A MATHEMATICAL MODEL AND APPLICATION ON THE PREVENTION OF GERRYMANDERING

by
SÜLEYMAN TOPDEMİR

Submitted to the Graduate School of Engineering and Natural Sciences
in partial fulfillment of the requirements for the degree of
Master of Science

Sabancı University

January, 2019

A MATHEMATICAL MODEL AND APPLICATION ON THE PREVENTION OF GERRYMANDERING

APPROVED BY:

Assoc. Prof. Dr. Kemal Kılıç
(Thesis Supervisor)

Prof. Dr. Şevket İlker Birbil

Asst. Prof. Dr. Fatma Tevhide Altekin

DATE OF APPROVAL: 14/12/2018

to the journey

Acknowledgements

First, I would like to thank my thesis supervisor Kemal Kılıç for the continuous support throughout my master's degree. I am grateful for his precious guidance in every respect and having the chance to work with him.

I also would like to thank Ş. İlker Birbil for his support and contributions in this research. I am also grateful to the jury member Tevhide Altekin for her valuable time.

My family and friends, I think you will never see these lines, but I want to express my gratitude in here anyway. I want to thank my all friends, luckily too many to count, for their support and the moments we shared.

STRATEJİK TAKSİMATIN ÖNLENMESİ UZERİNE MATEMATİKSEL MODEL VE UYGULAMA

Süleyman Topdemir

Endüstri Mühendisliği, Yüksek Lisans Tezi, 2019

Tez Danışmanı: Doç. Kemal Kılıç

Anahtar Kelimeler: stratejik taksimat, dar seçim bölgesi, daraltılmış seçim bölgesi, matematiksel model, tabu arama algoritması, İstanbul

Özet

Seçim bölgelerinin yeniden belirlenmesi problemi çeşitli ülkelerin üzerinde durduğu bir konu olarak literatürde yer almaktadır. Belirli bir partiye veya gruba siyasi avantaj sağlanılacak şekilde seçim bölge sınırlarının belirlenmesi pratiği stratejik taksimat olarak bilinmektedir ve seçim sonuçlarını doğrudan etkilemektedir. Bu tez dahilinde, seçim bölgelerinin yeniden belirlenmesi probleminin ne kadar suistimale açık olduğu gösterilmeye çalışılmıştır. Dar-bölge ve daratılmış-bölge seçim sistemleri için 2 farklı matematiksel model geliştirilmiş ve İstanbul ilçeleri üzerinde matematiksel modeller sınanmıştır.

Matematiksel modeller sadece kısıtlı sayıda birimin yer aldığı uygulamalarda çalışabilmektedir. Bu dezavantajı ortadan kaldırabilmek adına tabu arama algoritması isimli meta sezgisel algoritma geliştirilmiş ve İstanbul'un Asya kıtasında yer alan tüm mahallelerin dikkate alındığı bir problem bu algoritma ile çözdürülmüştür. Sonuçların görsel olarak yorumlanabilmesi için gerekli haritalar oluşturulmuştur.

A MATHEMATICAL MODEL AND APPLICATION ON THE PREVENTION OF GERRYMANDERING

Süleyman Topdemir

Industrial Engineering, Master's Thesis, 2019

Thesis Supervisor: Assoc. Prof. Kemal Kılıç

Keywords: gerrymandering, single-member district, multi-member district, mathematical model, tabu search algorithm, İstanbul

Abstract

The significance of the political districting also known as redistricting has been recognized by several countries across the world since electoral district boundaries can be manipulated for a political gain. This manipulation practice is known as gerrymandering and has serious influences on the results of an election. In this thesis, we tried to show how easy policymakers can misuse the redistricting practice to gain a political advantage such as increasing the number of their representatives in the parliament. Two different mathematical models have been developed for different types of election systems. Single-member district electoral system in which the only representative can be elected from each electoral district is one of them. Every county in İstanbul has been tried to be divided into their single-member districts considering the total number of representatives of the county. In addition to the first model, another formulation has been developed to also cover the multi-member district systems.

The main drawback of the mathematical models is that they are only working on the small cases in terms of the total number of political units. Tabu search algorithm has been developed to answer the cases that cannot be classified as small. The required algorithm steps such as initialization, neighborhood change structure etc. are explained in detail. The results of the mathematical models and the algorithm have been achieved and visualized aesthetically.

Contents

List of Figures

List of Tables

Chapter 1

INTRODUCTION

”Parliamentary democracy is a democratic form of government in which a party or a coalition of parties with the highest number of representatives in the legislature forms the government”(Britannica, 2019). The term of electoral district is related to this type of democracy and refers to a territorial subdivision for electing representatives to a legislative body. In other words, the residents in each electoral district are permitted to elect one or some of the candidates participated in the election of their district. There are different types of electoral districts in terms of the maximum number of candidates that represent their district in the parliament. The single-member electoral district, for instance, is one type of it. The term single-member refers to the case when only one representative who can be elected from an electoral district. The single-member electoral district is also known as the single-winner electoral system. The rule of winner-takes-all is in use for single-member electoral district systems. The multi-member electoral district is another type. In a multi-member district, more than one representatives can be elected pursuant to the proportion of the total votes of each candidate. Deciding on the type of electoral system has the potential to change the final total number of representatives of each party in the legislature. The advantages or disadvantages of these electoral systems are beyond the scope of this thesis.

In this thesis, the political districting problem which is abbreviated as PD for both of the electoral systems will be studied. PD problem is a specific version of the districting problem in which the aim is to aggregate the areal units into the set of districts with respect to an objective and subject to certain constraints on the zones. The design of districts for schools, social facilities, and sales/service territory are examples of other types of districting problems, also known as the zone design problem. Political districting consists of the partitioning of areal units, generally administrative units such as neighborhoods, into a prespecified number of districts that satisfy some criteria (Bacao *et al.*, 2005). Population equality, for example, is one of the most common criteria for the redistricting problem. Contiguity and the shapes of each district are other important criteria for redistricting.

A major consideration during the redistricting is preventing the *gerrymandering* that can be defined as a practice that involves manipulating the electoral district borders in favor of a particular party or group. The term gerrymander is used for the first time in Boston Gazette on 26 March 1812 to draw attention to an electoral district plan that has a salamander-shaped district prepared by the governor of Massachusetts Allbright Gerry in 1812. The caricature depicted the salamander shape redistricting plan is shown in Figure 1.1.

Figure 1.1: A caricature depicted the salamander-shaped district plan

One can easily claim that selecting districting criteria is the most important step of the redistricting problem since the resulting plan is determined by the chosen criteria. Districting criteria will be mentioned in Chapter 2 in detail. The other important factor of the problem is the geographical and demographic data. Here geographical data refers to the shape of each political unit that is assigned to an electoral district by combining each other. On the other hand, the term of demographic data refers to the population of the political units or a proportion of an ethnic group in the population. We used each indivisible political unit's spatial data, the population, and voting shares of available parties in the past elections and referendum as input.

1.1 Overview

This thesis provides a real-life application example with numerical experiments for the redistricting of electoral districts in İstanbul. Turkey was in the process of modifying its election system when we started this thesis. Currently, the multi-member electoral district is in use in Turkey. Policy makers work on a change to the single-member district for the electoral system. Motivated by this, we aim to partition the territory into a certain number of zones while satisfying some districting criteria. At first, a mathematical model with an objective function that maximizes the total number of representatives of a particular party while satisfying the districting criteria has been developed. We tried to obtain more than one redistricting plans that satisfy the required districting criteria. One of the major aims of this thesis is obtaining some redistricting plans in which a party can be represented by different number of representatives. For example, a party can dominate the other parties in a district according to a redistricting plan. Changing the boundries, this dominance can be reversed. The fair representation is beside the point we emphasize. However, utilizing the data excluding the total votes for each party in each neighborhood, a new study related to the fair representation can be achieved.

Secondly, the first formulation is enhanced to also answer for the multi-member district electoral system. Even though this electoral system is currently in use, a hybrid version of it could be an alternative for the policymakers according to the newspaper news. Existing electoral districts are planning to be divided into subdistricts in which a prespecified number of representatives can be elected. The difference between the single-member district and multi-member based hybrid electoral district is this prespecified number. In single-member version, the value must be equal to one. However, the value can be more than one and can differ between the subdistricts in another intended system. Note that, the number of representatives of each party will be allocated according to the D'Hondt method that will be mentioned in the further sections in detail.

Some applications cannot be solved by using these aforementioned formulations and we need some heuristics approaches. In this thesis, we have developed a Tabu Search algorithm that is known as a metaheuristic algorithm to also consider the cases that cannot be solved the exact approaches.

The mathematical models are originated from an exact approach to solve the PD problem in the literature (Nemoto and Hotta, 2003). The original formulation is minimizing the difference between the maximum and minimum population of the districts. The contiguity of the districts is satisfied by the corresponding constraints in the model directly. Taking into consideration the thesis objectives and the available mathematical model, we came up with a new formulation whose objective is maximizing the total number of representatives of a particular party while satisfying some criteria. Two models are still satisfying the contiguity of the districts criterion. One important difference between these models is that the population equality of the districts is enforced by the objective function in the original model, however, this criterion can be handled by the constraints that need a new parameter.

Another important contribution is that we also have a formulation for multi-member district electoral systems. The allocation of the representatives must be handled by the mathematical model for the multi-member district. The structure of the constraints related to the finding the number of representatives according to their total votes is generalized in the second model. We do not encounter the approach we use in the formulation in the PD literature.

Lastly, there is no known study on the political districting problem in Turkey in the literature. The problem has been mostly studied in the United States of America because of the regulations. We can claim that this thesis is the first scientific study on a PD problem in Turkey considering the mathematical models and the algorithms we developed.

The rest of the thesis is organized as follows. In Chapter 2, the districting criteria are introduced in detail. In Chapter 3, we present a comprehensive summary of the related literature. Details of the developed mathematical models are given in Chapter 4. In Chapter 5, we give the details of the algorithm we use. Chapter 6 is reserved for the computational results. Lastly, in Chapter 7, we conclude the study with the discussion of the results and future study.

Chapter 2

DISTRICTING CRITERIA

2.1 Population Equality

In the redistricting process, equal populated districts are one of the most looked for outcomes since it's related to the *one-person-one-vote*. The main reason to utilize such a criterion is to maintain that each citizen is represented in the legislature with equal weight. An exact equality cannot be satisified in every case; however, the values should be close enough to an average population value.

Technically speaking, the population of the determined districts must be in a narrow allowable range. The size of this range can be prespecified using some parameter values. For example, one can request that the population value of each district must be greater than 95 percent of the average population and less than 105 percent of the average population. ± 5 percent allowable deviation is used in this example. There are several metrics to measure the quality of a district in terms of population equality in Bozkaya (1999). General idea behind the approaches is similar. Some of the metrics are below. The main aim of these metrics is that determining how the population values of each district are different from each other. Mathematically, let n denote the number of districts, s_j population of district j, s_{min} population of the smallest district, s_{max} population of the largest district, $\bar{s}$ is the average population which can be calculated $\sum_j s_j/n$ and β the allowable percentage deviation from $\bar{s}$.

- Mean absolute deviation: $\sum_j |s_j - \bar{s}|/n$
- Mean squared deviation: $\sum_j |s_j - \bar{s}|^2/n$
- Maximum absolute deviation: $max_j|s_j - \bar{s}|$
- Extreme deviation: $(s_{max} - s_{min})/\bar{s}$
- Extreme ratio: s_{max}/s_{min}

In this thesis, we check the population values of each district one by one in the formulation. One of the metrics above can be utilized in a fitness function of a metaheuristic. Alternatively, an allowable percentage deviation can be used to control whether a district violates the population equality criterion or not. For such an approach, we only need the values of β, $\bar{s}$ and the population of the considered district. Note that, the population equality is one of the constraints in the formulation. In tabu search algorithm, this criterion is controlled in each iteration to determine whether a candidate solution will be eliminated or not.

2.2 Contiguity

The second criterion is the contiguity of the districts. A district is said to be contiguous if every part of the district is reachable from every other part without crossing the district boundary (Grofman, 1985). Contiguity is directly related to the feasibility of the solution. Consider a district in which a political unit has a connection with none of the other units. This kind of redistricting would be worse than the gerrymandering problem. That is to say, even if the shape of the districts is not proper, each unit in a district must be connected to each other somehow.

2.3 Compactness

The third criterion is the compactness. This term is defined as firmly put together, joined, or integrated; predominantly formed or filled in Dictionary (2019). The term can be understood better by using some visual materials since the term is ambiguous. In Figure 2.1 some district plans with more and less compactness are illustrated. In Niemi *et al.* (1990); Young (1988) some of the compactness measures are introduced. Each of them is developed heuristically since there is no exact score for the compactness criterion. The main idea behind the metrics for the compactness is similar. We are trying to figure out how close a geometric shape is to an ideal compact shape. Here the ideality assumption can differ among the compactness measurements.

Different measurements of a geometric shape can be used to calculate its compactness score. For example, the area and perimeter of the geometric shape are some of them and denoted by A and P. In Figure 2.2 illustrations of the scores are provided.

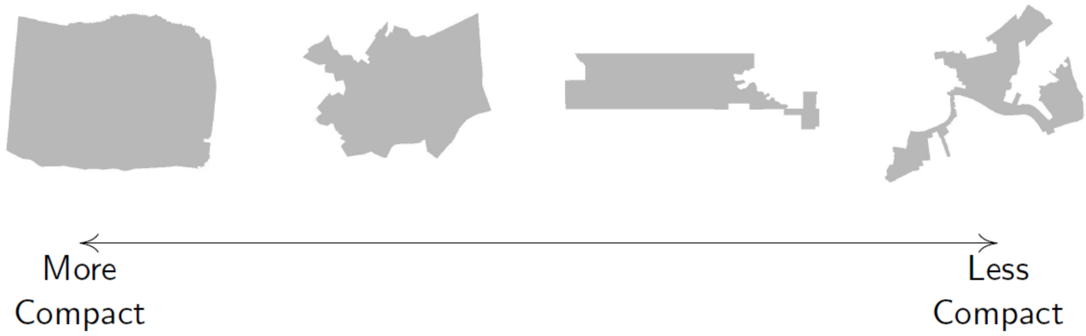


Figure 2.1: Illustration of more and less compact districts

2.3.1 Polsby-Popper Measure

This measurement type is introduced in Polsby and Popper (1991). The main assumption of this method is that an ideal district should be a circle. The formula for the Polsby-Popper measure is below. This value is equal to one when the geometric shape is a circle. In other cases, the value is smaller than one.

$$4\pi A/P^2 \tag{2.1}$$

2.3.2 Convex Hull Measure

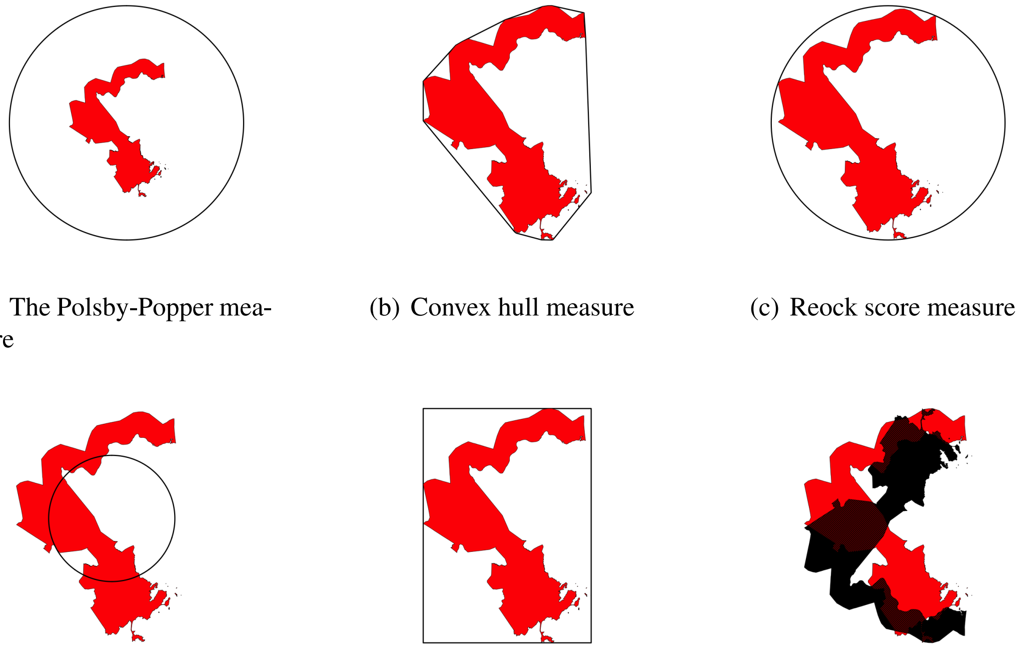
One of the compactness measurements is convex hull measure. The convex hull of a geometric object can be defined as the smallest convex set which covers the object. In the redistricting context, the geometric object refers to the spatial object. We need two areas to calculate the convex hull compactness measure; the area of the spatial object (district) and the area of the convex hull of this spatial object. The ratio of the first measurement to the second one is called convex hull score. Note that the area of the convex hull is equal to or greater than the area of the spatial object since the convex hull must enclose the geometric shape which can be a concave shape. That is to say, convex hull measure value can be in between 0 and 1 including 1. Let A_{ch} denote the area of the convex hull of a geometric shape. The convex hull score is calculated as follows:

$$A/A_{ch} \tag{2.2}$$

2.3.3 Reock Score Measure

Reock score is one of the area dispersion measure for compactness and the ratio of the area of the district to the area of the smallest circle that encloses the district (Corcoran and Saxe, 2014). Let A_{sc} denote the area of the smallest circle that covers the district, then the required score is calculated as follows:

$$A/A_{sc} \tag{2.3}$$



(a) The Polsby-Popper measure (b) Convex hull measure (c) Reock score measure

(d) Schwartzberg measure (e) Length Width measure (f) X-symmetry measure

Figure 2.2: Compactness measures

2.3.4 Schwartzberg Measure

In Schwartzberg (1965), Schwartzberg score is defined as the ratio of the perimeter of a district to the perimeter of a circle having the same area with the district. The score is calculated as follows:

$$P/\sqrt{4\pi A} \tag{2.4}$$

2.3.5 Length Width Measure

In order to calculate this compactness metric, we need to form a rectangle which touches the district on all four sides and covers the district such that the ratio of the length of the rectangle to the width of it maximum (Young, 1988).

2.3.6 X-symmetry Measure

The ratio of the total intersection area of the shape of district and its reflection across the x-axis to the area of the district. Let A_i denote the intersection area. The score is calculated as follows:

$$A_i/A \tag{2.5}$$

Chapter 3

LITERATURE REVIEW

There are several mathematical and numerical approaches to solve the PD problem in the literature. Integer programming, set partitioning, implicit enumeration and local search methods are some of these approaches. In Nygreen (1988), the first three approaches are compared on an application that covers a PD problem in Wales. In Ricca and Simeone (2008), four local search metaheuristics that are descent, tabu search, old bachelor acceptance, and simulated annealing are compared. According to the paper, different algorithms yield the best results for different criteria. For example, the simulated annealing algorithm performs the best result for the population equality criterion.
The desired outputs of an automated redistricting practise were explained for the first time in Vickrey (1961). In this short paper, gerrymandering is explained and illustrated with a small example. 1960s new computational technology was applied on the redistricting problem. The author aimed to answer how a redistricting process must be executed before proposing a greedy approach, multi kernel growth, as the solution of the problem. The process should be completely mechanical, and no one can predict in any detail the outcome of the process according to the author. Besides these questions and solution techniques, why we need such a process is explained in detail.

Even though the problem began with Vickrey's work, the first mathematical model was introduced in Hess *et al.* (1965) for the problem. Their claim was that a technique that is rapid and nonpartisan is needed for PD problem. The authors developed an integer programming model which is basically for a warehouse allocation problem minimizing the assignment cost in terms of distance between the centers and unit territories. However, at the time computational power was limited to solve even small cases. A heuristic algorithm which considers the criteria of population deviation, compactness and contiguity was needed. In the case of a tie in terms of given measures, a simple algorithmic approach was used in this article. However, they stated that more sophisticated approaches may be developed. In this work, contiguity is not an explicit constraint in the mathematical model and heuristic approach. The authors couldn't use the mathematical model to attain results due to the size of the problem and computational power at that time. In the heuristic approach, the contiguity constraint is satisfied by eliminating noncontiguous solutions. In the mathematical model below, I is the set of units and x_{ij} is the binary decision variable, which is 1 if the j^{th} unit is assigned to the i^{th} center. $|I|$ units are tried to be divided into $|H|$ districts. Here H is the set of districts we are trying to obtain. a and b represent the minimum and maximum allowable district populations, respectively, as a percent of the average district population. d_{ij} is the euclidean distance between i^{th} and j^{th} units. s_j denotes the population of the j^{th} unit. The case of the value of variable x_{jj} is equal to 1 indicates that j^{th} unit is the center of a district. Note that $\forall i, j \in I$ for all of these parameters and variables.

$$min \sum_{i \in I} \sum_{j \in I} d_{ij}^2 s_j x_{ij} \tag{3.1}$$

$$s.t.$$

$$\sum_{i \in I} x_{ij} = 1, \forall j \in I \tag{3.2}$$

$$\sum_{j \in I} x_{jj} = |H| \tag{3.3}$$

$$a\bar{s}x_{jj} \leq \sum_{i \in I} x_{ij} p_i \leq b\bar{s}x_{jj}, \ \forall j \in I \tag{3.4}$$

$$x_{ij} \in \{0, 1\}, \ \forall i, j \in I \tag{3.5}$$

In this formulation, the objective function is used to measure the compactness of the districts while the population equality criterion is taken into account by using constraints 3.4. These constraints also enforce that no basic unit can be assigned a unit which is the center of a district. Constraints 3.2 ensure that each unit is assigned to only one district. The prespecified number of districts is forced by the constraint 3.3.

The first exact approach for the solution of PD problem was developed in Garfinkel and Nemhauser (1970). This approach has two phases. In the first phase, the authors aim to attain all possible feasible district plans. In the second phase, a mathematical programming problem is solved to determine the best districts among the feasible solutions. The problem solved in this step is a version of the set-covering problem. A problem with 37 political units that need to be divided into 7 electoral districts is solved by using this two-phase exact approach. However, they found that an application with 55 political units cannot be solved.

In Nemoto and Hotta (2003), an exact approach was introduced, and the redistricting problem was reformulated in terms of *network flows*. First, the authors represent the problem as a graph. *G=(I,A)* denotes a contiguity graph with a set of nodes I and set of arcs between the nodes A. Here the political units and adjacency of the units are represented by the nodes and arcs in the graph respectively. After this representation, the authors developed a new network which is originated from the first one; $T = (\bar{I}, \bar{A})$. According to this network, each arc connecting node i and node j in G is replaced by the pair of arcs (i, j) and (j, i); n copies of this graph are formed, and node i in the h^{th} copy of the graph is denoted by w_i^h; $\forall i \in I, \forall h \in H$. Note that H denotes the set of districts; in other words, the copies of graph. In addition, $|H|$, which is equal to n, source-nodes and $|I|$ sink-nodes a.k.a tail-nodes are introduced. Each source-node s^h, $\forall h \in H$ is connected to all the nodes of the h^{th} copy of the graph with an arc (s^h, w_i^h) $\forall i \in I$, while for each sink-node t_i, $\forall i \in I$ there exists an arc (w_i^h, t_i), $\forall h \in H$. After introducing the additional nodes and arcs, $\bar{I}$ covers source-nodes (s^h), sink-nodes (t_i) and the nodes in the copies of the graph (w_i^h); $\forall i \in I$, $\forall h \in H$. $\bar{A}$ covers the arcs between the source-nodes and the nodes of the copies of the graph (s^h, w_i^h), the arcs between the nodes w_i^h and w_j^h , and lastly the arcs between the nodes of the copies of the graph and sink-nodes (w_i^h, t_i); $\forall i, j \in I$, $\forall h \in H$. The criteria of the contiguity and population equality are considered in the model.

$$min \qquad u - l \tag{3.6}$$

$$s.t.$$

$$l \leq \sum_{i \in I} p_i x_{ih} \leq u, \forall h \in H \tag{3.7}$$

$$\sum_{i \in I} y_{ih} = 1, \forall h \in H \tag{3.8}$$

$$f(s^h, w_i^h) = F y_{ih}, \forall i \in I, \forall h \in H \tag{3.9}$$

$$\sum_{a \in \delta^-(w_i^h)} f(a) = \sum_{a \in \delta^+(w_i^h)} f(a), \forall i \in I, \forall h \in H \tag{3.10}$$

$$\sum_{a \in \delta^-(w_i^h)} f(a) \leq F x_{ih}, \forall i \in I, \forall h \in H \tag{3.11}$$

$$x_{ih} \leq f(w_i^h, t_i), \forall i \in I, \forall h \in H \tag{3.12}$$

$$\sum_{h \in H} x_{ih} = 1, \forall i \in I \tag{3.13}$$

$$f(a) \geq 0, \forall a \in \bar{A} \tag{3.14}$$

$$u, l \geq 0 \tag{3.15}$$

$$x_{ih}, y_{ih} \in \{0, 1\}, \forall h \in H, \forall i \in I \tag{3.16}$$

The objective function of the formulation minimizes the maximum population difference between the districts. The objective function is directly related to the constraints 3.7. In these constraints, the model provides the maximum and minimum populations of the generated districts which are used in the objective function. Each source node must be assigned to one of the districts in the constraints 3.8. In 3.9, the flow amount from each source node to the nodes in the original problem must be equal to a predetermined value if the corresponding arc is used. The total flow on the entering arcs must be equal to the total flow on the leaving arcs for each node according to the contraints 3.10. In 3.11, if node i is not assigned to district h, then the total flow on entering arcs to this unit must be zero. If the flow between node w_i^h and tail node t_i is zero, then the node (unit) i cannot be assigned to the district h according to the constraints 3.12. Each unit must be assigned to a district, in other words, each node must be used in one of the districts (graph copies) 3.13. The constraints 3.14 and 3.15 enforce that the flow on arcs, u, and l values are nonnegative. The last constraints are for defining the domains of the decision variables.

In the early 1990s, several types of metaheuristic algorithms came into use for the redistricting problem. In Browdy (1990), the simulated annealing algorithm has been used. A generic version of the simulated annealing algorithm is given in Algorithm 1 at the end of this chapter. The paper states that the PD problem can be modeled as a constrained optimization problem with some constraints for contiguity, population equity, and compactness. However, this problem cannot be solved using an exact approach and the constraints must be relaxed. The formulation is relaxed by constituting an objective that covers the aforementioned constraints with their weights. They call this equation as *energy* and tried to minimize this energy value using the simulated annealing algorithm. In order to choose the appropriate weights, the study did not provide any particular method. Decision makers can choose the required values for the weights depending on the case they handled according to the author.

Tabu search heuristic algorithm was also used in the literature (Bozkaya et al., 2003). Some additional criteria such as socio-economic homogeneity and similarity to the existing plan. The political districting problem was formed a multi-objective optimization problem by using the weigths for each criterion. A tabu search with an adaptive memory structure was used in the study. The results of the city of Edmonton were provided using a geographical information system software.

Genetic algorithm with different encodings has been extensively used in the literature for the redistricting (Forman and Yue, 2003; Bacao *et al.*, 2005; Liu *et al.*, 2016; Vanneschi *et al.*, 2017). The problem is modeled as a multiobjective optimization model since the nature of the redistricting involves the multiple criteria. The fitness function of the algorithm covers the compactness measure and population equality. The solutions that have noncontiguous districts are eliminated if the fitness function only covers the compactness and population equality. However, the contiguity criterion takes part in the fitness function in Forman and Yue (2003). In this paper, the shape fitness function modifies the Schwarzberg measure to take into consideration the contiguity criterion by multiplying this value one plus the number of excess discontiguous pieces found in the district weighted by a parameter.

Different types of encodings are used for the genetic algorithm in the PD literature. A TSP-based encoding has been used in Forman and Yue (2003). As in the TSP, a single chromosome covers each unit. The step of transforming the chromosome to the district plans is called the conversion. At the conversion step, the algorithm starts with the first gene and travels along the chromosome summing the populations until some threshold population is met. A small example can be found in Figure 3.1.

1_{30}	2_{20}	3_{10}
4_{10}	5_{20}	6_{30}
7_{10}	8_{20}	9_{30}

(a) 3x3 territory

$$\mathbf{A} = 1 - 4 - 5 - 2 - 3 - 6 - 9 - 8 - 7$$

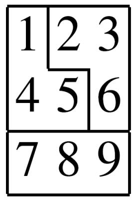
(b) an example of chromosome

Figure 3.1: A small example of TSP-based choromosome

The unit or track populations is given in the subscripts. An example of the TSP-based choromosome is also given. The 9 tracks are divided into 3 districts using the mentioned procedure in Figure 3.2.

Tract	Pop.	$\sum$Pop.	District
1	30	30	1
4	10	40	1
5	20	60	1
2	20	20	2
3	10	30	2
6	30	60	2
9	30	30	3
8	20	50	3
7	10	60	3

(a) Assignment of the tracks also known as units



(b) The result redistricting plan

Figure 3.2: The redistricting plan of the given choromosome

In Bacao *et al.* (2005), two different encoding schemes are proposed. The first one considers the political unit as a centroid. Another encoding scheme consists of the coordinates value of a political unit. Length of the first encoding is half of the second one since any coordinate can be represented by longitude and latitude values. At first, the algorithm chooses the centroid, after this step, the nearest political units are assigned to this centroid. Encoding scheme is similar to the k-means practice which usually used in machine learning community (Hartigan and Wong, 1979).

In Liu *et al.* (2016) a computational approach utilizing the parallel high-performance computing for the PD problem has been developed. They used the evolutionary algorithm since they have a proper structure for parallelization. The encoding scheme is different from the other papers. The main reason to choose a special encoding scheme is that the authors would like to use the advent of parallel computational power. The length of each chromosome is the number of total political units. Each gene represents the district to which corresponding political unit is assigned. They used two different initialization approaches that have similar steps. The fitness function of the algorithm covers the population equality and compactness criteria. At each iteration, the solutions are checked whether the solution is contiguous or not. They compare their results to the BARD redistricting software (Altman and McDonald, 2011) in which Simulated Annealing, Greedy, Tabu search and GRASP techniques can be utilized.

In Vanneschi *et al.* (2017), a genetic algorithm with variable neighborhood search was proposed. They used a hybrid algorithm that combines the NSGA-II technique with a variable neighborhood search algorithm. The encoding scheme is the same with the one in Liu *et al.* (2016). They compared their results with the outputs of the techniques such as graph partitioning, simulated annealing, genetic algorithm, and constrained polygonal spatial clustering.

Artificial Bee Colony(ABC) algorithm was used for the PD problem in Rincón-García *et al.* (2015) as another example of metaheuristics. They used a weight aggregation function strategy to handle the multi-objective nature of the problem. The function covers the compactness and population equality measures of the districts associated with their weights. Because of the multi-objective nature of the problem, they run the model 990 times and results are evaluated by using the Pareto sorting procedure in order to decide on the weight parameters values. The authors compared the ABC algorithm with the simulated annealing algorithm and their computational experiments showed that the ABC algorithm produces better quality and efficient solutions than the simulated annealing algorithm .

Algorithm 1 Simulated Annealing

```
s ← initialSolution
bestSolution ← s
t ← maxTemperature
while t > minTemperature do
    iter ← 0
    while t < maxIterations do
        s′ ← select a random solution s′ ∈ N(s)
        Δ ← f(s′) − f(s)
        if Δ < 0 then
            s ← s′
            if f(s′) < f(bestSolution) then
                bestSolution ← s′
            end if
        end if
        if rand(0, 1) < e^(−Δ/t) then
            s ← s′
        end if
        iter ← iter + 1
    end while
    t ← t(1 − α)
end while
return bestSolution
```

Chapter 4

MODELING

In this chapter, we present the details of the two parts of the study. In Section 4.1, the first part, the details of the mathematical model for single-member district, which corresponds to a model for single-member district election system provided. For this part, we also present the notation used in the mathematical model. In Section 4.2, mathematical model for single-member district is generalized. The first mathematical model can work on a system in which only one representative can be elected in each electoral district. However, the mathematical model in Section 4.2 allows that more than one representative can be elected in each electoral district. The second model is more complex than the first one in terms of the total number of variables and constraints. The formulations are originated from a mathematical model in Nemoto and Hotta (2003) and modified for our purposes. The required definitions had been provided in Chapter 3. However, the definitions are also given in this chapter for the sake of completeness.

4.1 Mathematical Formulation for the Single-Member District

G=(I,A) denotes a contiguity graph with a set of nodes I and set of arcs between the nodes A. Here the political units and adjacency of the units are represented by the nodes and arcs in the graph respectively. After this representation, the authors developed a new network which is originated from the first one; $T = (\bar{I}, \bar{A})$. According to this network, each arc connecting node i and node j in G is replaced by the pair of arcs (i, j) and (j, i); n copies of this graph are formed, and node i in the h^{th} copy of the graph is denoted by w_i^h; $\forall i \in I$, $\forall h \in H$. Note that H denotes the set of districts; in other words, the copies of graph. In addition, $|H|$, which is equal to n, source-nodes and $|I|$ sink-nodes a.k.a tail-nodes are introduced. Each source-node s^h, $\forall h \in H$ is connected to all the nodes of the h^{th} copy of the graph with an arc (s^h, w_i^h) $\forall i \in I$, while for each sink-node t_i, $\forall i \in I$ there exists an arc (w_i^h, t_i), $\forall h \in H$. After introducing the additional nodes and arcs, $\bar{I}$ covers source-nodes (s^h), sink-nodes (t_i) and the nodes in the copies of the graph

(w_i^h); $\forall i \in I, \forall h \in H$. $\bar{A}$ covers the arcs between the source-nodes and the nodes of the copies of the graph (s^h, w_i^h), the arcs between the nodes w_i^h and w_j^h , and lastly the arcs between the nodes of the copies of the graph and sink-nodes (w_i^h, t_i); $\forall i, j \in I, \forall h \in H$. The objective function of this model is modified. The sets and parameters are provided in Table 4.1. The required decision variables are introduced in Table 4.2.

Table 4.1: Sets and parameters for the first formulation

H	set of the districts or copies of graph
I	set of the political units
P	set of the parties
C	set of units that cannot take part of the same district for the compactness
$\bar{A}$	set of the arcs between the pair of nodes (s^h, w_i^h), (w_i^h, w_j^h), and (w_i^h, t_i); $\forall i, j \in I, \forall h \in H, i \neq j$
$\delta^-(w_i^h)$	the set of arcs entering node w_i^h; $\forall i \in I, \forall h \in H$
$\delta^+(w_i^h)$	the set of arcs leaving node w_i^h; $\forall i \in I, \forall h \in H$
v_{ki}	the number votes that party k has in unit i; $\forall k \in P$, $\forall i \in I$
s_i	number of voters in unit i; $\forall i \in I$
ρ	the party chosen
$\bar{s}$	the average number of voters
β	the allowable percentage deviation from the average number voters
M	A sufficiently big value
F	the volume of the flow from each source

Table 4.2: Decision variables for the first formulation

x_{ih}	binary decision variable indicating if unit i is assigned to district h $\forall i \in I, \forall h \in H$
$f(a)$	the volume of the flow on arc a; $\forall a \in \bar{A}$
y_{ih}	binary decision variable indicating if the h^{th} copy of G the flow enters through node i; $\forall i \in I, \forall h \in H$
c_{kh}	binary decision variable indicating the party k wins in district h; $\forall k \in P, \forall h \in H$
t_h	the number of voters in district h; $\forall h \in H$
o_{kh}	auxiliary variable which shows the total votes of each party k in each district h $\forall k \in P, \forall h \in H$
z_{kph}	binary decision variable for the comparing of total votes of party k and p in district h; $\forall k, p \in P,\ \forall h \in H, k \neq p$

$$max \sum_{h \in H} c_{\rho h} \tag{4.1}$$

$s.t.$

$$\sum_{h \in H} x_{ih} = 1,\ \forall i \in I \tag{4.2}$$

$$\sum_{i\in I} x_{ih}v_{ki} = o_{kh},\ \forall k \in P,\ \forall h \in H \tag{4.3}$$

$$o_{ph} - o_{kh} + Mz_{kph}\ \geq 0,\ \forall k,p \in P,\ \forall h \in H, k \neq p \tag{4.4}$$

$$\sum_{p\in P, p\neq k} z_{kph} - |P| + 2\ \leq c_{kh},\ \forall h \in H,\ \forall k \in P \tag{4.5}$$

$$\sum_{k\in P} c_{kh} = 1,\ \forall h \in H \tag{4.6}$$

$$\sum_{i\in I} x_{ih}s_i = t_h,\ \forall h \in H \tag{4.7}$$

$$t_h\ \leq \bar{s}(1+\beta),\ \forall h \in H \tag{4.8}$$

$$t_h\ \geq \bar{s}(1-\beta),\ \forall h \in H \tag{4.9}$$

$$\sum_{i\in I} y_{ih} = 1,\ \forall h \in H \tag{4.10}$$

$$f(s^h, w_i^h) = Fy_{ih},\ \forall i \in I,\ \forall h \in H \tag{4.11}$$

$$\sum_{a\in\delta^-(w_i^h)} f(a) = \sum_{a\in\delta^+(w_i^h)} f(a),\ \forall i \in I,\ \forall h \in H \tag{4.12}$$

$$\sum_{a\in\delta^-(w_i^h)} f(a)\ \leq Fx_{ih},\ \forall i \in I,\ \forall h \in H \tag{4.13}$$

$$x_{ih}\ \leq f(w_i^h, t_i),\ \forall i \in I,\ \forall h \in H \tag{4.14}$$

$$x_{ih} + x_{jh}\ \leq 1,\ \forall h \in H,\ \forall (i,j) \in C \tag{4.15}$$

$$f(a)\ \geq 0; o_{kh} \geq 0,\ \forall a \in \bar{A}; \forall h \in H,\ \forall k \in P \tag{4.16}$$

$$x_{ih}, y_{ih}, c_{kh}, z_{kph} \in \{0,1\},\ \forall h \in H,\ \forall i \in I,\ \forall k,p \in P, k \neq p \tag{4.17}$$

In the objective function, the number of representatives of a particular party is maximized. Each unit must be assigned to a district in the constraints 4.2. The number of representatives of each party from each district is calculated by using the constraints from 4.3 to 4.6. The total number of votes of each party in each district is calculated in 4.3. The constraints with the number 4.4 are utilized to find the number representatives of each party by considering the total number of votes of each party. In order to find the winner party in each district, at first we compare the total votes of a pair of party and keep the result information that can be 0 or 1, then we determine which party dominates the others using the information from the constraints 4.4. This determination part is in 4.5. The mentioned information is kept in the variable o_{ph} $\forall p \in P, \forall h \in H$. In 4.6, the constraints enforce that only one candidate or party can be elected in each district. Calculation of the population of each district is in the constraints 4.7. The constraints of 4.8 and 4.9 are developed for the population equality criterion. The rest of the mathematical model is originated from the aforementioned article. Each source node must be assigned to one of the districts 4.10. The flow amount from each source node to the nodes we have in the original problem must be equal to a predetermined value if the corresponding arc is used 4.11. The total amount of flow on entering arcs must be equal to the total amount of flow on leaving arcs for each node 4.12. In 4.13, if node i is not assigned to unit h, then the total flow on entering arcs to this unit must be zero. If the flow between node w_i^h and tail node t_i is zero, then the node (unit) i cannot be assigned to the district h according to the constraints 4.14. Some predetermined units cannot be assigned to a same district in order to attain more compact results in the constraints 4.15. This constraint can be deleted according to the circumstance. The constraints of 4.16 are for nonnegativity. The last constraints are for defining the domains of the decision variables.

4.2 Mathematical Formulation for the Multi-Member District

The formulation above considers only the single-member district system. Multi-member case is also developed to have a more generic formulation. The single-member district problems can be solved using the formulation given in this section. The difference between two formulations is that the contraints from 4.4 to 4.6 in the first formulation are modified to the constraints that are from 4.22 to 4.26 in the second formulation. Calculating the number of representatives of each party in each district is handled by this group of constraints in both models. Second approach is more sophisticated than the first one. D'Hondt method is utilized for the allocation of the number of representations of each party according to its vote proportion in each district. The reader should check the mentioned allocation method to comprehend the constraints from 4.22 to 4.26 in the second formulation. The additional sets and parameters are provided in Table 4.3. The additional decision variables are introduced in Table 4.4.

Table 4.3: Additional sets and parameters for the second formulation

J	the set of denominators(number of representatives)
M_1	A sufficiently big value for the difference of total votes
M_2	A sufficiently big value for number of representatives and parties

Table 4.4: Additional decision variables for the second formulation

c_{kh}	number of representatives of party k gained in district h; $\forall k \in P, \forall h \in H$
o_{kjh}	auxiliary variable which is derived from the o_{kh} $\forall k \in P, \forall j \in J, \forall h \in H$
e_{kjh}	binary decision variable if party k's denominator j does not have enough vote for a representative in district h; $\forall k \in P,\ \forall j \in J,\ \forall h \in H$
z_{pkjdh}	binary decision variable for the comparison of the votes $\forall k, p \in P,\ \forall j, d \in J,\ \forall h \in H, (k \neq p, d \neq j)$

$$max \sum_{h \in H} c_{\rho h} \tag{4.18}$$

$s.t.$

$$\sum_{h \in H} x_{ih} = 1,\ \forall i \in I \tag{4.19}$$

$$\sum_{i \in I} x_{ih} v_{ki} = o_{ph},\ \forall p \in P,\ \forall h \in H \tag{4.20}$$

$$o_{pjh} = o_{ph}/j,\ \forall p \in P,\ \forall j \in J,\ \forall h \in H \tag{4.21}$$

$$o_{pjh} - o_{kdh} + M_1 z_{pkjdh}\ \geq 0,\ \forall k, p \in P,\ \forall j, d \in J,\ \forall h \in H, (p \neq k, j \neq d) \tag{4.22}$$

$$\sum_{k \in P} \sum_{d \in J} z_{pkjdh} - |J| + 1\ \leq M_2 e_{pjh},\ \forall h \in H,\ \forall p \in P,\ \forall j \in J, (p \neq k, j \neq d) \tag{4.23}$$

$$\sum_{p \in P} \sum_{j \in J} e_{pjh} = |J||P| - |J|,\ \forall h \in H \tag{4.24}$$

$$c_{ph} = |J| - \sum_{j \in J} e_{pjh},\ \forall p \in P,\ \forall h \in H \tag{4.25}$$

$$\sum_{i \in I} x_{ih} s_i = t_h,\ \forall h \in H \tag{4.26}$$

$$t_h\ \leq \bar{s}(1 + \beta),\ \forall h \in H \tag{4.27}$$

$$t_h\ \geq \bar{s}(1 - \beta),\ \forall h \in H \tag{4.28}$$

$$\sum_{i \in I} y_{ih} = 1,\ \forall h \in H \tag{4.29}$$

$$f(s^h, w_i^h) = F y_{ih},\ \forall i \in I,\ \forall h \in H \tag{4.30}$$

$$\sum_{a \in \delta^-(w_i^h)} f(a) = \sum_{a \in \delta^+(w_i^h)} f(a),\ \forall i \in I,\ \forall h \in H \tag{4.31}$$

$$\sum_{a \in \delta^-(w_i^h)} f(a)\ \leq F x_{ih},\ \forall i \in I,\ \forall h \in H \tag{4.32}$$

$$x_{ih}\ \leq f(w_i^h, t_i),\ \forall i \in I,\ \forall h \in H \tag{4.33}$$

$$x_{ih} + x_{jh}\ \leq 1,\ \forall h \in H,\ \forall (i,j) \in C \tag{4.34}$$

$$f(a)\ \geq 0,\ \forall a \in \bar{A} \tag{4.35}$$

$$x_{ih}, y_{ih}, z_{pkjdh}, e_{kjh} \in \{0,1\},\ \forall h \in H,\ \forall i \in I,\ \forall k,p \in P,\ \forall j,d \in J \tag{4.36}$$

In the objective function, the number of representatives of a particular party is maximized. Each unit must be assigned to a district 4.19. The number of representatives of each party in each district is calculated by using the constraints from 4.20 to 4.25. In 4.20 the total number of votes for each party in each district is calculated. The constraints with number 4.21 are for comparison. Only the information of total votes of each party in each district is not sufficient to determine the total number of representatives of each party elected in each district for multi-member district. According to the D'Hondt method, we also need to calculate the half of the total votes of each party, one third of it and so on. At the end, we are going to compare the values as much the multiplication of the number of total representatives can be elected and total number of available parties. In 4.21 we are calculating these values. The working principle of the constraints 4.22 is similar to 4.4 and keeps the comparison information bu using the z variable with 5 indices. Using the constraints 4.23 and 4.24 we are ordering all the values we found from 4.21. Note that, right hand side value of 4.24 is equal to a value that is the difference between the number of alternatives and the number of representatives. In other words, we are trying to determine which values are the smallest ones. Total number of representatives of each party elected in each district is calculated by summing the each party s e columns and subtract this value from the total number of representatives can be elected. The constraints from 4.26 to 4.28 are related to the population equality criterion. The rest of the mathematical model is originated from Nemoto and Hotta (2003) and explained in the previous formulation in detail.

Chapter 5

TABU SEARCH ALGORITHM

The Section 5.1 covers an algorithm that is used to solve the cases which cannot be solved by an exact approach. We used the Tabu Search Heuristic algorithm that is commonly used in the literature for different types of problems such as vehicle routing (Barbarosoglu and Ozgur, 1999) and scheduling (Dell'Amico and Trubian, 1993). The pseudocode of the algorithm is provided in Algorithm 2.

5.1 Tabu Search Algorithm

Algorithm 2 Tabu Search Algorithm

Input: $TabuList_{size}$
Output: S_{best}
$TabuList \leftarrow \emptyset$
$S_{best} \leftarrow InitializationFunction()$
while $\neg StopCondition()$ **do**
    $CandidateList \leftarrow \emptyset$
    for all $S_{Candidate} \in SBest_{neighborhood}$ **do**
        if $\neg ContainsAnyFeatures(S_{Candidate}, TabuList)$ **then**
            $CandidateList \leftarrow S_{Candidate}$
        end if
    end for
    $S_{Candidate} \leftarrow ChooseBestCandidate(CandidateList)$
    if $ObjectiveValue(S_{Candidate}) \geq ObjectiveValue(S_{Best})$ **then**
        $S_{Best} \leftarrow S_{CandidateList}$
        $TabuList \leftarrow FeaturesDifferences(S_{Candidate}, S_{Best})$
        while $size(TabuList) > TabuList_{size}$ **do**
            $DeleteFeature(TabuList)$
        end while
    end if
end while
return S_{Best}

In the algorithm, $StopCondition()$ is the maximum number iteration. The first step in the while loop is determining the candidate solutions using the neighborhood structure. $ContainsAnyFeatures()$ checks whether the candidates are in the tabu list or not. The ones that are in the tabu list are eliminated. One of the most important function in the algorithm is $ChooseBestCandidate$ $(CandidateList)$. The input of this function is the candidate solutions. We need to find the best solution at each iteration. The quality of each solution is calculated using the Pareto optimality structure. The reason to use such an approach is the nature of the problem we have. We are trying to maximize the number of representatives of a particular party and the compactness of each district while satisfying the population equality constraint. Note that if a candidate is not satisfying the contiguity and population equality criteria, then it should be eliminated before checking the tabu list. After all, we only have two performance measures to be maximized; the number of representatives of a party we choose and the shape quality of the districts.

In practice, we realized that the number of representatives of a particular party can be increased steadily in the long run by considering the difference between the total votes of the party in the objective and the total votes of the party which has the greatest proportion of the votes. If these parties are the same, then we are interested in the difference between the votes of that party and the one which has the second most vote percentage. Assume that we have 2 parties; A and B and we are trying to maximize the total number of representatives of the party A. We know that two districts can be changed in each candidate solution. Before a change, assume that A has 3 and 7 total number of votes in two districts which are D1 and D2. On the other hand, B has 10 and 2 in these districts respectively. In this case, A and B are in a tie. A and B win a representative in D2 and D1 respectively. After the change, the numbers are updated as 5(D1)-5(D2) for A and 8(D1)-4(D2) for B. They still have a tie; however, something important changed. If we are interested in A, then we are trying to increase its total votes in each district while decreasing the total votes of other parties. The problem is the total number of votes of the parties in all districts are limited. We need to find a new approach to use our votes efficiently. In the most desired case, the total votes of A should not be extremely bigger than the others. The case 5(A)-4(B) is more desired than 7(A)-2(B) according to this reasoning.

The function $ChooseBestCandidate$ $(CandidateList)$ has a prioritization structure. At first, we are eliminating the ones that have the smaller number of representatives than the current solution. After this step, if no candidate is available, the algorithm would choose two new adjacent districts to be changed. If the number of candidate solutions is more than one , we need to check the compactness scores and the *difference* scores of each candidate by using the *Pareto Optimality* notion. In this step, we are trying to find a set of candidate solutions which dominate the rest strictly. For example, a candidate with (0.5,3) scores is strictly dominated by a solution with (0.6,1) scores. The first value indicates the compactness value and the second one is for the *difference* value. Note that we are aiming to increase the compactness value and to decrease the *difference* value of each district. We can still have more than one candidate solution after using the Pareto Optimality. This can be seen as a disadvantage in the algorithm; however, we used this case to change the primary objective of the algorithm. For example, we can choose the most compact solution of the Pareto set if we need more compact results at the end. On the other hand, we can choose the one that has the least *difference* value if the total number of representatives is more important than the compactness.

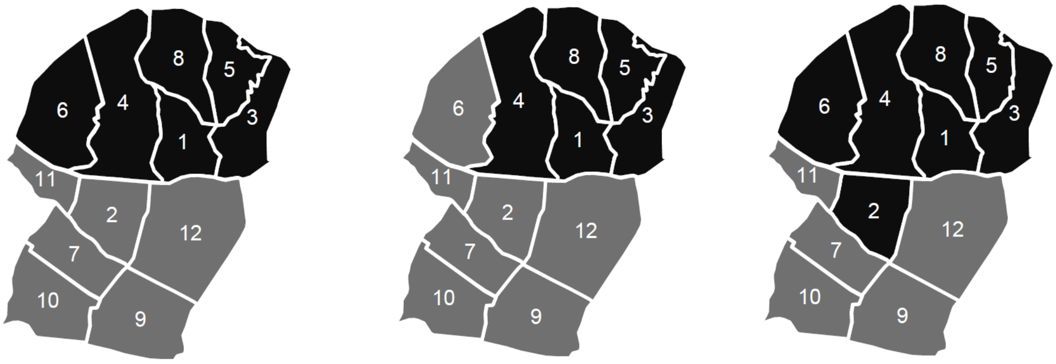
5.1.1 Initialization Function

We used a three step initialization approach to have the required initial solutions efficiently. At the first step, a random unit is chosen as the seed of the district and extended this district by adjoining to seed one of its adjacent units. The district is complete whenever its population attains the average population value or when no adjacent units are available. There are three possible cases after assigning all the existing units. The total number of districts can be lower than the desired value, equal to the value or greater than the value. If the number of districts is lower than the desired value, the algorithm gradually increases it by iteratively splitting the most populated district into two, while preserving contiguity. If the number of districts is greater than the desired value, which happens almost all the time, the algorithm reduces it by iteratively merging the least populated district with its least populated neighbor.

After these two steps, one more step is added to the initialization procedure to have feasible solutions. Population equality constraint cannot be achieved after the first two steps. In order to satisfy this equality, a unit which is in a highly- populated district is assigned to a district which is neighbor and least populated. We are trying to minimize the variance of population values of all districts by using the third initialization step.

5.1.2 Neighborhood Structure

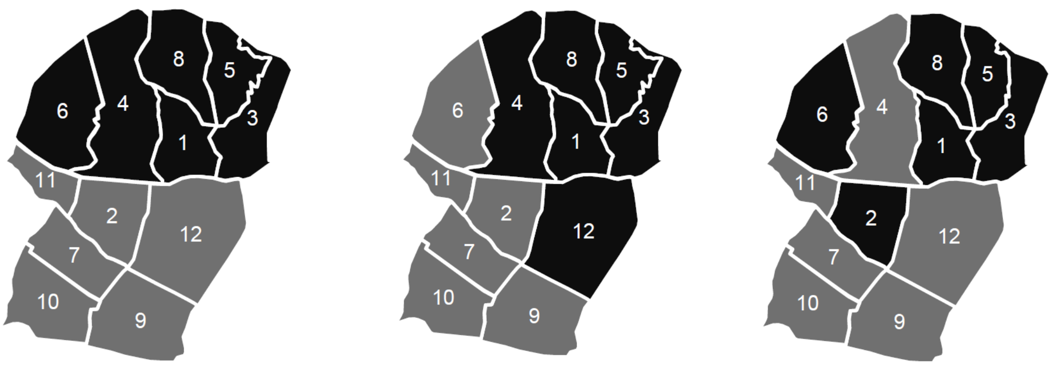
Two different neighborhood structures are utilized in the algorithm. All changes are between two neighboor districts. In the first one, one of the adjacent units between two districts is assigned to the neighbor district. The structure is shown in Figure 5.1. There are two adjacent districts that are colorized by black and gray. In the first example, the unit number 6 is assigned to the gray district. Unit number 2 is assigned to the black district in the second example.



(a) Initial Position (b) First example (c) Second example

Figure 5.1: Neighborhood change assignment based

Another neighborhood structure is swap-based. We are still interested in the adjacent units between two districts. These are unit number 1, 3, 4, 6 and 2, 11, 12 as it has been mentioned in the previous move and shown in Figure 5.2. One of each unit from two districts is assigned to the opposite district. For example unit number 6 and 12 are swapped in the second picture of Figure 5.2. As might be expected, some of changes are infeasible solutions. An example of the infeasible changes is shown in the last picture. In the algorithm, a contiguity control approach is used to eliminate these infeasible moves. Note that, the gray district is still contiguous. Both of the districts must be contiguous to accept a candidate solution for the next step in the algorithm.



(a) Initial Position (b) An example of feasible change (c) An example of infeasible change

Figure 5.2: Neighborhood change swap based

Chapter 6

COMPUTATIONAL RESULTS

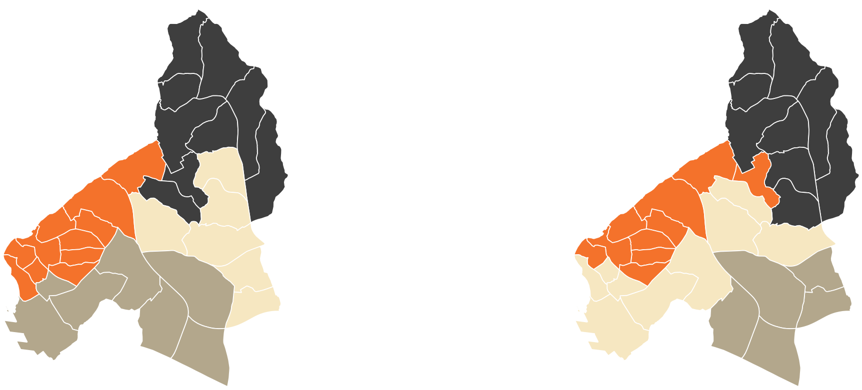
The required election data to test our formulations and the algorithm is downloaded from the Turkish Statistical Institute's website. Some previous elections' data are downloaded such as June 2015, November 2015 elections and April 2017 referendum. In the datasets of elections, we limited the number of parties as 5. The name of the parties are as follows.

- Justice and Development Party
- Nationalist Movement Party
- People's Democratic Party
- Republican People's Party
- Other Parties

In the referendum dataset we have only two options, total number of people that voted as *Yes* and voted as *No*. The computational experiments presented in this chapter are conducted on a workstation with Intel i5-5200U CPU @ 2.2 GHz and 4GB RAM running on Windows 10 operating system. The solution of the mathematical models are provided by using the IBM ILOG CPLEX Optimization Studio. R programming language with a software version 3.2.3 is used to implement the tabu search algorithm and for visualization of the results.

6.1 Mathematical Formulation Results

November 2015 election results are used in order to test the first mathematical model. There are 39 counties in İstanbul and each of them has different number of neighborhoods. A specific party, in general JDP or RPP, dominates the rest in most of the counties. In other words, a party has the largest proportion of the votes in all of the neighborhoods in a county. Such cases are not considered. We also eliminate the counties with just one representative since we cannot divide this county into more than one district. In Figure 6.1, two districting plans are provided. In the first districting plan, JDP and RPP win 2 representatives each considering the November 2015 elections. On the other hand, JDP wins 4 representatives and number of representative of RPP is 0 in the second plan. The compactness criterion is not used in the mathematical model. However, the results seem compact and can be applied easily. Another metrics are provided in Table 6.1. The population of the districts are close to each other in the table. Note that, the standart deviation of the population values of the last row is larger than the values which are in the first and second row since we use a higher β value in that run.



(a) Results when RPP in the objective (b) Results when JDP in the objective

Figure 6.1: Model results for Üsküdar

Table 6.1: The results of several runs for Üsküdar

Number of units	Number of Districts	β	The party	Obj. Value	pop for D1	pop for D2	pop for D3	pop for D4
33	4	0.1	JDP	4	85634	93315	90761	85588
33	4	0.1	RPP	2	93261	91717	89082	81238
33	4	0.2	RPP	2	93055	89243	101393	71607

6.1.1 Results for the Counties in Asia

There are 14 counties including Adalar in the Asia side of İstanbul. The number of neighborhoods and representatives of these counties can differ. The results of several runs are provided in Table 6.2. The last column is a measure for the population equality criterion. Note that each of these values is smaller than the corresponding β value. The measurement is the maximum deviation percentage from the average population value of the considered county. Üsküdar belongs to this table; however, we show its values in a different table for pointing out.

Table 6.2: The results of several runs for the Asian Counties

County Name	Number of Units	Number of Districts	β	The party	Objective Value	pop. equality measure
Ataşehir	17	2	0.1	JDP	2	0.066
Ataşehir	17	2	0.2	JDP	2	0.066
Ataşehir	17	2	0.1	RPP	1	0.005
Ataşehir	17	2	0.2	RPP	1	0.005
Ataşehir	17	2	0.3	RPP	1	0.206
Ataşehir	17	3	0.1	JDP	3	0.04
Ataşehir	17	3	0.1	RPP	1	0.018
Ataşehir	17	3	0.25	RPP	2	0.208
Çekmeköy	21	2	0.1	JDP	2	0.084
Kadıköy	21	3	0.1	RPP	3	0.086
Kartal	20	3	0.1	JDP	3	0.088
Kartal	20	3	0.4	RPP	1	0.391
Maltepe	18	3	0.1	JDP	2	0.072
Maltepe	18	3	0.15	RPP	2	0.107
Maltepe	18	3	0.2	RPP	3	0.187
Sancaktepe	19	3	0.1	JDP	3	0.097
Sancaktepe	19	3	0.1	RPP	1	0.075
Sultanbeyli	15	3	0.05	JDP	3	0.017
Ümraniye	35	5	0.05	JDP	5	0.044

6.1.2 Results for the Counties in Europe

There are 25 counties in the European side of the İstanbul. A party dominates the other parties in most of the counties. The results are available for 6 counties. We used a relatively high β value for some of the runs such as the first row in the table Table 6.3. In the last 3 rows, the reader can check the results of Sarıyer county for its interesting results. The objective function can be changed depending on the party in the objective while satisfying the population equality contraint with a relatively small β.

Table 6.3: The results of several runs for the European Counties

County Name	Number of Units	Number of Districts	β	The party	Objective Value	pop. equality measure
Avcılar	10	3	0.1	JDP	2	0.081
Avcılar	10	3	0.25	RPP	2	0.242
B.evler	11	4	0.15	JDP	4	0.142
B.evler	11	4	0.15	RPP	1	0.144
Basakşehir	10	3	0.1	JDP	3	0.035
Basakşehir	10	3	0.1	RPP	1	0.082
Fatih	57	4	0.2	RPP	1	0.174
Fatih	57	3	0.25	RPP	1	0.149
Kçekmece	21	5	0.2	JDP	5	0.168
Kçekmece	21	5	0.2	RPP	1	0.169
Sarıyer	38	2	0.05	JDP	2	0.016
Sarıyer	38	2	0.1	JDP	2	0.061
Sarıyer	38	2	0.05	RPP	1	0.031
Sarıyer	38	2	0.1	RPP	1	0.064

6.1.3 Tabu Search Algorithm Results

What would be the results if we ignore the current boundries of the counties? In such case we have 359 counties in Asia side and the problem cannot be solved using the mathematical formulation. In this part, we tried to divide the Asia into 31 districts that is the total number of representatives can be elected in all counties located in Asia. We have three alternative datasets to utilize in the algorithm. We choose the referendum dataset since the algorithm could not be used with November 2015 election dataset. Population values are extremely high in some regions such as Kadıköy, Üsküdar comparing with Şile. This case caused some problems in the initialization step and later steps. Some approaches can be used to prevent such problems; ignoring the counties that are populated low or changing the population values in an attempt to have almost equally populated neighborhoods. Second approach is used in this thesis.

In Figure 6.2, an initial solution is visualized on a map. The average compactness value of the districts in this plan is 0.3038. The compactness measure we take into consideration is convex-hull approach. When the algorithm is started for 2000 iterations, the result in Figure 6.3 has been achieved. The number of districts is increased from 12 to 17 for *No* objective. The average compactness score of the new districting plan is 0.3602. Improvement in the compactness can be detected visually in the maps we provided. 0.15 is used for β value in the algorithm. The elapsed time is approximately 72 minutes. Note that, after about 1200 iterations, we could not achive any improvement on the number of districts in which *No* alternative dominates the *Yes*. The size of the tabu list is set to 20, 50 and 100, however, we have almost identical solutions at the end. In this solution, tabu list size 50 is used.

We also set the *Yes* option to the objective and achieved some succesfull results. The objective is increased from 16 to 20 in this setting and the average compactness is increased from 0.3911 to 0.4120. The reason to have an initial solution with a high average compact score is that we use some of the results for initialization. In Figure 6.4, the initial solution for *Yes* objective can be seen and the results of the algorithm is shown in Figure 6.5. Same parameter setting is used for this run.

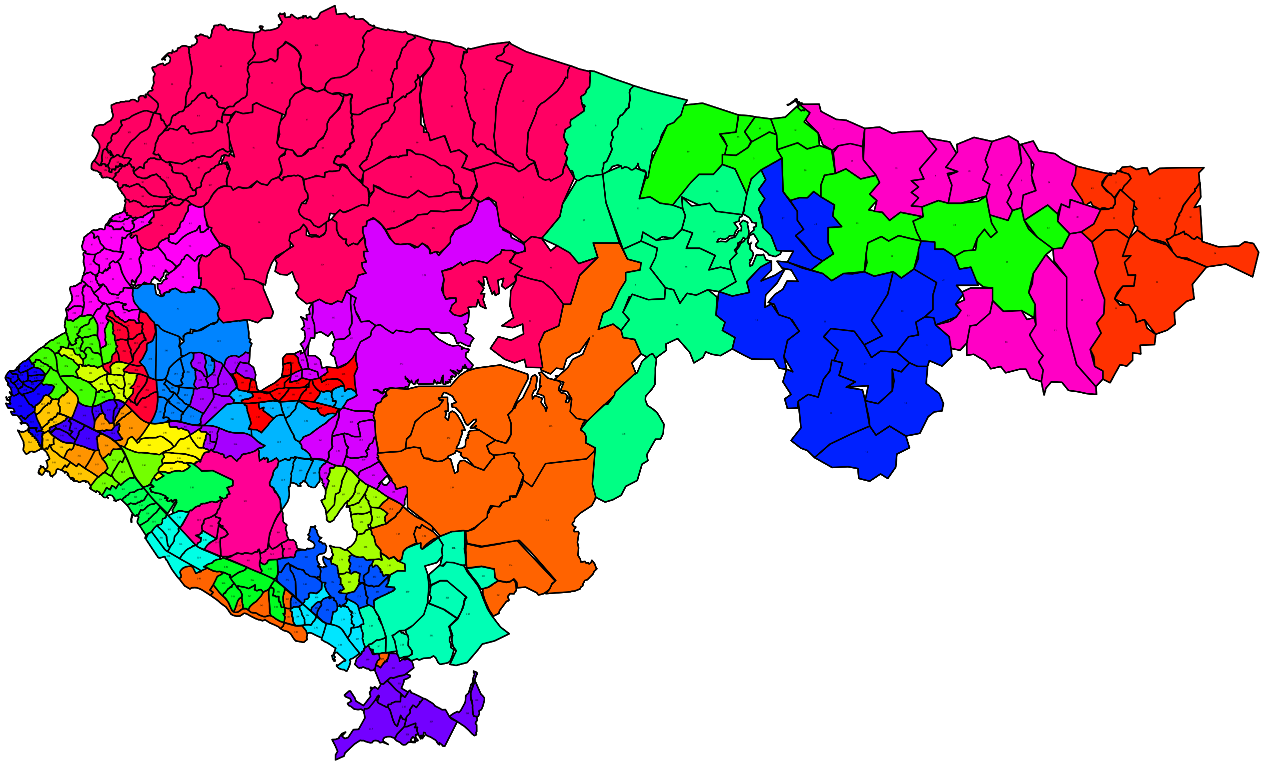


Figure 6.2: An initial solution with 12 *No* and 19 *Yes*

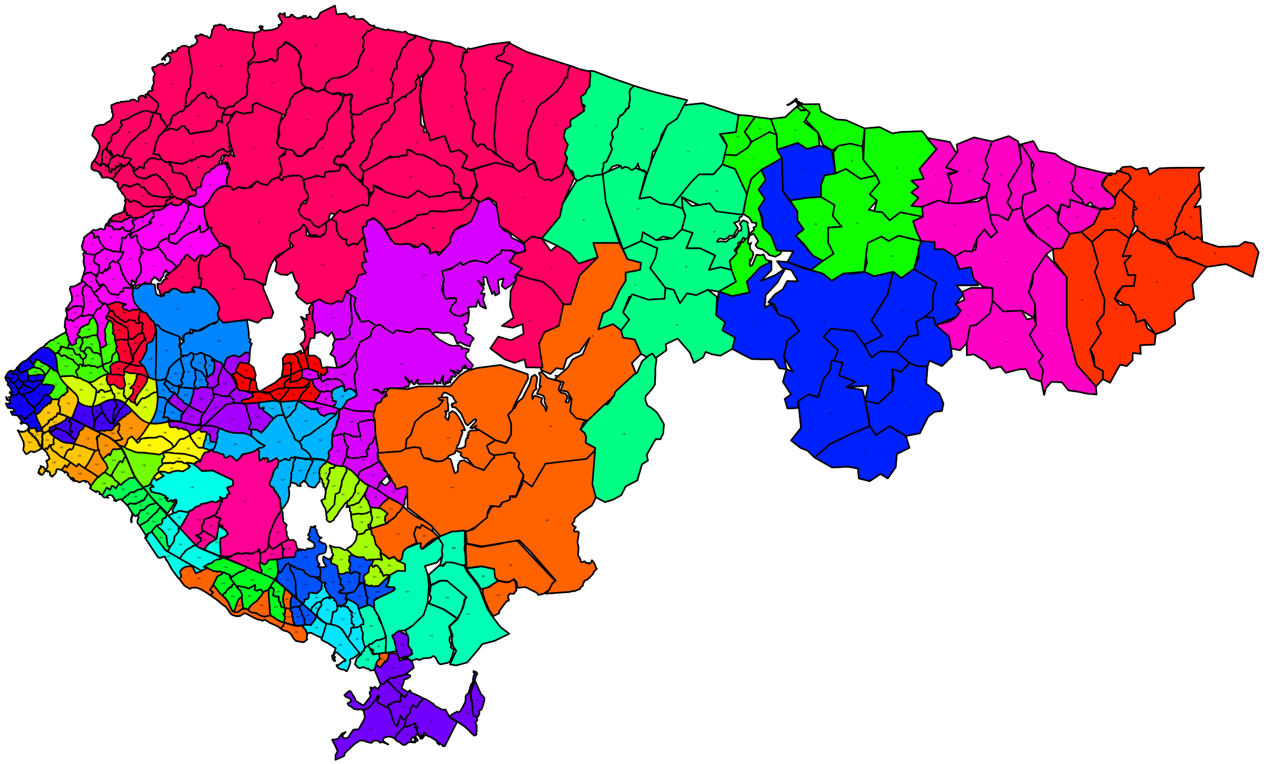


Figure 6.3: A result solution with 17 *no* objective and 14 *Yes*

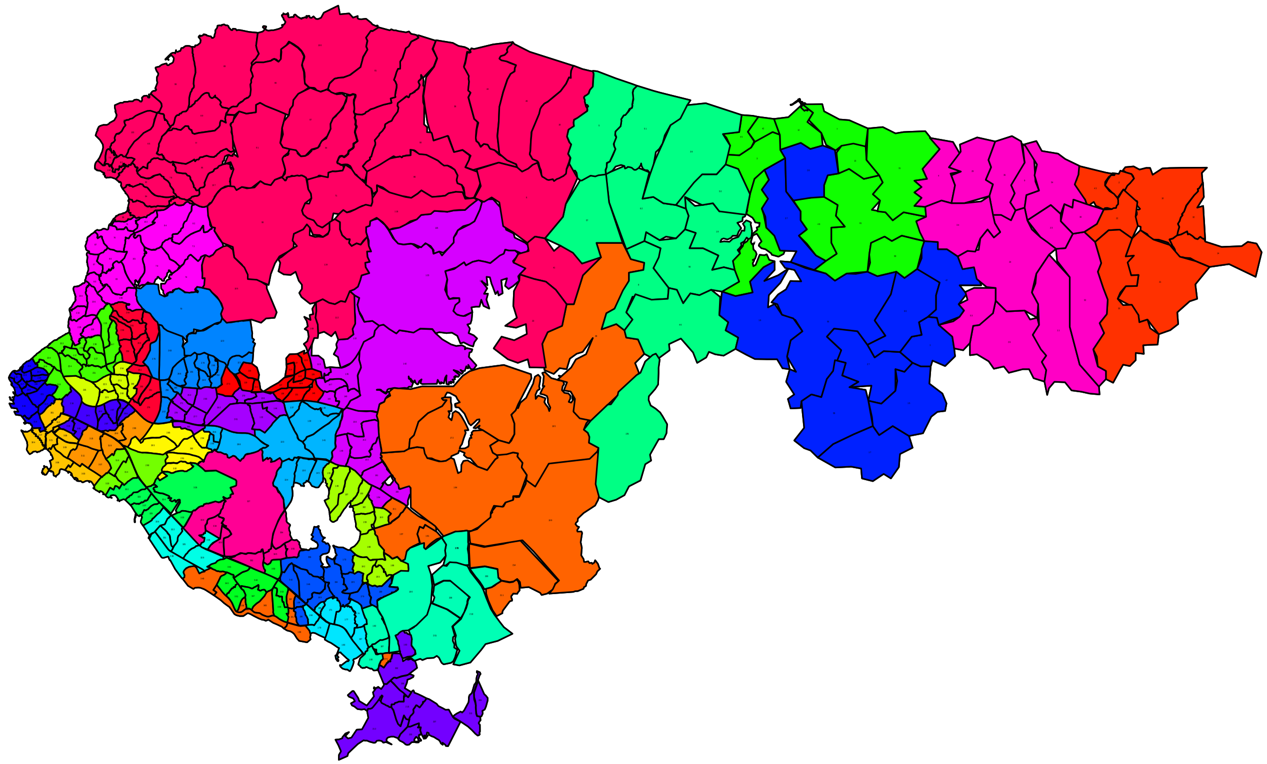


Figure 6.4: An initial solution with 16 *Yes* and 15 *No*

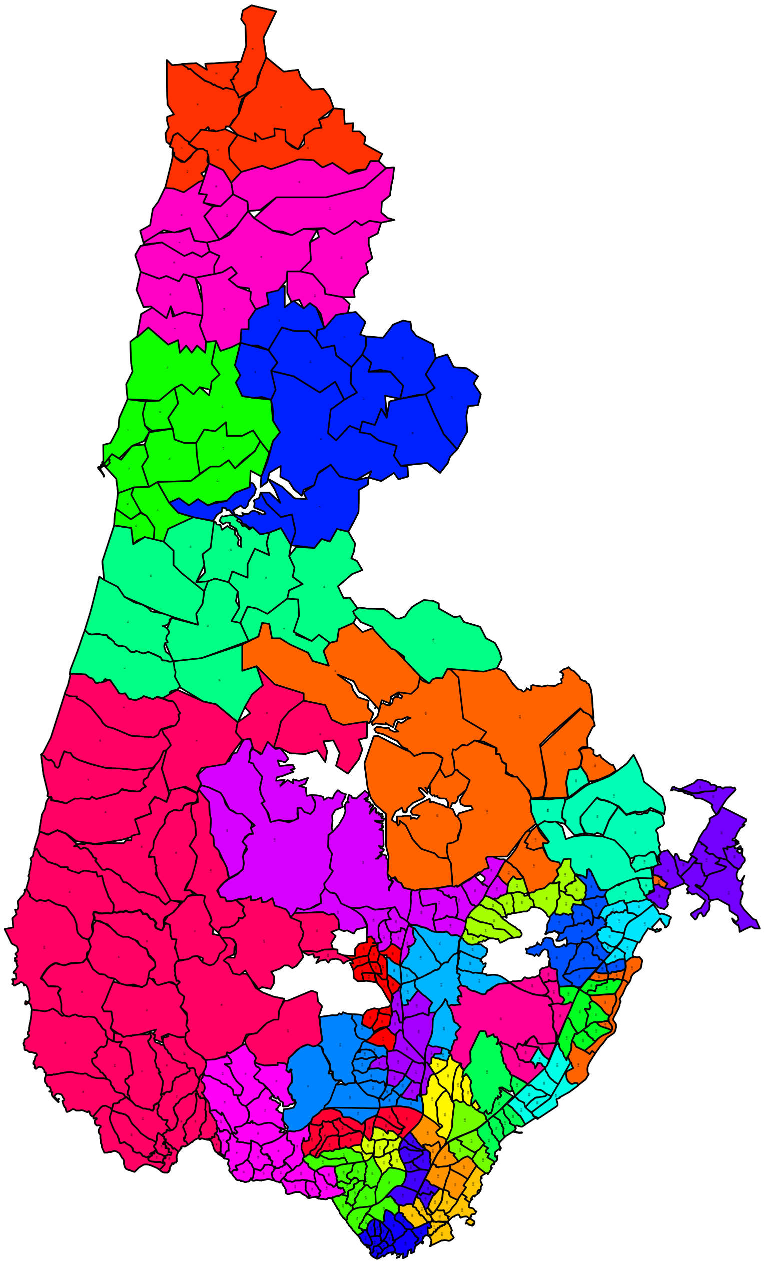


Figure 6.5: A result solution with 20 *Yes* and 11 *No*

Chapter 7

CONCLUSION

The main motivation to study the political districting problem for a master thesis is that topical issues and their possible reflections on the future. Manipulation of the political districts in favor of a particular party and unfair representation are some of the issues we aim to show in this thesis. Unfortunately, we could not provide a plan for how a redistricting should be made, in other words, we could not present a districting plan which is represented all the people in a district fairly. Since we have limited time and would like to emphasize on some other related topics in the thesis, the fair representation could not be studied.

Gathering the required election, referendum data and preprocessing of them was time-consuming. In addition to the election and referendum data, the spatial data for visualization is obtained from several sources partially. Almost half of the spatial data is constituted manually for visual representation of the results.

Several exact approaches are tested to obtain feasible solutions. The model in Nemoto and Hotta (2003) provides the satisfying results in terms of the objective function value and the elapsed time. The calculation of the number of representatives part of the formulation is combined with the original model.. In addition to this model, another calculation approach is developed for multi-member district problems.

The results of the mathematical models demonstrate that territory can be partitioned in a way that a party can gain a political advantage. In Üsküdar scenario, we verify this claim by showing the feasible alternatives with the different number of representatives. The number of these alternatives can be extended simply by using the mathematical model we proposed. One of the drawbacks of the formulation is that the compactness criterion is not used. However, an approach can be developed to also use this criterion such as not allowing to assign two units which are far from each other into the same district. After the results we have by using the mathematical formulation without the compactness criterion, we realized that we do not need such approach for the cases we are interested in.

For further studies, the mathematical model can be developed to avoid the prespecified population equality parameter. One approach is to have a multilinear objective function which covers all the criteria we mentioned including the total number of representatives of a particular party. A problem arising from this development can be determining the coefficients of the criteria in the objective. These values can be tuned according to the results of the formulation or changed depending on the needs of the decision maker.

Bibliography

Altman, M. and McDonald, M. P. (2011). Bard: Better automated redistricting. *Journal of Statistical Software. Forthcoming, URL http://www. jstatsoft. org*, 42(4):1–28.

Bacao, F., Lobo, V., and Painho, M. (2005). Applying genetic algorithms to zone design. *Soft Computing*, 9(5):341–348.

Barbarosoglu, G. and Ozgur, D. (1999). A tabu search algorithm for the vehicle routing problem. *Computers & Operations Research*, 26(3):255–270.

Bozkaya, B. (1999). Political districting: A tabu search algorithm and geographical interfaces.

Britannica, E. (2019). Parliamentary democracy. `https://www.britannica.com/topic/parliamentary-democracy`.

Browdy, M. H. (1990). Simulated annealing: an improved computer model for political redistricting. *Yale Law & Policy Review*, 8(1):163–179.

Corcoran, C. and Saxe, K. (2014). Redistricting and district compactness. *The Mathematics of Decisions, Elections, and Games, Providence: American Mathematical Society*, pages 1–16.

Dell'Amico, M. and Trubian, M. (1993). Applying tabu search to the job-shop scheduling problem. *Annals of Operations research*, 41(3):231–252.

Dictionary (2019). Compactness. `https://www.dictionary.com/browse/compactness`.

Forman, S. L. and Yue, Y. (2003). Congressional districting using a tsp-based genetic algorithm. pages 2072–2083.

Garfinkel, R. S. and Nemhauser, G. L. (1970). Optimal political districting by implicit enumeration techniques. *Management Science*, 16(8):B–495.

Grofman, B. (1985). Criteria for districting: A social science perspective. *UCLA L. Rev.*, 33:77.

Hartigan, J. A. and Wong, M. A. (1979). Algorithm as 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1):100–108.

Hess, S. W., Weaver, J., Siegfeldt, H., Whelan, J., and Zitlau, P. (1965). Nonpartisan political redistricting by computer. *Operations Research*, 13(6):998–1006.

Liu, Y. Y., Cho, W. K. T., and Wang, S. (2016). Pear: a massively parallel evolutionary computation approach for political redistricting optimization and analysis. *Swarm and Evolutionary Computation*, 30:78–92.

Nemoto, T. and Hotta, K. (2003). Modelling and solution of the problem of optimal electoral districting. *Communications of the OR Society of Japan*, 48:300–306.

Niemi, R. G., Grofman, B., Carlucci, C., and Hofeller, T. (1990). Measuring compactness and the role of a compactness standard in a test for partisan and racial gerrymandering. *The Journal of Politics*, 52(4):1155–1181.

Nygreen, B. (1988). European assembly constituencies for wales-comparing of methods for solving a political districting problem. *Mathematical Programming*, 42(1-3):159–169.

Polsby, D. D. and Popper, R. D. (1991). The third criterion: Compactness as a procedural safeguard against partisan gerrymandering. *Yale Law & Policy Review*, 9(2):301–353.

Ricca, F. and Simeone, B. (2008). Local search algorithms for political districting. *European Journal of Operational Research*, 189(3):1409–1426.

Rincón-García, E.-A., Gutiérrez-Andrade, M.-Á., de-los Cobos-Silva, S.-G., Lara-Velázquez, P., Mora-Gutiérrez, R.-A., and Ponsich, A. (2015). Abc, a viable algorithm for the political districting problem. pages 269–278.

Schwartzberg, J. E. (1965). Reapportionment, gerrymanders, and the notion of compactness. *Minn. L. Rev.*, 50:443.

Vanneschi, L., Henriques, R., and Castelli, M. (2017). Multi-objective genetic algorithm with variable neighbourhood search for the electoral redistricting problem. *Swarm and Evolutionary Computation*, 36:37–51.

Vickrey, W. (1961). On the prevention of gerrymandering. *Political Science Quarterly*, 76(1):105–110.

Young, H. P. (1988). Measuring the compactness of legislative districts. *Legislative Studies Quarterly*, pages 105–115.

Bozkaya, B. and Erkut, E. and Laporte, G. (2003). A tabu search heuristic and adaptive memory procedure for political districting.*European Journal of Operational Research*, 144(1): 12-26

Chapter 8

Appendix

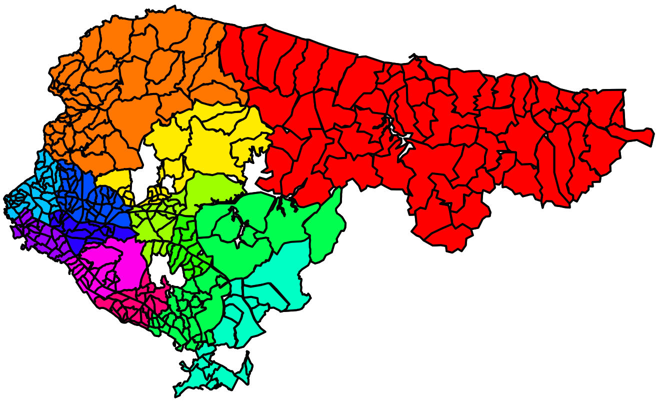


Figure 8.1: Asia map that shows the current counties

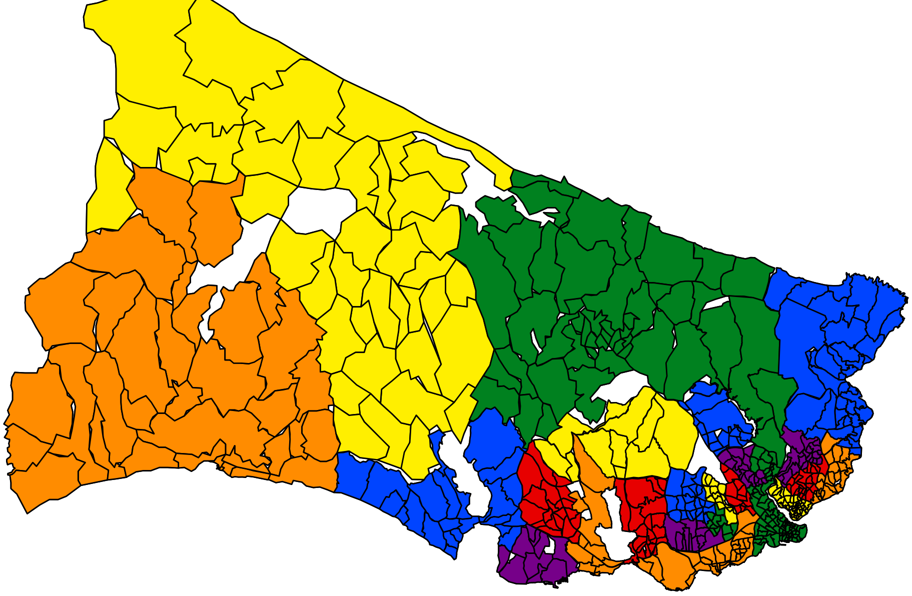


Figure 8.2: Europe map that shows the current counties