results. One parameter shows the average difference to the balanced activity measure value that needs to be satisfied, the second one considers the contiguity and the third one evaluates the compactness. With the help of these parameters points between 1 to 10 can be given to the algorithms so that a sum of points can be calculated. The algorithm with highest sum will be the most promising approach. The comparison shows that

AllocCrit and AllocMinDist are not usable. Additionally the calculations show that Eat-Up and EatupMinDist are not applicable to area segmentation processes using fixed centres, too. OutsideSmallestCritgetsNearest and SmallestCritGetsTrueNearest own too much problems to make them promising for Geomarketing strategies. Just SmallCritGetsNearest and AllocMinDistLocalSearch do have potential to be useful algorithm. The results show that AllocMinDistLocalSearch yields to the best area segmentation solution compared to the other algorithms. Consequently this algorithm will be used for the application to Geomarketing strategies. Before it is fully usable at first some improvements need to be done to fulfil the requirements from the field of Geomarketing completely.

7 Application of AllocMinDistLocalSearch to Geomarketing strategies

The aim of this thesis is the comparison of different approaches for area segmentation process to define the most promising one for application to Geomarketing strategies. Therefore different approaches were developed and implemented so that a comparison was possible. The comparison shows that the approach called AllocMinDistLocalSearch yields the best results concerning the requirements from the field of Geomarketing. During the AllocMinDistLocalSearch algorithm first basic areas are allocated to territory centres dependent on their distance to the centres. Afterwards in second step the basic areas are rearranged again to get balanced territories concerning an activity measure. The implementations have shown that there some improvements are necessary to make AllocMinDistLocalSearch applicable for Geomarketing strategies. These ones will be considered during the application and will be explained in more detail within the following sections. The algorithm will be applied to three defined kinds of strategies. One type will be the area segmentation itself which is applied during political distriction and by the optimization of existing sales districts for instance. Within these analyses already fixed territory centres are available. The second Geomarketing strategy is called Greenfield analysis. This one owns no given territory centres. Consequently at first some centres need to be defined. Afterwards the area segmentation algorithm can be applied. Whitespot analyses are a combination of area segmentation and Greenfield analyses. Within the Whitespot analyses already some territory centres exists. Additionally to the existing ones, new centres need to be created. To all these centres (given ones and new ones) the area segmentation of the basic areas will be done. All three kinds of analyses will be explained in more detail within the following sections.

7.1 Area segmentation and optimization

The area segmentation process is the main approach an can be found in both other used type of strategies too. Consequently this algorithm needs to be dependable and need to provide satisfying results because it affects the solution of the others too. Area segmentation processes are the core of all applied districtions for example political distriction, sales distriction and the optimization of sale territories. With the help of area segmentation processes a clustering of basic areas will be done into a predefined number of territories. Thereby one or more specific activity measures need to be considered. Within the case study just household numbers are considered during the allocation. During the comparison of different approaches it was determined that AllocMinDistLocalSearch yields to the best results. But the algorithm offers some weaknesses yet consequently enhancements are necessary to implement. Therefore the main ideas of AllocMinDistLocalSearch are taken. The process will be amplified by checking some constraints to satisfying the requirements from the field of Geomarketing. The most important constraint is the location of the territory centres within the territories which belong to it. Therefore it will be determined which basic area contains the territory centre. The resulting basic areas will be saved and considered all the time during the allocation. Remembering the structure of AllocMinDistLocalSearch it was recognizable that it consists of two steps. In first step basic areas are allocated to the territory centres dependent on their distance. In second step local search will be applied to realize the balance constraint. Within the first step the centre is always located within the territory even if the constraint is not applied. The problem of residing outside exists first during the rearrangement. Consequently before a basic area will be rearranged it will be checked whether it is the saved basic area, which contains the territory centre. If this is the case, the basic area will not be rearranged. Thus another basic area needs to be chosen. That means the basic areas containing the centres are preserved all the time for rearranging.

Besides the location of the territory centres, it will be postulated that the created territories need to be contiguous. For satisfying the coherence two different steps are need to implemented. The first step will be applied to the AllocMinDist approach which is used in first step of AllocMinDistLocalSearch. In most of the cases the created territories are coherent and compact after that step. But this result is not given all the time. Sometimes some basic areas are allocated in such a way that no contiguity within a territory exists. This problem was already mentioned during the analysis of the implementation of AllocMinDist. Consequently a check needs to be done after the first step of AllocMinDistLocalSearch to satisfying the coherence of the territories. Therefore each created territory will be checked whether all basic areas are contiguous. If this is the case, the next territory will be taken. If the check concludes with an incoherency basic areas will be rearranged to another territory. Therefore the part which is located apart will be defined. For doing so all basic areas of the territory are taken. From the amount of areas firstly the area is chosen which contains the territory centre. Afterwards all basic areas are determined which can be reached from that area using a graph. All basic areas which are not linked to the amount of basic areas that are reachable are the ones which need to be rearranged. Figure 36 illustrates the graph on an example.

Figure 36: *Using a graph for determining incoherence. All basic areas that can be reached by an edge of the graph (black lines linked by nodes) are contigous. Thus, one basic area needs to be rearranged. Example territory using zip-code8 areas.*

After the check all territories are contiguous so that a satisfying base exists for doing the rearrangement by local search. Previously all neighbouring territories are determined to define the one with the highest sum of activity measure. This territory has to deliver one basic area to the territory with the smallest sum of activity measure. Doing so maybe a basic area is given that causes an incoherency in one of the two areas for instance by diving the territory that gives the area into two half's. Consequently a check is necessary again whether the basic area can be given. Therefore the same approach is used like after step one. By comparing the resulting graphs of both territories it can be determined whether both ones are contiguous anymore after change. If the contiguity is given for both territories the change can be done. For creating the graph knowledge about neighbouring relationship need to be there. Two basic areas are neighboured if they are sharing edges or points. Determining the neighbour relationships can be done with the help of a function in PostGIS which determines all neighbours of a geometry. These information can be saved and used during the creation of the graph. Thus no additional access to the database is necessary. Another possibility is checking the coherency with the help of a PostGIS function. But this approach needs a lot of additionally database access consequently the run time raises much. Thus this approach was abandoned by performance issues.
Additionally to contiguity the compactness of the territories is important. For evaluating the quality of compactness the measure of Cox is used. During the evaluation of the algorithms

the compactness measure was just applied afterwards for determining the quality of compactness. Instead of this the measurement will be applied during the rearrangement to get more compact territories. Therefore the compactness of the territories need to be initialized after doing the first arrangement by AllocMinDist in first step. With the help of the initial values of compactness the rearrangement can be done in such a way that the compactness will be considered. Caused by the balance constraint it can be happen that basic areas need to be allocated in such a way that the compactness getting worse. Consequently a solution needs to be implemented to weight both constraints. Therefore a function was created which includes both constraints: compactness and balance. The function measures the changes of compactness and balance concerning both values before and after rearranging. The resulting values show whether the rearrangement produces a better resolution or not. Consequently all possible basic areas that may be rearranged can be checked so that the one can be chosen that provides the best result. Let $\Delta\phi$ denotes the changes of compactness and balance it can be written as

$$\Delta\phi = \phi_{new} - \phi_{old} \rightarrow min$$

ϕ_{new} and ϕ_{old} need to be calculated using the alteration of balance and compactness including weighting values. The resulting value of $\Delta\phi$ may be smaller or greater 0. If the value is smaller 0, the arrangement yields to a better resolution than before the rearrangement. Vice versa a value greater 0 denotes a degeneration. Consequently the value needs to be as minimal as possible. Let w denotes the weightings, Δc denotes the change of compactness and Δb denotes the change of the balance value thus the calculation of ϕ can be formulated as

$$\phi = \Delta c * w_1 + \Delta b * w_2 \rightarrow min$$

Δc will be calculated using the compactness measure of Cox. The resulting value will be normalized by

$$\Delta c = 1 - c \rightarrow min$$

The same will be done during the calculation of the change of balance value. For calculating the balance value the difference to the given sum of activity measure will be used that each territory should contain. That difference is normalized similar to the compactness measure.

$$\Delta b = 1 - b \rightarrow min$$

The normalizing is necessary for the application of the weighting values. The weighting values

can be defined by the user and need to be values between 0 and 1. The application of the algorithm using the weighting function shows that the weighting needs to be determined in benefit of balance if the threshold value is defined to be small. The threshold value is used for defining the allowed difference of the sum of activity measure comparing the territories. At the same time case studies show that a weighting for the benefit of balance does not mean automatically that worse balanced territories will be achieved. The weighting value just determines which basic areas will be preferred during the rearrangement.
The algorithm for area segmentation processes was applied to the same test data which were used during the implementation and comparison of different heuristic approaches. The results of different test cases using several weighting values and thresholds are shown in Figure 37.

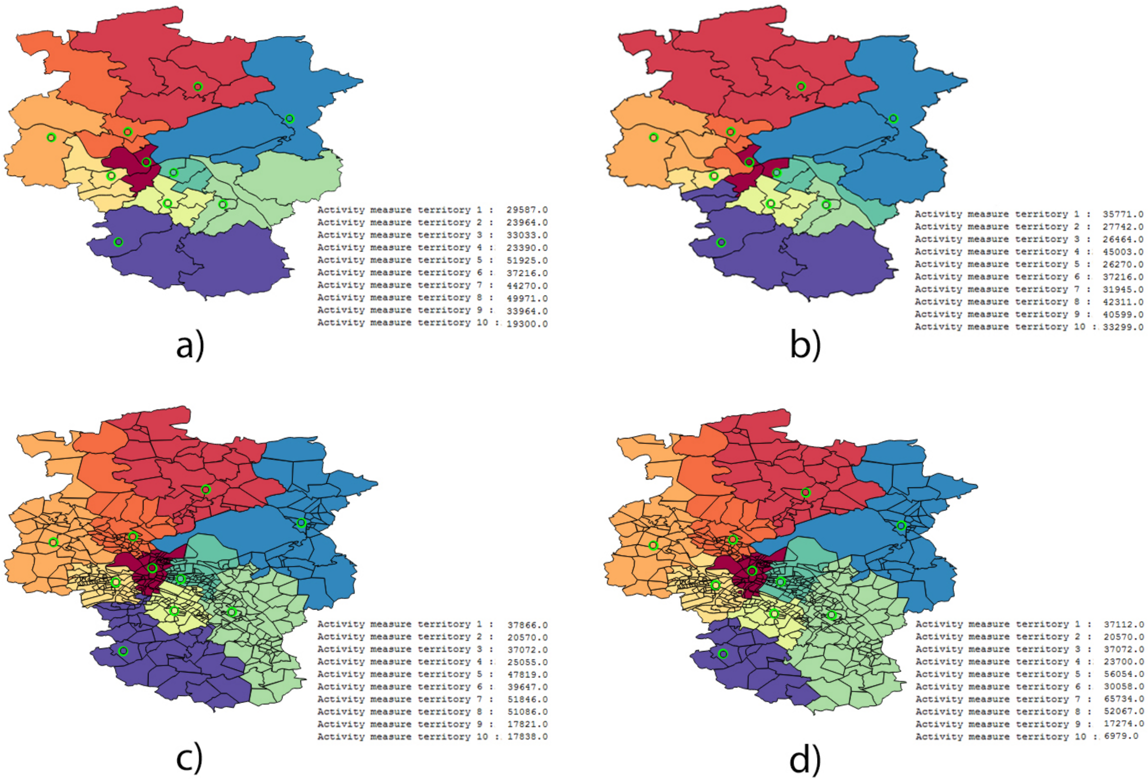


Figure 37: *Application of area segmentation algorithm to zip-code5 and zip-code8 areas. a) Weighting values: compactness 1, balance 0, threshold 50. b) Weighting values: compactness 0, balance 1, threshold 30. c) Weighting values: compactness 1, balance 0, threshold 50. a) Weighting values: compactnes 1, balance 0, threshold 100*

The results show that the resolution achieved by the area segmentation process depends on the chosen values for the weighting and the threshold. By using the compactness measure the territories are more compact. Their shapes were confirmed by some Geomarketing specialist which have considered them as compact. During the calculation of the compactness measure the areas and circumferences of the territories are used. These ones can be calculated with the help of functions provided by PostGIS. Performance tests showed that these calculations need

a lot of time because they are done several times during the rearrangement. Consequently a solution was implemented which uses no access to the database. Instead of this the areas, circumferences and shared edges of basic areas were saved at the beginning. Thus only one time a database access was necessarily. Afterwards every calculation can be done with the help of the stored information.
The rearrangement of basic areas during the local search offer some problems. It can be recognized in some cases that the predefined threshold value will be not reached because always the same basic areas are rearranged again and again so that no correction of the sum of activity measure will be achieved. Several approaches were tried to find a solution for this problem but no optimal one could be found. Within AllocMinDistLocalSearch the territory with smallest sum of activity measure was determined. This one took a basic area from the territory with biggest sum of activity measure. Test have shown that the problem of endless rearranging occurs more often in such approach than using the approach which is implemented now. Now the sum of all territories are checked whether the threshold value of balance is satisfied. If one territories does not fulfill the threshold value a basic area will be rearranged. If a territory satisfy the value nothing happens and thus the next territory will be checked. As soon as all territories satisfying the threshold value the algorithm ends. Caused by the endless rearrangement a break needs to be implemented to get a solution in the end. Although other approaches were tried too for optimizing the rearrangement process the implemented one yields to the best results. Nevertheless improvements are necessary to get satisfying results all the time. The workflow diagram of the implemented algorithm can be found in Figure 38.

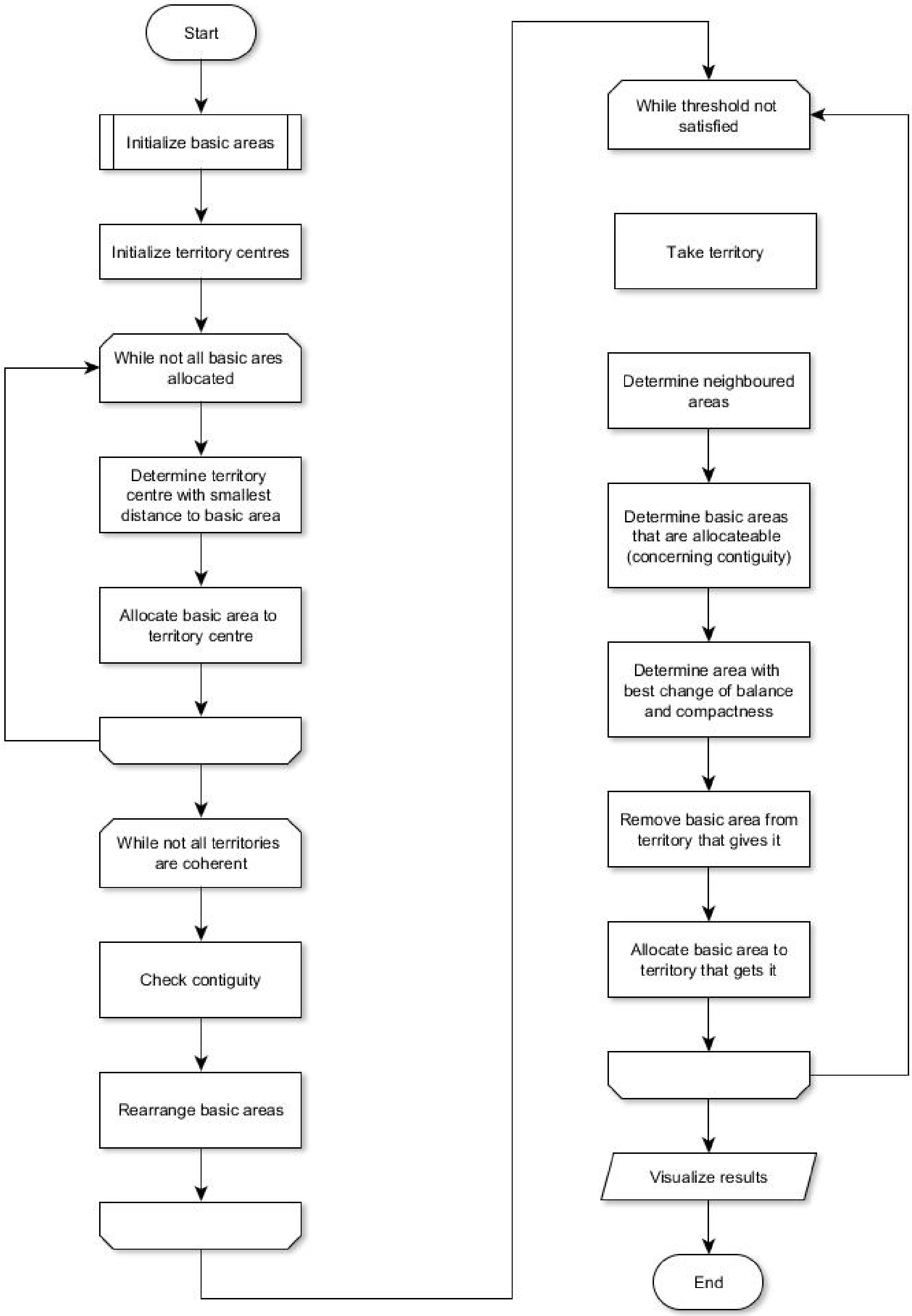


Figure 38: *Workflow diagram of implemnted area segmentation approach*

The initialization process consists of several sub process. Besides the geometries also the neighbours need to be initialized. Additionally centroids of each basic area are calculated used for determining the distances between territory centres by using orthodroms. Furthermore areas and circumferences are saved to use them during the rearrangement.

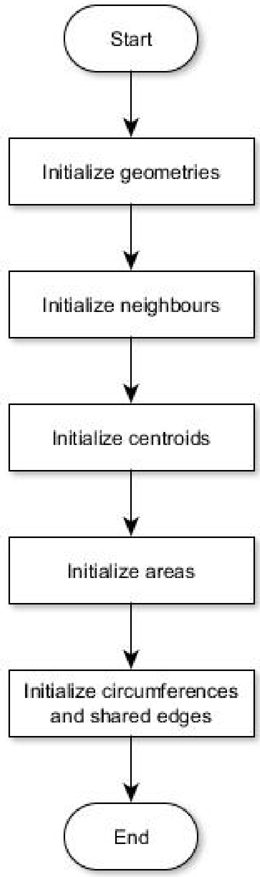


Figure 39: *Processes of the initialisation*

7.2 Greenfield analysis

Greenfield analyses are a special type of strategies for planning new locations. When a company will be founded at the beginning no branch exists. Consequently it may be needed to determine possible places for locations so that a meaningful area segmentation considering an activity measure can be done. Besides the planning of new locations for a company also political distriction using no predefined centres is an example of an approach similar to the Greenfield analyses. Doing the analysis it is first necessary to determine possible locations before the area segmentation can be done. Consequently an algorithm is needed which consists of two steps similar to location-allocation approaches. For defining possible places for the locations an approach similar to EatUpMinDist will be used. The comparison of the algorithms showed that this approach is not usable using fixed centres. But doing Greenfield analysis in the beginning no centres exists consequently the approach may be promising. The EatUpMinDist algorithm needs to be adapted in such a way that a meaningful starting point can be determined. Therefore a basic area resides on the outer border of the investigation area, will be chosen. Starting from that chosen area the first territory will be grew up at the boundary of the territory by allocating basic areas until a threshold value is reached.

The threshold value is calculated by summing the activity measures of all basic areas and dividing them through the number of territories that will be created. For generating the next territory the next basic area at the boundary will be chosen which is not allocated yet. Doing so the condition needs to be satisfied that the basic area is neighboured by the territory that was created before. As soon as a basic area was found the territory can be created. This will be done until the predefined number of territories is reached. If no basic area at the boundary exist anymore, another basic area is chosen which is neighboured by the territory created before independently on the location of that area. If no not yet allocated basic area, which is neighboured by the territory, is available anymore a basic area will be taken which is not neighboured but yet not allocated. Following these steps all territories can be created. Figure 40 shows the created territories after the first step using the case study investigation area.

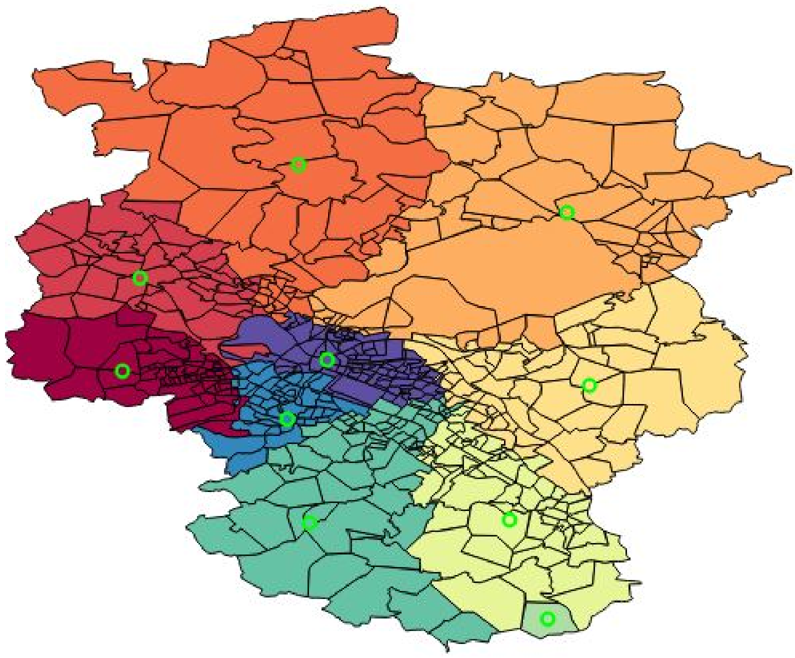


Figure 40: *Created intial territories for defining initial territory centres during a Greenfield analysis*

The problem using this approach is given by the threshold value to abort the allocation of basic areas to one territory. It was already mentioned during the explanation of Eat-up and EatUpMinDist. By exceeding the threshold value for abortion it may happen that already all basic areas are allocated before the predefined number of territories is reached. To prohibit this phenomena additionally abortion criteria need to be defined to ensure that all territories contain at least one basic area. This can be done by checking the resulting number of not yet allocated basic areas.
After the creation of initial territories is done territory centres can be determined by setting the centres into the middle of each territory. Afterwards these centres can be used doing the area segmentation. The used algorithm is the same which is used in section 7.1 "Area

segmentation and optimization". This one will be applied using the predefined territory centres. As soon as the area segmentation is passed final territory centres will be determined. The following figures show the workflow and results of that calculation using different values for weighting and threshold. In both application 10 territories and thus 10 locations are created.

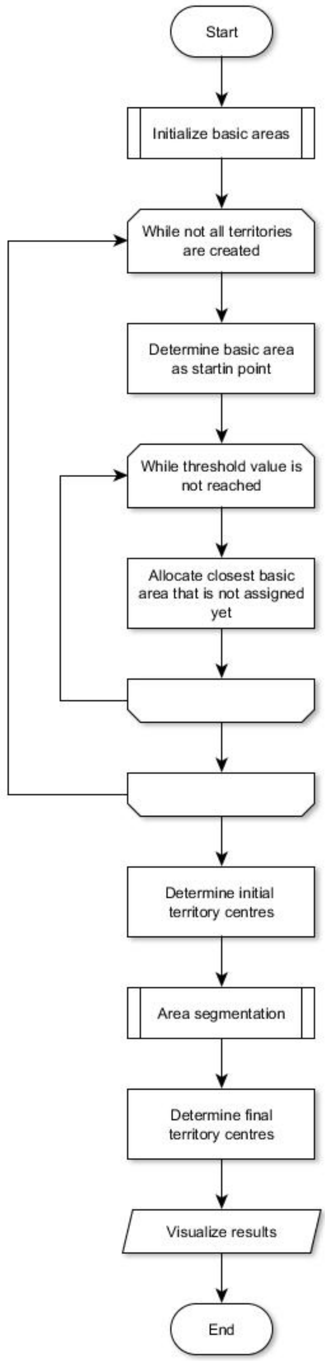


Figure 41: *Workflow of algorithm used for Greenfield analysis*

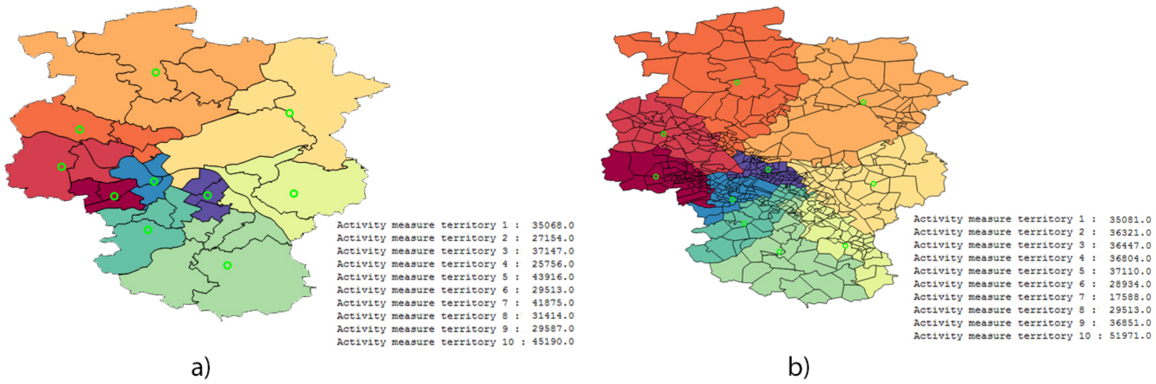


Figure 42: *Results of Greenfield analyses. a) Zip-code5 areas. Weighting values: compactness 1, balance 0, threshold 50. b) Zip-code8 areas. Weighting values: compactness 0.1, balance 0.9, threshold 50.*

Although the results of that case study look promising such resolutions are not satisfied all the time. Using other parameters or investigation areas the area segmentation process may be yield to worse results. Some possibilities are no well balanced territories or have a corrupted compactness. The origin of this can be the following two reasons: First, one problem is caused by the determination of the initial locations. They may be located in such a way that no area segmentation containing a satisfying result may be done. Second, the problems of the used area segmentation algorithm which are already mentioned in section 7.1, also take an influence of the results here. Consequently it can be concluded, that the Greenfield algorithm yields to results but another solution which will be more robust may be necessary. Some additional test cases using several investigation areas and parameters may confirm or disprove that.

7.3 Whitespot analysis

Whitespot anaylses are a combination of area segmentation processes and Greenfield analyses. But instead starting on a greenfield the white spots between existing locations need to be determined to define places where new locations can be set. Consequently several numbers of locations are given at the beginning of the calculation. Additionally to these ones a predefined number of new locations will be set. In reality, mostly 1 up to 3 new locations will be created. All locations (given one and new created one) will be used during the application of the area segmentation. Consequently it is at first necessary to define the white spot for placing new locations. Afterwards the area segmentation process can be done using all locations. The implementation of the Whitespot analysis consists of three main steps:

1. Creation of territories for given locations
2. Creation of territories containing new locations so that initial new locations can be determined

3. Application of area segmentation to given and initial locations

As soon as the area segmentation process is done the final positions of the new locations can be determined. Considering the first two steps in more detail it can be recognized that an approach similar to EatupMinDist will be applied. At first for each given location the territories are created using EatupMinDist. The threshold value for aborting the allocation is calculated by the sum of of the activity measure of all basic areas divided by the sum of given and new locations.

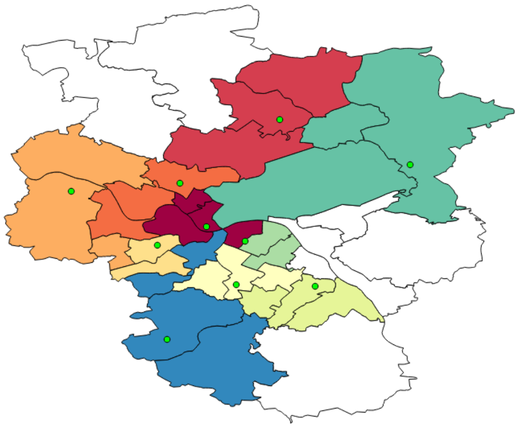


Figure 43: *Allocation of basic areas to given locations within a Whitespot analysis*

Afterwards the places of new locations need to be determined. In order to place the new locations, it is necessary first to necessary to define a starting point for creating the territory. Therefore three different approaches were implemented and compared determining the one which yields to best results. First approach uses the not yet allocated basic areas by their identification number. All number will be ordered and the one with smallest ID will be taken. Consequently no consideration of location of the basic area or value of activity measure are done during the determination of the starting point for new locations. The second algorithm uses just basic areas that are located at the boundary of the investigation area. These basic areas are ordered by their IDs and again the one with the smallest number is taken. That approach considers the location of basic areas after all but do not consider any activity measure. That is why the third approach was developed. That one considers all basic areas that are not allocated yet. Thereby all basic areas from that amount which are contiguous will be respected as a territory. Afterwards the sum of the basic areas will be calculated for each possible territory. The one with the highest sum will be taken. Within that taken territory the basic area with highest activity measure is determined. This one is used as starting point. Considering the figure above this process will be explained using the figure as example. The figure shows that seven basic areas are not allocated yet. Using these areas three contiguous territories can be created. One territory contains two basic areas, the second one contains four areas and the last one comprises just one basic area. For each territory the sum of activity measure will be calculated. In the case of the example it will be determined that the territory containing four not yet allocated basic areas owns the highest value of activity measure. Consequently one basic area from that territory will be taken as starting point. Using a basic area as starting point a territory can be created using the same approach like EatUpMinDist uses. This process is equal within all three approaches. The allocation ends as soon as the threshold value is reached or no basic areas are allocatable anymore. It is not possible to allocate basic areas either if all basic areas are allocated or if no basic area exists that is coherent to the created territory. After the allocation to the territory is done the new location can be set into the middle of the area. Figure 44 shows the resolution of creating a new location using the third approach. The given locations are marked by green points, the new one is marked by a pink circle.

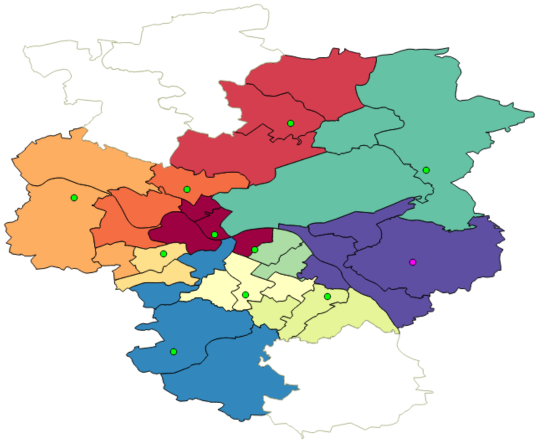


Figure 44: *Result after creating a new territory containing the new location within a Whitespot analysis*

The implementation of all three approaches showed that the third approach yields to the best results. Additionally it considers the locations and activity measures of basic areas so that it can be seen as the most advisable approach. Thus that one will be used for the Whitespot analyses. The figure shows that after creating territories for ten given and one new location, already some basic areas are not assigned yet. Consequently a check was implemented to allocate the resulting areas to existing territories. Afterwards the given and the new locations are used doing the area segmentation process. This one uses similar to Greenfield analyses the area segmentation algorithm from section 7.1 "Area segmentation and optimization".
Within the following figures workflow and results using different calculation parameters are shown.

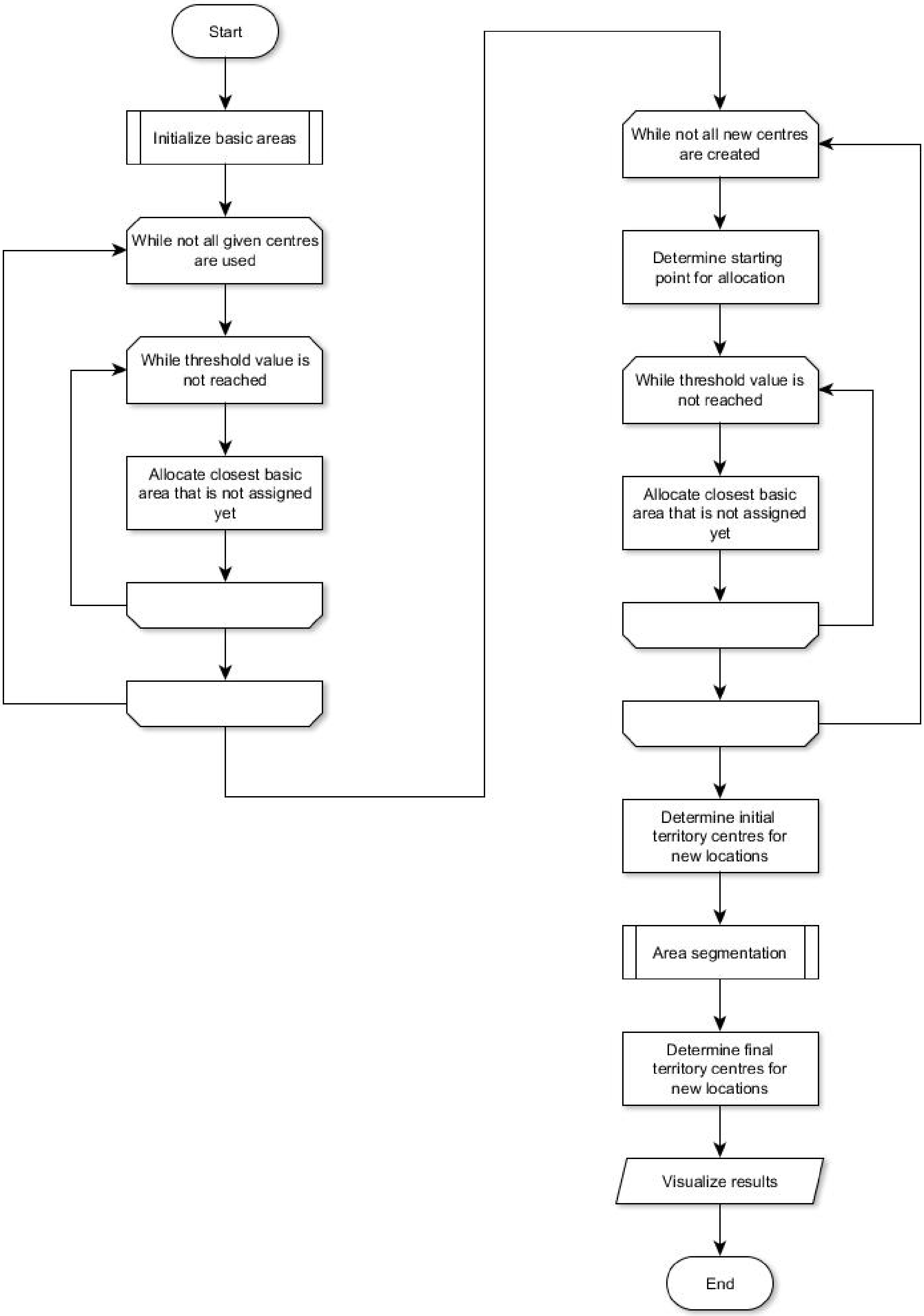


Figure 45: *Workflow of algorithm used for Whitespot analysis*

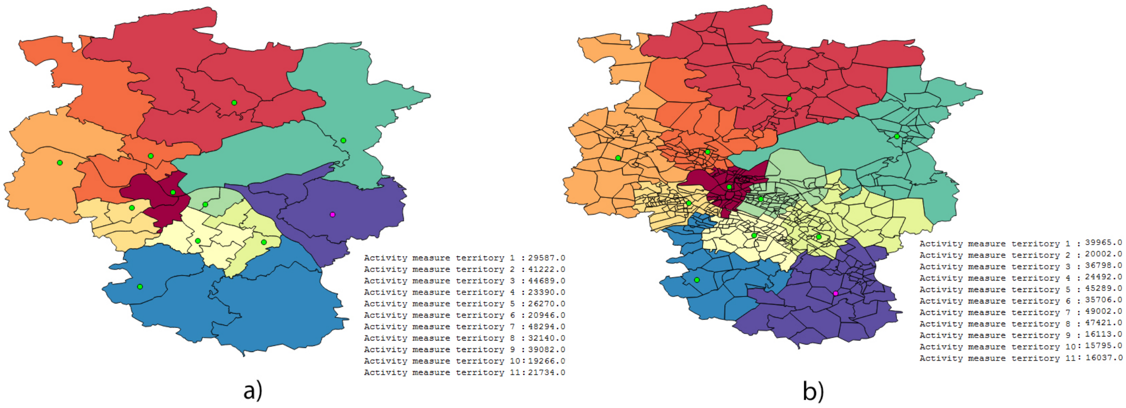


Figure 46: *Results of Whitespot analyses using 10 given locations (green circle) and creating 1 new location (pink circle). a) zip-code5 areas. a) zip-code8 areas.*

The outcome of Whitespot analysis show results that are quite good. The created territories are contiguous as well as balanced. By implementing three different approaches determining the starting point during the creation of new locations the best one could be chosen and used. Consequently the created territories containing new locations are located quite well. Although the results look good some problems exist which are caused by the mentioned weaknesses of the area segmentation approach. By using that process also for Whitespot analyses they take influence to these results, too. Nevertheless the implemented algorithm is a usable approach for doing Whitespot analyses.

8 Evaluation

After all implementations were done an evaluation of the used algorithm was accomplished. Therefore different investigation areas are used applying several parameters to make an analysis of the results possible. The results have been compared with results from other Geomarketing software.
First a comparison to a calculation in SIM Tool was done. Therefore microm provided some results doing a Greenfield analysis to zip-code8 areas of Hamburg. The area segmentation process was done using approximately 2500 basic areas. Thereby 200 new locations should be created. Afterwards the basic areas should be allocated to these locations. The calculation using SIM runs 1:30h. The same calculation was done using the Greenfield algorithm which was implemented within this thesis. For making the results comparable the same dataset was used. The aim of the implemented algorithm was to create coherent, compact and well balanced territories within an adequate running time. The results of the Greenfield analysis using the algorithm implemented for this thesis achieved the result within 5 minutes.. Contrary to the territories which were created by SIM, the created territories were all contiguous. Additionally the territories were more balanced than the ones of SIM. Nevertheless the balance can be optimized further. The result of the calculation using the algorithm implemented in this thesis is shown in Figure 47.

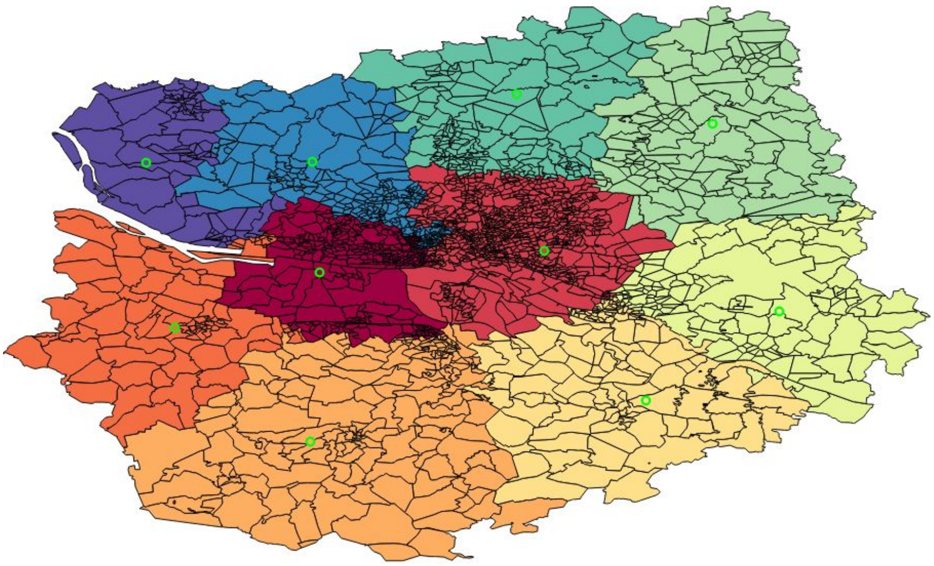


Figure 47: *Result of Greenfield analysis doing to zip-code8 areas of Hamburg. 10 locations and territories are created. The check of unity was not used in that calculation.*

Additionally a comparison to another Geomarketing software called Map&Market was done. Thereby whole Germany was used as investigation area, consequently approximately 8000 basic areas were assigned to the locations. The area segmentation was done using a Greenfield analyses creating 10 new locations. Doing this calculation a lot of problems occurred. The biggest problem are the islands in the North Sea and Baltic Sea. The islands are not connected to any other basic areas. Consequently they are not continuous to a territory in sense of the definition used in that thesis because no shared edges of two basic areas exists. That is why the check of contiguous yields to errors. To solve that problem a variable was integrated that shows whether the check of coherence should be done or not. In case of using islands within the amount of basic areas no totally contiguity can be reached thus the variable needs to indicate that no check will be done. Although the check of coherence was disabled during the calculation most of the basic areas are allocated in such a way that they are contiguous. The algorithm was improved thus the possibility creating incoherent territories become smaller. Thus the results form an expected output. The result is illustrated in Figure 48.

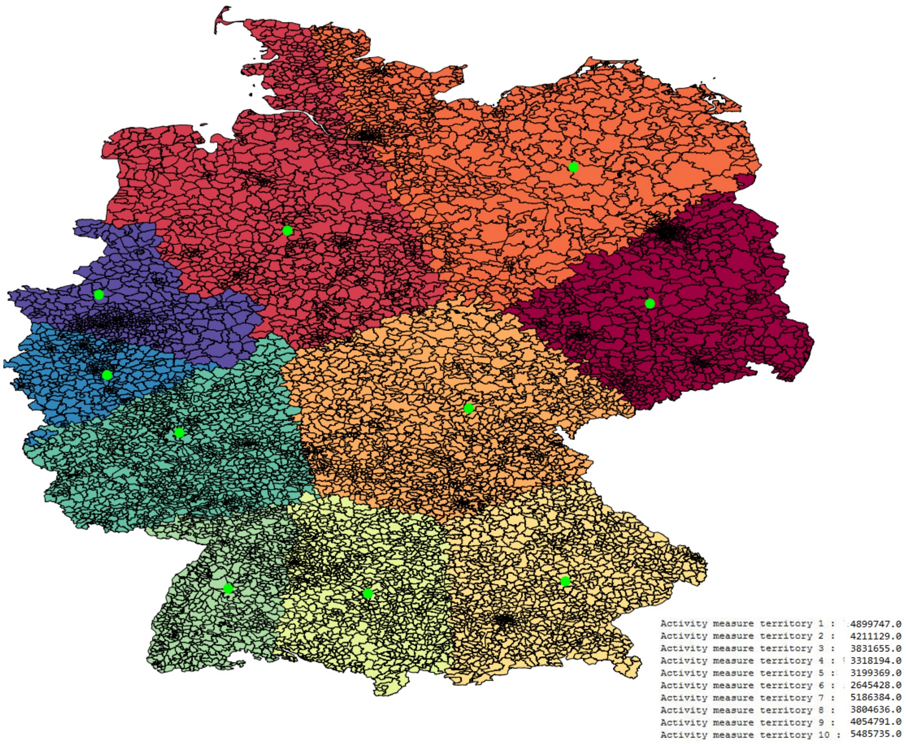


Figure 48: *Result of Greenfield analysis doing to zip-code5 areas of Germany. 10 locations and territories are created. The check of unity was not used in that calculation. Parameters: compactness 1, balance 0, threshold 30.*

Although a lot of improvements concerning the performance were done the calculation time applying the Greenfield algorithm to whole Germany is still too high. This is caused by two facts. One fact are the number of database accesses which are necessary during the calculation. Before the area segmentation starts the polygons including their properties and neighbouring relationships are stored into local variables. Using 8000 basic areas these preparatory works needs up to half an hour. After this the allocation starts which needs running time again. Especially the local search needs a lot of time, thus the calculation time is higher than the one of Map&Market. Map&Market needs 50 minutes, the algorithm

which was implemented here needs 1:20h. Nevertheless the approach of the thesis owns some advantages compared to the one used in Map&Market. Thus the compactness is considered and contiguity will be achieved if the data is applicable doing so. However some more improvements are still necessary. Some preparation within the database may yield to faster calculations as well. During the calculation of the two test cases some problems could be identified especially concerning performance. To make the algorithm much faster a lot of improvements were done. First of all the number of accesses to the database were reduced. Therefore a lot of queries were removed. Instead of this just one access to the database in the beginning is used now which stores all necessary data. This one is just one example of enhancement that were done.
Additionally to the two comparisons of use cases the algorithms were tested to different investigation areas using several parameters. Thereby some weaknesses could be determined. Especially the implementation of the local search yields to some problems. They are caused by rearranging same basic areas again and again so that no better balance will be achieved. That is why an abortion was implemented after a determined running time or number of rearrangements. This problems leads to two weaknesses. First it may be happen that the predefined balance threshold will be not reached. Consequently the territories are not well balanced. Second the running time is maybe still high although no better balance is achieved. That is why an advancement of that algorithm is still necessary in future work. However the implemented algorithms for area segmentation, Whitespot analyses and Greenfield analyses are good prototypes doing area segmentation. Compared to other applications they yield to convincing results within an adequate running time. Nevertheless some improvements are necessary concerning the balance and the performance. Additionally it is not satisfied that an optimal solution of area segmentation is found, but therefore the running time is still fast.

9 Discussion and Perspective

9.1 Summary

Within that thesis area segmentation processes were considered. Therefore in the beginning some related works were described. Afterwards basic knowledge of the topic were explained additionally to criteria that are often used. Especially contiguity, balance and compactness were illustrated in more detail because these criteria are applied mostly. For the reason that the implemented algorithms are applicable for Geomarketing strategies, these three criteria will be considered here, too. For determining promising algorithms that will be implemented previous work and approaches were examined. Thereby three different types of models could be identified. For each type advantages and disadvantages were shown in section 4.2 "Comparison of model types". In conclusion the heuristic approaches could be defined as most usable algorithms for the application to Geomarketing strategies. Afterwards existing heuristics are surveyed. Additionally some new ones were developed by combining or improving existing approaches so that in the end 8 different algorithm were defined for implementation. Before the implemented algorithms were described, first some information about the investigation and the data were given. Afterwards each algorithm was explained including its workflow and results. As soon as all algorithms were implemented a comparison of them was done in section 6 "Comparison of implemented approaches". Thereby three variables were defined that denote the criteria compactness, balance and contiguity. The comparison shows that the approach called AllocMinDistLocalSearch yields to the best results concerning the three parameters. Nevertheless some improvements were necessary to satisfying the requirements from the field of Geomarketing during the calculation. Per example a check of contiguity needed to be implemented to get coherent territories. The improvements were integrated into the algorithm so that an application to Geomarketing strategies was possible. The algorithm was used calculating area segmentations and optimizations, Greenfield analyses and Whitespot analyses. All three types of calculations yields to good results. However some weaknesses were identifiable during the application so that additionally improvements are necessary within the future. These were discussed in section 8 "Evaluation". Additionally in that section the results were compared to ones provided by other Geomarketing software. That confrontation shows some advantages of the implemented algorithm for example the check of contiguity or the performance of some test cases. Consequently the goal of developing an usable algorithm which contains some advantages compared to other algorithms was reached. At the same time it is recognizable that some enhancements are necessary to make the calculation more stable.

9.2 Perspective

The comparison of the implemented algorithm to existing Geomarketing tools showed that the results are promising. The calculations runs in the most of the cases faster and the results are looking more convincing. Especially the implementation of an used threshold value and a weighting value of balance and compactness presents a great improvement compared to other tools. However, some additional enhancement will be necessary within the future. The greatest problem of the implemented algorithm is the realization of the local search. It is possible that always the same basic areas will be rearranged without getting better balanced activity measures. Consequently it is needed to find an approach which do not show that problem and yields to balanced territories always. It may be recommendable to compare the results to one other approach for example the one implemented by Kalcsics et al. [KNS05]. In his work an algorithm was developed using consistent tree decomposition which is based on an location-allocation approach. The drawback of the implemented heuristic approach of that thesis is that just one solution will be determined. Within the location-allocation approach the best resolution will be calculated. Consequently a comparison of the results of both algorithms would be advisable to evaluate the quality of the results of the heuristic approach. It is recommend to analyse which approach yields to the best result considering the ratio of quality and running time. It was already mentioned in section 4.2 "Comparison of model types" that location-allocation algorithms need a long running time to get a result of the area segmentation. It may be important to compare the running time on the basis of a specific example to proof that conclusion.

Furthermore the application of the algorithm to a larger dataset like zip-codes of whole Germany shows that problems exist if island are used as basic areas. In such cases it is necessary at the moment to turn off the check of contiguity. For future work it will be a good idea finding an approach which considers coherence also if island are used. Especially the check of contiguity is one of the great advantages of that algorithm to other tools, consequently it should be usable within mostly all investigations. But the check of coherent territories needs a lot of performance so that the running time raises up if huge territories are created. That is why there also an improvement is needed to get a satisfying result within an adequate running time.

The area segmentation is based on an allocation of basic areas by distance. Surveys show that using just linear distances may yield to results which no Geomarketing analyst would create in reality because geographical barriers as mountains or rivers would be considered during the creation by hand. That is why in future besides linear distance additionally driving distance and driving time will be considered as well.

As soon as all improvements are implemented the prototype of the developed algorithm of that thesis will be a powerful and dependable tool doing Geomarketing analyses effectively. The developed prototype is a good base to simplifies analyses from the field of Geomarketing for analysts to use it within a software tool.

References

[BL14] A. Butsch and P. Ludwig. *Ergebnisse eines visuellen Tests im Vergleich zu klassischen Kompaktheitsmaßen am Beispiel der Landtagswahlkreise in Baden-Württemberg.* Fakultät für Wirtschaftswissenschaften, Institute for Operation Research, KIT, 2014.

[Bod73] L. D. Bodin. *A districting experiment with a clustering algorithm.* Number Vol. 19, pp. 209-214. Annals of the New York Academy of Sciences, 1973.

[DB] Statistisches Bundesamt Der Bundeswahlleiter.
URL: http://www.bundeswahlleiter.de/de/bundestagswahlen/BTW_BUND_13/,
last visit: 30th of October 2015.

[Dec77] R. F. Deckro. *Multiple Objective Districting: A General Heuristic Approach Using Multiple Criteria.* Number Vol. 38, No. 4 Part 2 ,pp. 953-961. Palgrave Macmillan Journals, Operational Research Society, 1977.

[For64] E. Forrest. *Apportionment by computer.* Number Vol. 23, No. 7 ,pp. 23-35. American behavioral scientist, 1964.

[FP98] B. Fleischmann and J. N. Paraschis. *Solving a large scale districting problem: a case study.* Number Vol. 15, No. 6 ,pp. 521-533. Comp. Opns. Res., 1998.

[FS95] J. M. Frühling and W. Steingrube. *Geomarketing: Neue Begriffe = neue Methoden?* Number 39(3-4),pp. 184-194. Zeitschrift für Wirtschaftsgeographie, 1995.

[GLW97] J. A. George, B. W. Lamar, and C. A. Wallace. *Political District Determination Using Large-scale Network Optimization.* Number Vol. 31, No. 1, pp 11-28. Frauenhofer Institut Techno- und Wirtschaftsmathematik ITWM, 1997.

[Gmb] Lutum+Tappert Dv-Beratung GmbH.
URL: http://www.geomarketing.de/,
last visit: 03th of November 2015.

[Gmb12] GfK Geomarketing GmbH. *Praxiswissen Geomarketing.* GfK Geomarketing GmbH, 2012.

[GN70] R. S. Garfinkel and G. L. Nemhauser. *Optimal political districting by implicit enumeration techniques.* Number Vol. 16, No. 8, pp. B495-B508. Institute for Operations Research and the Management Sciences, 1970.

[HM08] M. Herter and K.-H. Mühlbauer. *Handbuch Geomarketing.* Herbert Wichmann Verlag, 2008.

[HS71] S. W. Hess and S. A. Samuels. *Experiences with a sales districting model: criteria* *and implementation.* Number Vol. 18, No. 4, Part II. Institute for Operations Research and the Management Sciences, 1971.

[HWS65] S. W. Hess, J. B. Weaver, and S. A. Samuels. *Nonpartisan Political Redistricting by Computer.* Number Vol. 13, No. 6, pp. 998-1006. Institute for Operations Research and the Management Sciences, 1965.

[ibm] Zim-Tec Solutions in business mapping.
URL:http://www.zim-tec.de/html/standard/home.htm,
last visit: 16th of November 2015.

[Kö12] Peter Köhler. *Kompaktheitsmessung von Flächen und punktförmigen Basisgebieten.* Fakultät für Wirtschaftswissenschaften, Institute for Operation Research, KIT, Bachelorarbeit, 2012.

[KMNG02] J. Kalcsics, T. Melo, S. Nickel, and H. Gründra. *Planning sales territories - a facility location approach.* Number pp. 141-148. Operations Research Proceedings 2001, Springer Verlag Berlin, 2002.

[KNS05] J. Kalcsics, S. Nickel, and M. Schröder. *Towards a Unified Territory Design Approach Applications, Algorithms and GIS Integration.* Frauenhofer Institut Techno- und Wirtschaftsmathematik, 2005.

[Mü] Dennis Müller.
URL: http://www.omkt.de/geomarketing/,
last visit: 29th of October 2015.

[Mar81] P. G. Marlin. *Application of the transportation model to a largescale districting' problem.* Number Vol. 139, No. 3, pp. 521-532. Comp. Opns. Res., 1981.

[MJN98] A. Mehrotra, E. L. Johnson, and G. L. Nemhauser. *An optimization based heuristic for political districting.* Number Vol. 44, No. 8, pp. 1100-1114. Institute for Operations Research and the Management Sciences, 1998.

[mMSG] microm Micromarketing-Systeme and Consult GmbH.
URL: http://www.microm-online.de/,
last visit: 03th of November 2015.

[mMSuCG] microm Micromarketing-Systeme und Consult GmbH.
URL: http://www.mapchart.de/,
last visit: 03th of November 2015.

[Nit98] M. Nitsche. *Micromarketing: Daten-Methoden-Praxis.* Wirtschaftsverlag Ueberreuter, Wien/Frankfurt, 1998.

[Nyg88] B. Nygreen. *European assembly constituencies for wales comparing methods for* *solving a political distrcting problem.* Number Vol. 42, pp. 159-169. Math. programm, 1988.

[PDS14] KIT Institute for Operations Research Prof. Dr. S.Nickel. Lizard - library of optimization algorithms for distrcting.
URL: http://dol.ior.kit.edu,
last visit: 03th of November 2015, 2014.

[Sch01] M. Schröder. *Gebiete optimal aufteilen: OR-Verfahren für die Gebietsaufteilung als Anwendungsfall gleichmäßiger Baumzerlegung.* 2001. Dissertation, Fakultät für Wirtschaftswissenschaften der Universität Fridericiana zu Karlsruhe.

[Tap07] W. Tappert. *Geomarketing in der Praxis.* Bernhard Harzer Verlag, 2007.

[You88] H. P. Young. *Measuring the Compactness of Legislative Districts.* Number Vol. 13, No. 1,pp. 105-115. Legislative Studies Quarterly, 1988.

[ZS83] A. A. Zoltners and P. Sinha. *Sales territory alignment: a review and model.* Number Vol. 29, No. 11. Institute for Operations Research and the Management Sciences, 1983.