Computers & Operations Research 146 (2022) 105892

Contents lists available at ScienceDirect

Computers and Operations Research

journal homepage: www.elsevier.com/locate/cor

Constraint-based electoral districting using a new compactness measure: An application to Portugal

Tiago Almeida, Vasco Manquinho *

INESC-ID, Instituto Superior Técnico, Universidade de Lisboa, Lisboa, Portugal

ARTICLE INFO

Keywords:
Electoral districting
Gerrymandering
Multi-objective combinatorial optimization
Compactness

ABSTRACT

Since the 1960s, several approaches have been proposed to prevent the redrawing of electoral districts so that they are beneficial to a particular party or political faction (gerrymandering). In order to avoid gerrymandering, most approaches focus solely on redrawing electoral maps that try to maximize the compactness of the electoral districts. However, this problem is computationally hard, and several criteria must be satisfied when drawing electoral maps.

In this work we focus on drawing electoral districts for Portugal. For that, a new compact Boolean formulation is proposed for solving the electoral districting problem. This formulation satisfies all standard criteria for building electoral maps, particularly the contiguity and representation criteria. Moreover, a new compactness measure is also proposed that does not depend on geographic centers. Additionally, a heuristic formulation is also devised for problem instances where the optimum values are hard to find.

The new compactness measure and formulations are applied in the context of Portuguese parliament elections, assuming a change in the Portuguese electoral system. Experimental results on drawing electoral maps show that the proposed formulations are more effective than previous ones in drawing the electoral districts in the Portuguese context. Moreover, based on results from the last elections, gerrymandering scenarios are devised, showing that electoral outcomes can change depending on the drawing of the electoral maps.

1. Introduction

Most European democracies elect their members for the legislative body through a closed-list party system. In this case, the elected members of each party are defined through a proportional representation (PR) system (such as the D'Hondt method D'Hondt, 1882) that makes the distribution according to a predefined ordered list. Hence, in these voting systems, people do not vote on a particular candidate but in a party-list from which a given number of people are to be elected according to the number of votes the party received.

A different electoral system is one based on single-member electoral districts. In this system, the territory is split into a given number of electoral districts (e.g., the number of members to be elected), and each electoral district elects just one person. In this electoral system, the design of the electoral districts can significantly impact the election outcome. In particular, one can define a map of electoral districts that favors a given political party or coalition.

There are several works to build electoral district maps based on clustering (Bodin, 2006; Bottman et al., 2007; Svec et al., 2007; Ricca et al., 2008; Guest et al., 2019), local search (Vickrey, 1961; Hayes, 1996) or metaheuristics (Browdy, 1990; D'Amico et al., 2002; Macmillan, 2001; Bação et al., 2005; Joshi et al., 2012). The focus of these works is mainly on maximizing the compactness of the electoral districts by optimizing the distance to the geographic center of the electoral district. However, this compactness metric might still produce electoral districts with *odd* shapes. Observe that the notion of odd shape is not precisely defined but is usually associated with a highly irregular and unbalanced shape such as a long and narrow area with many sides.

Integer Linear Programming (ILP) models have also been proposed for electoral districting (Garfinkel and Nemhauser, 1970; Hess et al., 1965; Validi et al., 2020) that try to maximize compactness. However, a common difficulty of these ILP models is to ensure the contiguity of the generated electoral districts. In this paper, we focus on drawing electoral districts for the Portuguese context by extending previous work on ILP models for electoral districting in several ways: (i) new and more compact models to ensure the satisfaction of contiguity constraints; (ii) new measure of compactness based on the size of the frontiers between electoral districts; (iii) new symmetry breaking techniques that improve upon previously proposed models, as well as our models; (iv) new multi-objective combinatorial optimization (MOCO) models to generate

* Correspondence to: INESC-ID, Rua Alves Redol, 9, 1000–029 Lisboa, Portugal.
E-mail addresses: tiagomcalmeida@tecnico.ulisboa.pt (T. Almeida), vasco.manquinho@tecnico.ulisboa.pt (V. Manquinho).

https://doi.org/10.1016/j.cor.2022.105892
Received 27 January 2021; Received in revised form 13 April 2022; Accepted 22 May 2022
Available online 29 May 2022

possible gerrymandering scenarios; (v) extensive experimental results carried out using real data from continental Portugal.

The paper is organized as follows. First, Section 2 defines the electoral districting problem to be solved, and a brief review on previous and related work to this problem. Next, Section 3 proposes new constraint-based formulations and techniques to solve the electoral districting problem. Extensive experimental results on real data from Portugal are presented in Section 4 where we evaluate our work in different scenarios. Finally, a summary of our contributions and future work in Section 5 concludes the paper.

2. Background

This section reviews several voting systems, defines the electoral districting problem, as well as the related issue of gerrymandering. Next, ILP-based exact methods for electoral districting are briefly described, as we refer to the literature for a more complete survey (Ricca et al., 2013). Finally, related research work on heuristic methods and issues related to electoral districting conclude this section.

2.1. Electoral districting and voting systems

An electoral district is a territorial subdivision of a country (does not necessarily have to match existing administrative subdivisions) that elects members to that country's legislative body.[1] The number of elected members in each district can vary depending on the electoral system in place.

The electoral system is the set of rules that determine when the elections occur, who can vote, and the voting system (how candidates are elected). There are many voting systems, but the most used are party-list proportional representation (e.g., the Portuguese and Spanish systems for legislative elections), first-past-the-post voting (also known as *winner takes all*, used in several elections in the United States of America), the two-round system (standard in head of state elections such as the French or Portuguese presidential elections) and ranked voting where voters rank several candidates by preference (e.g., used to elect members to the legislature of Australia or the president of India) (Cox, 1997; Farrell, 2011).

A single-member district is an electoral district where only one candidate is elected to a body with multiple members (e.g., a legislature). The voting system usually used in these districts is the first-past-the-post (FPTP), where the person with the most votes wins the district even if she did not achieve a majority (50% of the votes). Although less common, instant-runoff voting (IRV) (e.g., Australia) and the two-round system (e.g., France) are also a possibility, and these systems guarantee a majority to the winner. One argument against single-member districts is that they tend to favor two-party systems (Duverger's law), Cox (1997) and Riker (1982) resulting in fewer minority parties in parliament. In order to introduce some proportional representation, some countries use parallel voting (e.g., Italy) that combines single-member districts with party-list proportional representation.

In order to generate new electoral maps, electoral districts must be created by joining territorial units (indivisible) to form clusters. There are multiple criteria to evaluate the quality of the new map, but the core ones are:

1. Each elected member should represent approximately the same number of people. Ideally, each elected member would represent the theoretical best value B of people where B equals the total number of electors[2] divided by the number of officials to be elected.
2. All the new electoral districts should be contiguous. An electoral district ED is contiguous if and only if to go from any point A_1 inside ED to any other point A_2 inside ED as well, it is not necessary to leave ED.[3]
3. All districts should be compact, meaning that odd shapes (see examples in Fig. 1) should not exist.[4] To measure compactness, multiple ideas have been proposed, such as the Schwartzberg score (Schwartzberg, 1966), the Polsby–Popper score (Polsby and Popper, 1991), the Reock score (Reock, 1961) or simply summing the Euclidean distances between the geographical centers (centroids) of each territorial unit inside a district (Niemi et al., 1990; Young, 1988).

Although less crucial, there are other criteria (sometimes used in political districting) such as the conformity to administrative boundaries, that is, the respect of existing administrative subdivisions (used, as much as possible, in the United Kingdom) and the respect of natural boundaries in cases where mountains or rivers may be a problem for the contiguity of the districts (Ricca and Simeone, 1997). Finally, the design of political districts can also be constrained by specific laws such as the Voting Rights Act in the United States of America. However, the computational modeling of these criteria is very hard. Hence, one can check the compliance of the computational solutions with these specific laws afterward.

The odd shape criterion in electoral districts is often associated with gerrymandering. The term gerrymandering first appeared in 1812 associated with electoral districting in the Boston area and is often defined as the practice of redrawing the boundaries of an electoral district to make it more beneficial to a particular party or political faction. Although less usual, gerrymandering can also be used to increase/decrease the voting power of a racial minority (racial gerrymandering) (Lublin, 1999). As a result of gerrymandering, the electoral districts often end up with odd shapes. Fig. 1 illustrates the practice of gerrymandering. In particular, the map of districts 1 and 12 was struck down in 2017 by the United States Supreme Court, and a new electoral map had to be produced.

2.2. Exact methods

Considering that the electoral districting problem is NP-Complete (Altman, 1997), all exact methods are exponential in the worst case, unless $P = NP$. Garfinkel and Nemhauser (1970) were the first to use an exact approach for the political districting problem. They use enumeration techniques, and the algorithm is divided into two stages. The first stage constructs all sets of feasible districts (contiguous districts that meet the population requirements). In the second stage, they solve a set cover problem to choose the solutions that minimize the population deviation with respect to the ideal value.

The first Integer Linear Programming (ILP) formulation was proposed by Hess et al. (1965). Let n and k denote respectively the total number of territorial units and the number of electoral districts to be drawn. The goal is to identify k territorial units as the centers of the k electoral districts, and each territorial unit must be assigned to exactly one district center. The model has n^2 binary variables $x_{i,j}$ where $i, j \in \{1, \dots, n\}$ and $x_{i,j} = 1$ if territorial unit i is assigned to district with center at territorial unit j, and 0 otherwise. A variable $x_{j,j}$ equals 1 if and only if territorial unit j is chosen as the center of a district. Let $d_{i,j}$ denote the distance between units i and j and let V_i denote the population in territorial unit i, where $i, j \in \{1, \dots, n\}$. Finally, let L and U define the lower and upper bound on the population allowed in

[1] There can also be (possibly different) electoral districts within administrative subdivisions of a country to elect local legislatures (e.g., state elections in the United States of America).

[2] Alternatively, the total population can be considered.

[3] It is worth mentioning that contiguity can be hard to determine in the presence of bodies of water or if one wants to maintain boundaries of administrative districts with exclaves.

[4] Odd shapes in electoral districting are usually associated with gerrymandering by the public.

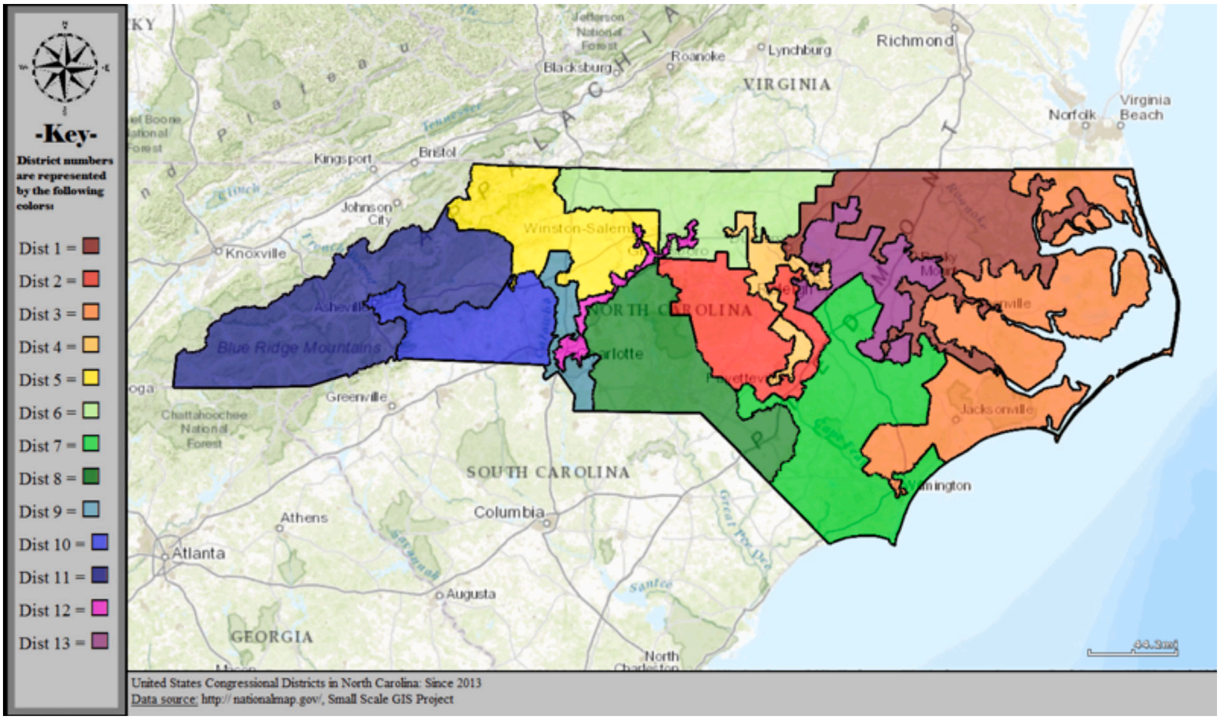

Fig. 1. North Carolina Congressional Districts in 2013.
Source: Extracted from Lewis et al. (2014).

each electoral district. Hence, the ILP model can be built as shown in Formulation 1.

The objective function in the Hess model (1) considers the Euclidean distance between the centers of the territorial units and the population in each unit. It penalizes adding territorial units with large populations to a district far away from the center. The constraints are as follows: (i) a territorial unit must be assigned to only one district center (2), (ii) there must be exactly k district centers (3), (iii) the population limits inside each electoral district must be between the lower and upper bounds (4). Although not in the original model, some adaptations (Oehrlein and Haunert, 2017; Ricca et al., 2013; Validi et al., 2020) add that $x_{i,j} \leq x_{j,j}$, $\forall i, j \in \{1 \dots n\}$, meaning that if a territorial unit i is assigned to a district centered at j, then j must be the center of a district.

A key feature of electoral districting is that each electoral district should be contiguous. Hence, several approaches use the notion of a contiguity graph to represent the neighboring relation between each territorial unit. A contiguity graph is a connected weighted graph $G = (V, E)$ where each vertex corresponds to a territorial unit considered in the districting (e.g., census blocks in the United States of America). An edge between two vertices exists if and only if those territorial units have a common frontier (i.e., neighbor territorial units). For each pair of neighbor territorial units u and v, there is an edge $(u, v) \in E$ with an associated weight that usually represents the (geographical) distance between the center of each geographic unit (either calculated through Euclidean geometry or road distance). In some cases, a weight is also associated with each vertex representing the population living inside the corresponding territorial unit.

Observe that the model in Formulation 1 does not guarantee contiguity of the electoral districts. While trying to optimize the compactness of the electoral districts (Mehrotra et al., 1998; Hess et al., 1965) it might produce some contiguous districts, but this is not guaranteed for all. Therefore, a post-processing algorithm (e.g., local search) must be applied to try to fix this issue.

Defining a model that satisfies the contiguity restriction on the electoral districts is not trivial. However, different techniques have

$$\min \sum_{i=1}^{n} \sum_{j=1}^{n} d_{i,j}^2 \cdot V_i \cdot x_{i,j} \tag{1}$$

$$\sum_{j=1}^{n} x_{i,j} = 1 \qquad \forall i \in \{1 \dots n\} \tag{2}$$

$$\sum_{j=1}^{n} x_{j,j} = k \tag{3}$$

$$L x_{j,j} \leq \sum_{i=1}^{n} V_i \cdot x_{i,j} \leq U x_{j,j} \qquad \forall j \in \{1 \dots n\} \tag{4}$$

$$x_{i,j} \in \{0, 1\} \qquad \forall i, j \in \{1 \dots n\} \tag{5}$$

Formulation 1: Objective function and constraints in Hess et al. (1965).

been proposed (Kotthoff et al., 2015), including by extending the Hess model (Validi et al., 2020). Recently, Validi et al. (2020), presented and tested four different approaches to add contiguity to the Hess model. The contiguity formulations CUT and LCUT (Validi et al., 2020) can add an exponential number of constraints to the ILP formulation. Hence, the authors devise a procedure where CUT and LCUT constraints are added incrementally when needed as new integer solutions are found. However, in this work, we focus on compact formulations that can be solved by any off-the-shelf ILP or MOCO solver without developing a specific algorithm to handle the formulation. Hence, we briefly review the more compact flow-based formulation using Boolean variables, the authors named the MCF model.

In the MCF formulation, the authors start by creating a bi-directional version of the contiguity graph $G = (V, E)$ by replacing each undirected edge $\{i, j\} \in E$ by its directed counterparts (i, j) and (j, i). The set of edges pointing away from vertex i is denoted by $\delta^+(i)$ while, inversely, the set of edges pointing towards vertex i is denoted $\delta^-(i)$. Let $f_{i,j}^{a,b}$ be a Boolean variable that denotes if edge $(i, j) \in E$ is on

$$f^{a,b}(\delta^+(b)) - f^{a,b}(\delta^-(b)) = x_{a,b} \quad \forall a \in V \setminus \{b\}, \forall b \in V \tag{6}$$

$$f^{a,b}(\delta^+(i)) - f^{a,b}(\delta^-(i)) = 0 \quad \forall i \in V \setminus \{a,b\}, \forall a \in V \setminus \{b\}, \forall b \in V \tag{7}$$

$$f^{a,b}(\delta^-(b)) = 0 \quad \forall a \in V \setminus \{b\}, \forall b \in V \tag{8}$$

$$f^{a,b}(\delta^-(i)) \leq x_{i,b} \quad \forall a, i \in V \setminus \{b\}, \forall b \in V \tag{9}$$

Formulation 2: Contiguity Constraints for the Hess model (Validi et al., 2020).

the path to vertex a from its district center b. If this is the case, then $f_{i,j}^{a,b} = 1$. Otherwise, $f_{i,j}^{a,b}$ must be assigned value 0. In practice, one can interpret the directed graph as a flow network and $f_{i,j}^{a,b}$ denotes the flow passing at edge (i,j) considering b as the source and a as the sink. Moreover, let $f^{a,b}(S)$ be a shorthand for $\sum_{(i,j)\in S} f_{i,j}^{a,b}$ for a given $S \subseteq E$. Hence, by adding the constraints in Formulation 2 to the Hess model in Formulation 1, the new formulation only generates contiguous electoral districts.

In more detail, constraint (6) states that if a territorial unit a has b as its district center, then the flow coming out of b to a must equal 1. Moreover, constraints (7) ensure the flow conservation on the network. Finally, if b is a district center, then its incoming flow is 0 (8), and if i belongs to the flow path from b to a, then its flow is limited to 1 (9).

Besides the described MCF formulation, there are other flow-based constraint formulations for contiguity. For instance, a Mixed-Integer Programming (MIP) formulation has been applied to political districting in Japan considering the established administrative regions (Nemoto and Hotta, 2003; Ricca et al., 2013). Other MIP formulations adapted to different optimization criteria have also been proposed (Shirabe, 2009). A similar approach to the MCF formulation has been proposed that uses concepts of transportation network with functional capacities to model contiguity (Dugosija et al., 2020). Moreover, this formulation is also adapted to maximize the expected results of a given party in order to gerrymander the districting process.

2.3. Additional related work

There are a plethora of heuristic methods for electoral districting. A common approach of heuristic methods is to successively apply swap operations (Vickrey, 1961; Hayes, 1996). These algorithms start with an initial electoral map and repeatedly trade territorial units between electoral districts and evaluate if it resulted in an improvement, according to the objective function. Since these methods can quickly converge to an optimal local solution, an alternative is to use simulated annealing (Browdy, 1990; D'Amico et al., 2002; Macmillan, 2001).

Metaheuristic approaches for electoral districting include the usage of tabu search (Bozkaya et al., 2003) or genetic algorithms (Bação et al., 2005). While some authors combine compactness with a population deviation measure in the objective function (Bação et al., 2005; Joshi et al., 2012), others use a multi-objective metaheuristic framework to optimize both criteria (Vanneschi et al., 2017).

Another approach for electoral districting is the usage of clustering (Bodin, 2006; Bottman et al., 2007; Svec et al., 2007; Ricca et al., 2008; Guest et al., 2019). Some clustering algorithms (Bodin, 2006) focus on the contiguity of the electoral districts and population equilibrium but have a hard time finding compact districts since, in some cases, they do not control the shape of the clusters. In order to avoid this issue, Guest et al. (2019) applied the weighted k-means clustering algorithm to generate compact districts with roughly the same population.

There is also extensive work in social choice research related to election methods and how to condition the election results in different voting procedures (Bartholdi III et al., 1992; Hemaspaandra et al., 2007), as well as theoretical results on its computational complexity (Betzler and Uhlmann, 2009; Erdélyi et al., 2015a,b; Faliszewski et al., 2009; Liu et al., 2009). However, this paper assumes the first-past-the-post voting method, and our focus is on defining auditable formulations for the problem of electoral districting.

Some computational methods have also been proposed to evaluate if a particular outcome is neutral (Herschlag et al., 2017, 2020). The goal is to generate an ensemble (either using Markov Chain Monte Carlo DeFord et al., 2019; Fifield et al., 2020 or evolutionary approaches Tam Cho and Liu, 2016; Liu et al., 2016) that satisfies some design criteria. Afterward, a probability distribution is considered, and an evaluation is made on the degree to which a given electoral districting is neutral or not.

Single and multi-objective optimization has been used in many different domains to define districts or partition a given territory. For instance, multi-objective optimization has been used to define districts for public transportation in the Paris area (Tavares-Pereira et al., 2007), healthcare administrative authorities over geographic areas in England Datta et al. (2013), or census units in Canada (Datta et al., 2012).

The management of the forests and ecological areas also has extensive work on partitioning with constraints similar to the electoral districting problem. For instance, several harvest scheduling in agriculture areas (Murray, 1999; McDill et al., 2002; Constantino et al., 2008) or forest planning (Carvajal et al., 2013) also deal with contiguity constraints. The same also occurs in defining reserved areas for species (Önal and Briers, 2006; Önal et al., 2016). However, the models proposed for these problems suffer from the same drawbacks as those presented in the previous section.

3. New formulations for the electoral districting problem

In this section, we present new formulations for the electoral districting problem. First, we focus solely on maximizing compactness. Later, other objective functions are added, and multi-objective combinatorial optimization formulations are used. Next, the formulation sizes are discussed in Section 3.4 while optimizations are proposed in Sections 3.5 and 3.6.

3.1. Shortest-path contiguity formulation

We start by presenting a new formulation for the electoral districting problem based on the shortest paths in the contiguity graph of territorial units. Although this formulation guarantees that all districts are contiguous, it is not exact because some feasible electoral districting maps might not be considered. However, it is simple, effective and it seldom fails to include the optimal global solution. Moreover, even in those cases, it manages to find compact solutions, making it a valid option to use in scenarios where more complex complete formulations fail to produce an answer.

Let $\mathcal{V}$ denote the set of territorial units numbered from 1 to n and let V_i denote the number of registered voters[5] at territorial unit i. We say that two territorial units are adjacent (or neighbors) if they share a common border. Let N_j denote the set of neighbors of territorial unit j. If two territorial units i and j are adjacent, then $i \in N_j$ and $j \in N_i$. Let $\mathcal{K}$ denote the set of electoral districts to be built numbered from 1 to k.

In this first model, we consider two sets of Boolean variables:

[5] Alternatively, the total population can be used.

$$\max \sum_{i\in\mathcal{K}}\sum_{j\in\mathcal{V}}\sum_{j'\in N_j} L_{j,j'}\cdot b_{i,j,j'} \tag{10}$$

$$\sum_{i\in\mathcal{K}} x_{i,j} = 1 \qquad \forall j\in\mathcal{V} \tag{11}$$

$$\sum_{j\in\mathcal{V}} x_{i,j} \geq 1 \qquad \forall i\in\mathcal{K} \tag{12}$$

$$L \leq \sum_{j\in\mathcal{V}} V_j\cdot x_{i,j} \leq U \qquad \forall i\in\mathcal{K} \tag{13}$$

$$x_{i,j} - b_{i,j,j'} \geq 0 \qquad \forall i\in\mathcal{K},\ j\in\mathcal{V}, j'\in N_j \tag{14}$$

$$x_{i,j'} - b_{i,j,j'} \geq 0 \qquad \forall i\in\mathcal{K},\ j\in\mathcal{V}, j'\in N_j \tag{15}$$

$$x_{i,j}, b_{i,j,j'} \in \{0,1\} \qquad \forall i\in\mathcal{K}, j\in\mathcal{V}, j'\in N_j \tag{16}$$

Formulation 3: Base formulation.

- $x_{i,j}$ denotes if territorial unit $j\in\mathcal{V}$ is assigned to electoral district $i\in\mathcal{K}$.
- $b_{i,j,j'}$ denotes if territorial units $j, j'\in\mathcal{V}$ both belong to electoral district $i\in\mathcal{K}$.

Formulation 3 contains the base formulation for this model. In more detail, the constraints are as follows: (i) each territorial unit must belong to a single electoral district (11), (ii) each electoral district must have at least one territorial unit (12), (iii) the number of voters in each electoral district must be at least L and at most U (13), and (iv) if two neighboring territorial units j and j' are not in the same electoral district $i\in\mathcal{K}$ then $b_{i,j,j'}$ is set to 0 (14) (15). Finally, let $L_{j,j'}$ be the border length between neighboring territorial units j and j'. Our objective function (10) aims at maximizing the total length of the internal borders (i.e., minimizing the border length between different electoral districts). Note that any optimal solution must have $b_{i,j,j'}$ set to 1 if two neighboring territorial units j and j' are in the same electoral district. This occurs since the goal of the objective function is to maximize. As an alternative, we could force it by adding the constraints $b_{i,j,j'} - x_{i,j} - x_{i,j'} \geq -1$, $\forall i\in\mathcal{K},\ j\in\mathcal{V}, j'\in N_j$.

Observe that our goal is to obtain compact electoral districts, but using a different metric than other approaches (Hess et al., 1965; Validi et al., 2020). While other formulations try to minimize the total distance between the geographic centers of territorial units to the geographic center of the electoral district, we focus on the border length as a proxy measure for the shape of the electoral district. Despite some eventual drawbacks such as the coastline effect (Duchin and Tenner, 2018), the usage of border length has several advantages over geographic centers. For instance, in some cases, the geographic center might be outside the territorial unit due to: (i) oddly shaped territorial units or (ii) territorial units that contain enclaves. Moreover, since the border length is a local measure, we can develop models that focus on territorial units' adjacency instead of considering non-adjacent relations such as the distance from a territorial unit to the center of the electoral district. Therefore, we believe the border length is also a valid measure for compactness, and our experimental results confirm that.

Note that the base formulation in Formulation 3 must be extended with additional constraints to guarantee the contiguity of the electoral districts. For that, consider the contiguity graph of the territorial units, where the weight of 1 is defined for every edge between two adjacent territorial units. Given this weighted graph, one can compute the matrix D of shortest distances between all pairs of graph vertices (territorial units) (Cormen et al., 2001) in polynomial time. Based on the information from the shortest paths between all pairs of territorial units, one can define that between two non-neighboring territorial units j and j' that belong to the same electoral district, then one of its shortest paths in the contiguity graph must also belong to the electoral district. This can be achieved with the following constraints:

$$-x_{i,j} - x_{i,j'} + \sum_{j''\in N_{j'}, D_{j,j''}<D_{j,j'}} (x_{i,j''}) \geq -1 \forall i\in\mathcal{K}, j, j'\in\mathcal{V}, j'\notin N_j \tag{17}$$

Notice that if two non-neighboring territorial units j and j' belong to the same electoral district i (i.e. $x_{i,j} = x_{i,j'} = 1$), then there must exist at least another territorial unit j'' in electoral district i such that j'' is neighbor of j' and the distance between j and j'' is lower than the distance between j and j' ($D_{j,j''} < D_{j,j'}$). In other words, a shortest path between j and j' must exist inside the electoral district. Finally, note that (17) only applies if j and j' belong to the same electoral district i. Otherwise, the constraint is trivially satisfied.

Intuitively, one might think that it is beneficial for compactness to include shortest paths between territorial units. In our experimental results, it is rare that the most compact solution does not include a shortest path between territorial units in the electoral district. However, this always depends on the neighboring relation between the territorial units. In Fig. 2 we present one scenario where the optimal solution is excluded due to the shortest path constraints in (17).

Note that the distance between territorial units A and B (through the shortest path) is 2 (passing through C). Therefore, according to the shortest path constraints in (17), the only way for A and B to be in the same electoral district is also containing C. However, in this context, since C is a high-density territorial unit, this is not feasible, and the optimal solution (right map in Fig. 2) is not considered due to (17). Although it is not easy to visually identify which one is better, the map on the right has a 5.22% increase in the objective function.

3.2. Tree-based contiguity formulation

The previous formulation allows to generate electoral maps that are contiguous, but it excludes some feasible solutions. Therefore, the optimal solution might also be excluded, as shown in the example from Fig. 2. This section uses alternative contiguity constraints to ensure that the territorial units assigned to each electoral district form a tree in the contiguity graph. As a result, all electoral districts will be contiguous.

Recall the base formulation in Formulation 3. In order to represent the tree, for each electoral district we extend this formulation with the following new sets of Boolean variables:

- $p_{j,j'}$ denotes if territorial unit $j'\in N_j\cup\{0\}$ is predecessor of $j\in\mathcal{V}$
- $d_{j,l}$ denotes if territorial unit $j\in\mathcal{V}$ is at depth l in the tree

In this case, observe that only neighbors can be considered as predecessors of a given territorial unit j in variables $p_{j,j'}$. Moreover, for the particular case of the root of a given tree, those vertices have 0 as their predecessor. Variables $d_{j,l}$ encode the depth in the tree for each vertex j. The root node is considered to be at depth 1, and the depth is increased by one at each level of the tree. Let $\mathcal{M} = \{1,\dots,m\}$ denote the set of possible depths in a tree where m is the maximum number of territorial units that can be assigned to an electoral district. Considering there are limits on the number of voters in each electoral district, one can easily calculate a priori a proper value for m. In the worst case, for a problem instance with n territorial units and k electoral districts, one can safely define that $m = n - k$.

Formulation 4 presents the base formulation from Formulation 3 extended with the contiguity constraints of the tree-based encoding. In more detail, the constraints for the tree-based encoding are as follows: (i) each territorial unit j must have a neighbor as predecessor in the tree that represents a given electoral district or 0 if j is a root (18), (ii) if two neighbors j and j' belong to different electoral districts, then they cannot have a predecessor relation (19), (iii) for each pair of neighbors j and j', the predecessor relation can only be established in one direction (20), (iv) in each electoral district, there can only be one root node (21), (v) each territorial unit can only be assigned a single depth (22), (vi) a root node is always at depth 1 (23), (vii) the depth

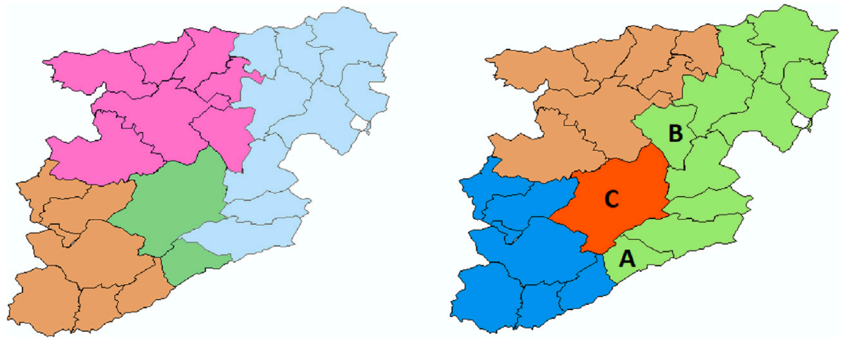

Fig. 2. Redistricting the same area with the shortest path constraints (left map) and the excluded optimal solution (right map).

$$\boxed{\begin{array}{lll}
\max & \sum_{i\in\mathcal{K}}\sum_{j\in\mathcal{V}}\sum_{j'\in N_j} L_{j,j'}\cdot b_{i,j,j'} & \\
 & \sum_{i\in\mathcal{K}} x_{i,j} = 1 & \forall j\in\mathcal{V}\\
 & \sum_{j\in\mathcal{V}} x_{i,j} \geq 1 & \forall i\in\mathcal{K}\\
 & L \leq \sum_{j\in\mathcal{V}} V_j\cdot x_{i,j} \leq U & \forall i\in\mathcal{K}\\
 & x_{i,j} - b_{i,j,j'} \geq 0 & \forall i\in\mathcal{K},\ j\in\mathcal{V}, j'\in N_j\\
 & x_{i,j'} - b_{i,j,j'} \geq 0 & \forall i\in\mathcal{K},\ j\in\mathcal{V}, j'\in N_j\\
 & x_{i,j}, b_{i,j,j'} \in \{0,1\} & \forall i\in\mathcal{K}, j\in\mathcal{V}, j'\in N_j
\end{array}}$$

$$\left(\sum_{j'\in N_j} p_{j,j'}\right) + p_{j,0} = 1 \qquad \forall j\in\mathcal{V} \qquad (18)$$

$$-x_{i,j} + x_{i,j'} - p_{j',j} \geq -1 \qquad \forall i\in\mathcal{K}, j\in\mathcal{V}, j'\in N_j \qquad (19)$$

$$p_{j,j'} + p_{j',j} \leq 1 \qquad \forall j\in\mathcal{V}, j'\in N_j \qquad (20)$$

$$-x_{i,j} - x_{i,j'} - p_{j,0} - p_{j',0} \geq -3 \qquad \forall i\in\mathcal{K}, j, j'\in\mathcal{V}, j\neq j' \qquad (21)$$

$$\sum_{i=1}^{m} d_{j,i} = 1 \qquad \forall j\in\mathcal{V} \qquad (22)$$

$$-p_{j,0} + d_{j,1} \geq 0 \qquad \forall j\in\mathcal{V} \qquad (23)$$

$$-p_{j,j'} - d_{j',l} + d_{j,l+1} \geq -1 \qquad \forall j\in\mathcal{V}, j'\in N_j, l\in\mathcal{M}\setminus\{m\} \qquad (24)$$

$$-d_{j,m} - p_{j',j} \geq -1 \qquad \forall j\in\mathcal{V}, j'\in N_j \qquad (25)$$

$$\sum_{j\in\mathcal{V}} p_{j,0} = k \qquad (26)$$

$$p_{j,j'} \in \{0,1\} \qquad \forall j\in\mathcal{V}, j'\in N_j\cup\{0\} \qquad (27)$$

$$d_{i,l} \in \{0,1\} \qquad \forall i\in\mathcal{K}, l\in\mathcal{M} \qquad (28)$$

Formulation 4: Base formulation from Formulation 3 (boxed) extended with tree-based formulation for contiguity of electoral districts.

of a territorial unit is one more than its predecessor (24), (viii) the nodes at depth m must be leaves of the tree, i.e., these nodes cannot be predecessors of other nodes (25), (ix) the number of roots must be exactly the same as the number of electoral districts (26).

We note that tree-based contiguity constraints have been previously used. Kotthoff et al. (2015) propose to encode a tree with a proper ranking among all territorial units inside the electoral district. In practice, this corresponds to a topological sorting of the predecessor relation graph among the territorial units in each electoral district. However, our formulation shows that this ordering is not necessary.

Another tree-based formulation has been proposed by Duque et al. (2011). Their work differs from ours since they use different constraints to prevent cycles and integer (i.e., non-binary) variables. In addition, they propose an order-based formulation that also builds a tree. Nevertheless, they establish an order for the territorial units in each district, while we rely on a predecessor relationship between adjacent territorial

$$\min \sum_{i\in\mathcal{K}} l_{i,p} \qquad (29)$$

$$K_{i,p,p'} l_{i,p,p'} + \sum_{j\in\mathcal{V}} (v_{p,j} - v_{p',j}) x_{i,j} \geq 1 \qquad \forall i\in\mathcal{K}, \forall p, p'\in\mathcal{P}, p\neq p' \qquad (30)$$

$$K_{i,p,p'} l_{i,p,p'} + \sum_{j\in\mathcal{V}} (v_{p,j} - v_{p',j}) x_{i,j} \leq K_{i,p,p'} \qquad \forall i\in\mathcal{K}, \forall p, p'\in\mathcal{P}, p\neq p' \qquad (31)$$

$$-l_{i,p} + \sum_{p\in P} l_{i,p,p'} \geq 0 \qquad \forall i\in\mathcal{K}, \forall p, p'\in\mathcal{P}, p\neq p' \qquad (32)$$

$$l_{i,p} - l_{i,p,p'} \geq 0 \qquad \forall i\in\mathcal{K}, \forall p, p'\in\mathcal{P}, p\neq p' \qquad (33)$$

$$l_{i,p}, l_{i,p,p'} \in \{0,1\} \qquad \forall i\in\mathcal{K}, \forall p, p'\in\mathcal{P}, p\neq p' \qquad (34)$$

Formulation 5: Gerrymandering constraints and objective function.

units. Moreover, they define a heterogeneity distance for all pairs of territorial units. This distance does not have to be a geographic distance but can be a social measure distance (e.g., mean of housing prices) between territorial areas. In our work, we do not require to define a distance between all pairs of territorial units, resulting in a more compact formulation.

3.3. Gerrymandering constraints

The formulations proposed in the previous sections are solely concerned with optimizing compactness. However, gerrymandering can often be used to produce electoral maps favoring some political party (or being less favorable to some other party). This section extends the previous formulations by considering which party is likely to win each electoral district in a first-past-the-post (FPTP) electoral system based on results from the last elections.

Let $\mathcal{P}$ denote the set of political parties to be considered in the election. In order to add gerrymandering objectives, we consider the following new sets of Boolean variables:

- $l_{i,p,p'}$ denotes if party $p\in\mathcal{P}$ is likely to lose the election at electoral district i to party $p'\in\mathcal{P}$
- $l_{i,p}$ denotes if party $p\in\mathcal{P}$ is likely to lose the election at electoral district i against at least one of the other parties

Let $v_{p,j}$ be an estimate of the votes to be obtained from party $p\in\mathcal{P}$ in territorial unit $j\in\mathcal{V}$. Observe that the vote estimation can be obtained from previous election results. However, the procedure to best determine vote estimation is out of the scope of this paper since our focus is on problem formulation.

Formulation 5 presents the additional set of constraints and objective function to consider when trying to gerrymander the electoral district map. First, constraints (30) and (31) allow to determine if a political party $p\in\mathcal{P}$ has fewer votes than another party $p'\in\mathcal{P}$ in electoral district $i \in \mathcal{K}$. Note that constant $K_{i,p,p'}$ can be any integer value such that $K_{i,p,p'} > \sum_{j \in \mathcal{V}} |v_{p,j} - v_{p',j}|$. Next, constraints (32) and (33) are used to determine if a political party $p \in \mathcal{P}$ loses the elections in electoral district $i \in \mathcal{K}$. Finally, to gerrymander the construction of the electoral districts, one can define an objective function that maximizes the number of electoral district wins of a particular party $p \in \mathcal{P}$. For that, the objective function (29) minimizes the sum of all $l_{i,p}$ variables.

Note that if the goal is to minimize the electoral district wins (maximize losses), then we simply have to create a maximizing objective function instead of a minimizing one. Furthermore, we define a Multi-Objective Combinatorial Optimization (MOCO) formulation that gerrymanders the electoral district and maximizes compactness by joining the tree-based formulation from Formulation 4 and the gerrymandering formulation from Formulation 5. Observe that we use the tree-based formulation since the shortest-path formulation from Section 3.1 does not guarantee optimality. Experimental results from this MOCO formulation applied to the case of electoral districting of Portugal are presented in Section 4.

The work of Swamy et al. (2019) includes an optimization criteria minimizing the expected voting differences between the two main parties to create competitive districts. If one were to create advantageous electoral maps towards a party, maximizing the voting differences in favor of a party is also an option.

3.4. Formulation sizes

In this paper, we focus on new formulations for generating electoral districts. This section analyzes the size of the formulation being generated by the newly proposed encoding. In particular, we focus on the size of the complete MOCO formulation joining the variables and constraints from the base formulation in Section 3.1 with the ones defined in Section 3.2.

Let n denote the number of territorial units and b denote the number of borders between them in the contiguity graph (i.e., the number of edges). Moreover, let k be the number of electoral districts to be built and m the maximum number of territorial units in an electoral district.

We start by analyzing the number of variables. First, in our formulations, the number of variables to represent the assignment of territorial units to electoral districts ($x_{i,j}$ variables) is $O(n \times k)$ with $1 \leq k \leq n$. Nevertheless, it is expected that k to be much smaller than n. Observe that the Hess model uses $O(n^2)$ variables to represent the assignment of territorial units to electoral districts.

On the other hand, the number of variables to encode the contiguity of the electoral districts ($p_{j,j'}$ variables) is b. In the worst case, we have $b = O(n^2)$. Still, we expect the number of borders to be much smaller than n^2, i.e., the number of edges in the contiguity graph is usually very limited.[6] Finally, the number of variables to represent the depth of each territorial unit in the tree-based representation ($d_{j,l}$ variables) is $O(n \times m)$. In the worst case, we have $m = n - k$ resulting in $n^2 - nk$ variables. However, one can expect m to be much smaller than $n - k$ since voters are not usually so concentrated.

On the number of constraints, we have that (21) introduces $O(n^2 \times k)$ constraints and (24) adds $O(b \times m)$. Hence, one can bound the number of overall constraints by $O(n^3)$.

Finally, observe that the contiguity constraints by Validi et al. (2020) (see Formulation 2) to be added to the Hess model uses $O(n^3)$ variables.

[6] Most of the maps of territorial units correspond to a planar graph.

3.5. Additional techniques

A trivial upper bound on the maximum depth m of each tree representing a district is defined as $m = n - k$. In the worst case, there is a district with m territorial units, while the remaining $k - 1$ districts only have one territorial unit. However, this is only expected to occur if there are territorial units with enough voters to be a district on its own. Otherwise, tighter bounds can be defined.

Another trivial approach is to consider that the district with the most territorial units is composed of the territorial units with fewer voters. Let U denote the maximum number of voters assigned to a district and let V_j denote the number of voters in territorial unit j. Furthermore, let O define the ordered list of the n territorial units indexes in a non-decreasing order of voters. Therefore, one can define m as follows:

$$m = max\{u : \sum_{j=1}^{u} V_{O[j]} \leq U\} \tag{35}$$

Clearly, a district with $u + 1$ territorial units cannot exist since it would not satisfy constraint (13). However, a limitation of this method is that the territorial units with fewer voters might not be contiguous. In the worst case, the units form a graph that corresponds to a linked list. Notwithstanding, in this case, if we consider the root of a district to be a territorial unit in the middle of the list, there is still always a solution since our constraints defined in Section 3.2 do not preclude it. Hence, we can safely divide m by 2 without removing any feasible district.

Drawing from Validi et al. (2020), it is also possible to determine that some pairs of territorial units cannot be in the same district. Consider the following weighted directed graph $G_w = (V, E)$ defined as follows:

- For each territorial unit $j \in \mathcal{V}$ there is a corresponding vertex $j \in V$
- For each pair of adjacent territorial unit j and j', we define two edges (j, j') and (j', j) in E such that $w(j, j') = V_{j'}$ and $w(j', j) = V_j$

Let $\delta(j, l)$ denote the shortest path from j to l in G_w. If $V_j + \delta(j, l) > U$, then territorial units j and l cannot be in the same district. Note that it would require more voters than the upper limit U for these territorial units to be in the same contiguous district. Hence, the following constraint could be safely added to our formulation:

$$\forall i \in \mathcal{K} : -x_{i,j} - x_{i,l} \geq -1 \tag{36}$$

The same principle can also be applied in the Hess model, resulting in requiring territorial units j and l to always be in different districts as follows:

$$\forall i \in \{1 \dots n\} : -x_{j,i} - x_{l,i} \geq -1 \tag{37}$$

3.6. Symmetry constraints

Symmetry constraints are simply additional constraints that allow removing equivalent solutions. The use of symmetry constraints is known to be very effective in some domains (Aloul et al., 2006; Marques-Silva et al., 2008; Sakallah, 2009), but it depends on the pruning capacity and the extra effort necessary to maintain these constraints. Hence, in some cases, it might not be helpful (Orvalho et al., 2019).

Consider four territorial units A, B, C, D and two districts. Suppose that a solution is reached where A, B form district 1 and C, D form district 2. Notice that we would have essentially the same solution if we were to assign A, B to district 2 and C, D to district 1. In this case, we say that these two solutions are symmetric to the district assignment.

To avoid these symmetries between districts, one can add constraints that cut these symmetric assignments, just allowing one of them to be a solution to our model. Observe that the model remains valid

Table 1
Identifiers, names and current number of elected officials for each region.

ID	Region name	Elected
01	Aveiro	16
02	Beja	3
03	Braga	19
04	Bragança	3
05	Castelo Branco	4
06	Coimbra	9
07	Évora	3
08	Faro	9
09	Guarda	3
10	Leiria	10
11	Lisboa	48
12	Portalegre	2
13	Porto	40
14	Santarém	9
15	Setúbal	18
16	Viana do Castelo	6
17	Vila Real	5
18	Viseu	8
Total		215

since we are only cutting symmetric solutions. In our domain, we can add constraints such that the number of voters in each district is non-decreasing according to the district identifier. Recall that V_j denotes the number of voters in territorial unit j. Then, we can add the following constraints:

$$\forall i \in \mathcal{K} \setminus \{k\} : \sum_{j \in \mathcal{V}} V_j \cdot x_{i,j} \leq \sum_{j \in \mathcal{V}} V_j \cdot x_{i+1,j} \tag{38}$$

Observe that there are also symmetries inside districts. For example, consider the same four territorial units A, B, C, D divided between the same two districts. It is equivalent assigning A, B to district 1 and C, D to district 2 with A and C as roots or the same assignment with B and D as roots and A and C as leaves.

A possible solution to avoid inner district symmetries is to add a constraint that would only allow the territorial unit with the lowest ID to be the root of the district. It can be defined as follows:

$$\forall i \in \mathcal{K}, \forall j \in \mathcal{V}, j' \in N_j, j < j' : -p_{j,j'} - p_{j',0} \geq -1 \tag{39}$$

These techniques can be easily adapted to the Hess model (Hess et al., 1965) (see Section 2.2 and Formulation 1) in order to cut the same type of symmetries:

$$\forall j, j' \in V, j < j' : -x_{j,j'} - x_{j',j'} \geq -1 \tag{40}$$

Note that this optimization is only valid if the objective function used to maximize compactness does not depend on the root of the districts.

4. Experimental results

In this section, our formulations proposed in Section 3 are compared with previous approaches in order to test their efficiency. Therefore, Section 4.1 describes the benchmark instances used in the experimental evaluation using a scenario of electoral districting in Portugal. Next, Section 4.2 considers the problem of maximizing compactness in electoral districting, and Section 4.3 tests real-world scenarios where a party tries to gerrymander the electoral maps.

The results presented use real data from continental Portugal.[7] The continental part is divided into 18 regions where each region elects a given number of officials to the national parliament. Fig. 3 shows the map with the 18 regions.[8] and Table 1 the number of officials per region. Observe that the number of elected officials depends on the population in each region. Regarding the computational infrastructure, all experimental results were obtained on an Intel Xeon Silver 4110 processor running Debian Linux with 64 GB of RAM. Moreover, single-objective problem instances were solved using CPLEX 12.6.0,[9] while multi-objective formulations are solved using the `sat4jMoco` solver.[10]

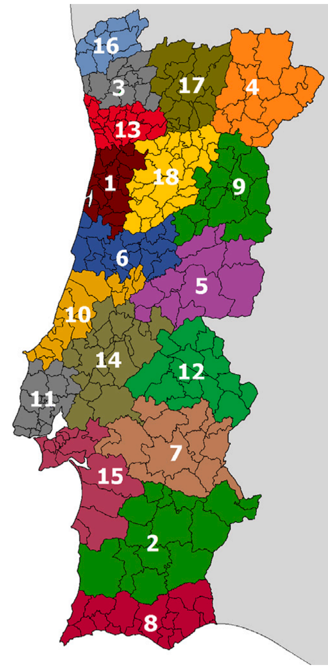

Fig. 3. Map of Portuguese regions.

4.1. Electoral districting in Portugal

Portugal uses a party-list proportional representation voting system. However, in the context of this work, we study the scenario of changing it to a parallel voting system. This change was proposed by the two major Portuguese parties in 1998 (Bação et al., 2005) making it a strong candidate for a future electoral system revision in Portugal.

In a parallel electoral system, a percentage of the parliament seats is distributed using single-member districts. The rest uses a party-list proportional representation method (most commonly, the D'Hondt method) at the national level. The idea behind it is to bring the voters closer to politics by voting directly to elect a parliament member who is typically more concerned and connected with their local electoral area. Nevertheless, some proportionality is still maintained through the national circle at the party level (avoiding the tendency to create a two-party system of a first-past-the-post system).

In the parallel voting system scenario, the number of single-member districts to be created in each region is half of the current number of elected officials (see Table 1), rounded up. Hence, the total number of single-member districts becomes 112 in continental Portugal, as shown in Table 2.

To generate the electoral maps (instances) tested in the following sections, we follow a set of rules, typical for redistricting, which are:

[7] We do not consider electoral districting of the archipelagos of Azores and Madeira since a different approach would need to be applied.

[8] In Portuguese, each of the map regions in Fig. 3 is named a *"distrito"* To avoid confusion with electoral districts, we use the term region to designate these areas.

[9] https://www.ibm.com/analytics/cplex-optimizer.

[10] https://gitlab.ow2.org/sat4j/moco.

Table 2
Identifiers, names and number of electoral districts to be created for each region under a parallel voting system.

ID	Region name	Electoral districts
01	Aveiro	8
02	Beja	2
03	Braga	10
04	Bragança	2
05	Castelo Branco	2
06	Coimbra	5
07	Évora	2
08	Faro	5
09	Guarda	2
10	Leiria	5
11	Lisboa	24
12	Portalegre	1
13	Porto	20
14	Santarém	5
15	Setúbal	9
16	Viana do Castelo	3
17	Vila Real	3
18	Viseu	4
Total		112

1. The number of people registered to vote in each electoral district must not diverge more than 15% from the theoretical best value β.
2. The new electoral districts should be as compact as possible.
3. All electoral districts must be contiguous.
4. There must be conformity to administrative boundaries, i.e., new electoral maps must respect the current administrative divisions. The largest possible administrative divisions should be kept whenever possible without disregarding the first rule.

The population margin is difficult to set. The bigger the margin from a theoretical ideal value, the less equal is the voting power between electoral districts. Some countries, such as the United States of America (USA), prefer a margin as low as possible. When necessary, in most states in the USA, electoral districts can be drawn from census blocks to create population balance.

In our scenarios, the maximum margin value is set at 15%. This value lies between the 10% used in countries such as Italy, Australia, or Ukraine and the maximum value presented in one of the prominent propositions for Portugal back in 1998 (Bação et al., 2005) which was 25% (such value is also used in Canada and Germany).

Whenever possible, we try to preserve the municipality's administrative divisions (the smaller subdivisions in Fig. 3) inside each region. However, that is not possible in most cases, either because the number of electoral districts is larger than the number of municipalities of the region or because the population differences between municipalities do not allow us to join them while respecting the population similarity rule. In the cases mentioned earlier, we start by splitting the region into different areas at the municipality level. Next, each area is redistricted at the civil parish level to define the electoral map.

Let $\vec{d}_k^s$ denote the vector of positive integers with size s such that the sum of its elements is k (i.e. $\sum_{i=1}^{s} d_k^s[i] = k$). Moreover, the values of vector $\vec{d}_k^s$ are sorted in a non-increasing order (i.e. $\forall i \in \{1 \dots s-1\}$: $d_k^s[i] \geq d_k^s[i+1]$) and the maximum difference between any two values in the vector is 1 (i.e. $\forall i \in \{1 \dots s\}$: $|d_k^s[i] - d_k^s[j]| \leq 1$).

In practice, $\vec{d}_k^s$ is a vector that splits k into s values such that the sum of the s values equals k and the maximum difference between the s values is 1. For instance, if $k = 14$ and $s = 4$, then we have that $\vec{d}_{14}^4$ is $\langle 4, 4, 3, 3 \rangle$. Observe that if k is a multiple of s, then all values in the vector are $\frac{k}{s}$.

In Formulation 4, all areas will have a number of people bounded between a lower limit L and an upper limit U. However, if we change these limits, one can use Formulation 4 to split a given region into areas where each area represents several electoral districts at the municipality level. Next, each area is redistricted at the civil parish level. Suppose we want to split a region according to vector $\vec{d}_k^s$. Then we instantiate Formulation 4 with splitting the region into s areas such that the population lower limits of each area $i \in \{1 \dots s\}$ are defined by $\vec{d}_k^s[i] \cdot L$ and the upper limits are defined by $\vec{d}_k^s[i] \cdot U$.

For instance, let $\vec{d}_{14}^4 = \langle 4, 4, 3, 3 \rangle$ and lets solve Formulation 4 according to the split in $\vec{d}_{14}^4$. In this case, we can define that the region is to be split into only 4 areas where two areas represent 4 electoral districts and the other two areas represent 3 electoral districts. For that, we only need to change the population bound constraints as previously explained. If the resulting formulation is feasible, then we use the split found at the municipality level and proceed to define the electoral districts at the civil parish level for each of these areas.

In order to find the split of a region at the municipality level where k electoral districts are to be defined, we apply the following steps:

for ($s \leftarrow 2$; $s < k$; $s \leftarrow s+1$) **do**
  Generate $\vec{d}_k^s$
  Solve Formulation 4 to split the region in s areas with vector $\vec{d}_k^s$
  if Formulation 4 is feasible for $\vec{d}_k^s$ **then**
    return the optimal area split found
  end if
end for
return No split is performed on the region

In our procedure, we first try to split the region into two areas with a similar number of electoral districts. If this is infeasible, we increase the number of areas to divide the region. If the procedure fails to split at the municipality level, then no split is performed and the whole region is considered at the parish level. In the case of continental Portugal, we could always find a split for each region.

After this first step, we end up with 31 areas of continental Portugal where a given number of electoral districts are to be defined. These 31 areas are characterized in Table 3 and define the benchmark instances used to evaluate the proposed ideas of the paper.

In the following subsections, we provide experimental results to answer the following research questions:

Q1. Can we improve previous formulations with the additional techniques from Sections 3.5 and 3.6?
Q2. How does our tree-based formulation compare with previously proposed formulations?
Q3. The shortest-path based formulation can just provide an approximation. However, how far are these results from optimal? What about the performance?
Q4. Can we solve the districting problem for Portugal just considering compactness?
Q5. Can we gerrymander the political districts? What flip can we make to the electoral results?

4.2. Maximizing compactness

In order to answer the first research question, we implemented the original Hess formulation (Hess et al., 1965) (see Formulation 1) with the contiguity constraints by Validi et al. (2020) (see Formulation 2), abbreviated simply as Hess formulation in the following sections. Next, the optimizations proposed in Section 3.5 (constraints (36)) and Section 3.6 (constraints (38) and (39)) were introduced. Both implementations are evaluated in the 31 benchmark instances defined in the previous section, where the goal in each instance is solely to optimize compactness. Therefore, both formulations were evaluated using CPLEX 12.6.0 with a time limit of three hours (10 800 s).

Table 4 presents the results for the Hess formulation with and without the proposed techniques. Whenever that limit is exceeded,

Table 3
Correspondence between instance number, Portuguese region, the type of administrative boundaries used, the number of territorial units and districts to be created inside the area.

Instance number	Portuguese region	Administrative boundary	Territorial units	Electoral districts
1	01	Parish	25	2
2	01	Parish	30	2
3	01	Parish	43	2
4	01	Parish	49	2
5	02	Municipality	14	2
6	03	Parish	100	3
7	03	Parish	143	3
8	03	Parish	104	4
9	04	Municipality	12	2
10	05	Municipality	11	2
11	06	Parish	68	2
12	06	Parish	87	3
13	07	Municipality	14	2
14	08	Municipality	16	5
15	09	Municipality	14	2
16	10	Parish	57	2
17	10	Parish	53	3
18	11	Parish	20	8
19	11	Parish	30	8
20	11	Parish	84	8
21	12	Municipality	15	1
22	13	Parish	11	5
23	13	Parish	22	5
24	13	Parish	47	5
25	13	Parish	163	5
26	14	Municipality	21	5
27	15	Parish	12	4
28	15	Parish	43	5
29	16	Municipality	10	3
30	17	Municipality	14	3
31	18	Municipality	24	4

Table 4
Comparison between the execution times of the Hess model (Hess et al., 1965) with the contiguity constraints by Validi et al. (2020) with and without the optimizations proposed in Sections 3.5 and 3.6. Highlighted in orange are cases where the time limit (3 h) was exceeded.

Instance number	Territorial units	Time without optimizations	Time with optimizations
1	25	6 881.66	24.98
2	30	10 800.00	36.55
3	43	10 800.00	939.45
4	49	10 800.00	3 903.59
5	14	8.96	0.72
6	100	10 800.00	10 800.00
7	143	10 800.00	10 800.00
8	104	10 800.00	10 800.00
9	12	4.06	0.46
10	11	1.60	0.04
11	68	10 800.00	10 800.00
12	87	10 800.00	10 800.00
13	14	7.60	0.47
14	16	4.06	0.05
15	14	10.07	0.69
16	57	10 800.00	1 486.81
17	53	10 800.00	10 800.00
18	20	9.71	0.50
19	30	10 800.00	40.89
20	84	10 800.00	10 800.00
21	15	0.15	0.04
22	11	0.51	0.03
23	22	5 761.53	7.95
24	47	10 800.00	10 800.00
25	163	10 800.00	10 800.00
26	21	1 290.08	4.05
27	12	2.11	0.49
28	43	10 800.00	10 800.00
29	10	0.53	0.09
30	14	8.27	1.27
31	24	8 095.44	3.52

the background of the individual cell is highlighted. Otherwise, the time spent to find the optimal solution is shown. The experimental results clearly show the effectiveness of cutting symmetries in the Hess formulation. Note that we can increase the number of solved instances within the time limit and significantly improve the performance in the remaining instances. Observe that the performance never worsens and, in some cases, improves by two orders of magnitude. Although not shown, in instances where the time limit is reached, a better approximation is usually obtained using the techniques proposed in the paper.

Next, we compare the improved Hess formulation with the proposed tree-based (see Section 3.2) and shortest path-based (see Section 3.1) formulations. Table 5 shows the computational time for each approach, as well as the number of variables used in each formulation and the final population deviation to the theoretical best value. Whenever the time limit is reached, that entry's background is highlighted. Moreover, the fastest computational time between the optimized Hess formulation and the tree-based formulation is also highlighted in bold.

Observe that both our formulations produce formulas that are much smaller than the ones using the Hess formulation. This occurs since the number of electoral districts to create is much smaller than the number of territorial units in each instance. Furthermore, the shortest-path formulation uses fewer variables than the tree-based since it does not need the extra variables to encode the contiguity of the electoral districts. Overall, it is clear that the proposed tree-based formulation improves upon the Hess formulation being able to solve more instances. Additionally, when both formulations can find an optimal solution, the tree-based formulation is sometimes much faster. Finally, observe that CPLEX can solve all instances of the shortest-path based formulation. However, in some cases, the optimal solution of the shortest-path formulation is worse since this formulation excludes some feasible solutions. There is one instance (number 13, highlighted in red) where the shortest-path formulation excludes all feasible solutions. Nevertheless, as shown in the last column of Table 5, for this set of instances, the quality of the solution using the shortest-path formulation is not usually significantly impacted. Moreover, there are 5 instances where the tree-based formulation reached the time limit, and CPLEX could not find an integer solution. As a result, for these instances and the one where the shortest-path formulation cut all solutions, there is no comparison on the quality of the solutions.

Simulated annealing is a well-known heuristic procedure that has been shown to be effective in solving districting problems. Hence, we have also implemented a tool using this technique for solving our instances. The simulated annealing results were obtained considering 10^8 iterations (taking between 1 to 5 h of CPU time per instance) and maximizing the internal frontiers of each electoral district. Moreover, we also incorporated a penalty when a solution did not satisfy the contiguity constraints or the population bounds in each electoral district. With these penalties in the objective function, the simulated annealing approach was able to produce better results than just maximizing the internal frontiers. Table 6 compares the quality of the solution provided by the simulated annealing algorithm with our approaches. Each cell contains the value of the best solution found for each method. When an approach does not provide a feasible solution, `NA` is presented. In bold, we highlight the best results for each instance.

These results show that our shortest-path based formulation clearly outperforms the simulated annealing approach by providing the best solution in 20 out of the 31 instances. In contrast, the simulated annealing only provides the best quality solution in 13 out of 31. Moreover, observe that it is hard for the simulated annealing approach to find feasible solutions due to failing to define contiguous districts, in particular in instances with a larger number of territorial units. This approach was unable to find a feasible solution for 12 out of 31 instances compared to just 1 with the shortest-paths method.

Table 5
Comparison between the optimized Hess formulation, our proposed Tree-based (exact) and Shortest-path formulations. For each formulation we present the number of variables, solving CPU time and maximum deviation of population between electoral districts in the solution. Finally, the last column presents the compactness increase of using the Tree-based approach. Highlighted in orange are cases where the time limit (3 h) was exceeded (TLE). The fastest execution times are highlighted in bold between Hess and Tree-based formulations.

Ins.	Hess formulation optimized			Tree-based formulation			Shortest-path formulation			Obj. Incr.
	Var.	Time	Max. Dev.	Var.	Time	Max. Dev.	Var.	Time	Max. Dev.	
1	70 681	25.0	5.3	562	**0.5**	5.3	156	0.5	5.3	0.0%
2	119 766	36.6	9.4	766	**5.6**	9.4	188	0.3	9.4	0.0%
3	349 555	939.5	5.9	1 310	**171.1**	5.9	268	0.6	5.9	0.0%
4	554 746	3903.6	13.0	1 722	**32.0**	13.0	322	6.7	11.4	**0.8%**
5	10 021	0.7	11.9	240	**0.6**	11.9	78	0.0	11.9	0.0%
6	5 270 263	10 800.0		5 950	**5 710.4**	4.3	1050	10.8	4.5	**0.1%**
7	15 275 776	10 800.0		9 833	10 800.0		1524	30.8	13.2	**TLE**
8	5 851 726	10 800.0		5 152	10 800.0		1424	85.7	6.2	**TLE**
9	7 080	**0.5**	9.2	204	0.6	9.2	72	0.1	9.5	**0.2%**
10	4 252	**0.1**	9.0	167	0.4	9.0	56	0.0	6.0	**5.2%**
11	1 493 713	10 800.0		2 948	**6 049.8**	14.2	454	0.9	12.0	**2.1%**
12	3 232 176	10 800.0		4 786	10 800.0		888	3.7	9.1	**TLE**
13	10 021	0.5	11.0	254	0.5	11.0	78	0.0		**INF**
14	14 620	**0.1**	12.7	388	0.6	12.7	220	0.3	12.7	0.0%
15	10 087	0.7	13.3	248	**0.3**	13.3	82	0.0	13.3	0.0%
16	688 327	**1486.8**	13.9	2 116	10 800.0	13.9	374	0.7	13.3	**0.2%**
17	870 865	10 800.0		2 026	**42.1**	14.1	516	1.5	10.1	**0.9%**
18	37 246	**0.5**	12.6	720	1.6	12.7	512	0.9	12.7	0.0%
19	123 368	40.9	12.5	1 050	**7.1**	12.5	744	584.8	12.5	0.0%
20	3 027 238	10 800.0		4 966	10 800.0		2360	315.4	10.8	**TLE**
21	13 304	0.1	0.0	252	**0.0**	0.0	44	0.0	0.0	0.0%
22	4 981	**0.0**	10.8	239	0.1	10.8	155	0.0	10.8	0.0%
23	43 120	8.00	14.1	594	**3.9**	14.1	330	1.1	14.1	0.0%
24	483 880	10 800.0		1 891	**286.9**	12.4	780	3.1	7.0	**2.4%**
25	23 354 590	10 800.0		11 267	10 800.0		2925	329.2	9.3	**TLE**
26	39 293	**4.1**	14.7	581	4.5	14.7	325	0.6	14.7	0.0%
27	7 080	0.5	14.9	240	**0.5**	14.9	128	0.4	14.9	0.0%
28	371 749	10 800.0		1 646	**59.4**	14.3	715	9.5	14.3	0.0%
29	3 718	**0.1**	13.6	180	0.6	13.6	84	0.0	13.6	0.0%
30	10 021	1.3	13.2	265	**0.4**	13.2	117	0.2	13.2	**1.4%**
31	58 226	3.5	14.0	612	**1.4**	14.0	296	0.1	13.1	**14.4%**

The benchmark instances used in the evaluation represent real-world regions of Portugal. Hence, it is possible to draw an electoral map of Portugal under a parallel voting system. As a result, Fig. 4 presents the complete electoral map for continental Portugal using the shortest-path formulation. Instance 13 is unsatisfiable using this formulation at the municipality level. However, a solution can be easily obtained at the parish level and is the one presented in Fig. 4. Note that all the electoral districts are contiguous, meaning that a color change also represents a change in electoral district. Overall, there are no major compactness issues (such as the ones in Fig. 1), thus showing the capabilities of the proposed objective function for compactness, as well as the formulations proposed in this paper.

4.3. Gerrymandering electoral maps

Since the establishment of democratic elections in Portugal in the 1970s, two major political parties have dominated the political spectrum: the Socialist Party (PS) and the Social Democratic Party (PSD). Hence, the drawing of electoral maps is bound to be constrained by people from these two parties that might decide to use other criteria than compactness. This section considers scenarios where a political party tries to gerrymander the electoral maps to its favor based on previous electoral results and evaluate the overall impact on the number of elected officials for parliament. Therefore, we focus our analysis on identifying the worst-case scenarios of gerrymandering. This type of analysis has been the subject of several works in the literature (Apollonio et al., 2009), as well as the issue of fairness (Balinski, 2008).

In order to maximize the results of a party, for each instance in Table 3, two MOCO formulations are created to find maps favoring each of the main Portuguese parties (one that maximizes PS and another that maximizes the results for PSD) as well as maximizing compactness using the tree-based formulation proposed in Section 3. Hence, in this section, we consider the MOCO formulation that results from the union of Formulations 4 and 5. Observe that it is usually not possible to maximize both the compactness of electoral districts and the expected results of a given party. Hence, these new formulations correspond to multi-objective combinatorial optimization problems for which one tries to generate the respective Pareto front. Since these are hard constrained problem instances, we use the `sat4jMoco` solver to generate the Pareto frontier, or a good approximation when the Pareto front cannot be identified within the time limit.

To instantiate the MOCO formulation for gerrymandering, it is necessary to have some estimation of the party results in each territorial unit. In this paper, we assume the results from the last parliament elections. It is clear that this is a weak estimation, and a data science analysis could be performed on previous results (including local elections). However, in the context of this paper, our goal is solely to evaluate the proposed formulations and check the variance of gerrymandering obtained with this initial data.

The districting was done independently for all Portuguese regions. Therefore, despite a complete map not being presented, the full electoral results (number of seats won by each party) with the extreme solutions for each objective are shown in Table 7. Observe that the Pareto front was not completely identified in some instances, and these cases are highlighted.

Note that in our results, given that the associated colors of parties PS and PSD are, respectively, pink and orange, districts colored in shades of pink are wins for party PS and colored in shades of orange are wins for party PSD.

Fig. 5 shows the 3 possible redistricting options for the region of Viseu (number 18). The left map in Fig. 5 is the most compact solution, and on this map, the total border length between territorial units in the same district is 606 840 m. Thus, the 4 electoral districts are split evenly

Table 6
Comparison between the simulated annealing method with our proposed Shortest-path (heuristic) and Tree-based (exact) formulations. For each formulation we present the value of the best solution found. The best solutions found for each instance are highlighted in bold. NA is presented when the method was unable to provide a feasible solution.

Ins.	Simulated annealing	Shortest-path formulation	Tree-based formulation
1	**168 101**	**168 101**	**168 101**
2	**301 448**	**301 448**	**301 448**
3	419 406	**435 990**	**435 990**
4	512 434	536 587	**540 876**
5	**730 646**	**730 646**	**730 646**
6	NA	630 784	**631 160**
7	NA	**1 119 906**	NA
8	NA	**597 844**	NA
9	**516 866**	515 705	**516 866**
10	**485 633**	461 863	**485 633**
11	683 354	778 081	**794 202**
12	NA	**1 146 290**	NA
13	**603 826**	NA	**603 826**
14	**301 490**	**301 490**	**301 490**
15	**493 959**	**493 959**	**493 959**
16	564 352	682 660	**684 252**
17	NA	691 249	**697 371**
18	**99 566**	**99 566**	**99 566**
19	NA	**78 380**	**78 380**
20	NA	**1 037 203**	NA
21	**719 949**	**719 949**	**719 949**
22	NA	**22 415**	**22 415**
23	NA	**110 666**	**110 666**
24	NA	283 514	**290 208**
25	NA	**1 181 490**	NA
26	454 028	**486 197**	**486 197**
27	**81 445**	**81 445**	**81 445**
28	NA	**919 352**	**919 352**
29	**189 934**	**189 934**	**189 934**
30	**439 980**	433 785	**439 980**
31	582 972	530 506	**606 840**

between the two parties. The center map is the most favorable towards party PS with an objective value of 521 578 m, a 14.05% decrease in compactness, which is clearly noticeable by the electoral districts' thin elongated shape, a symptom of gerrymandering. The decrease in compactness is rewarded with one more district win. Finally, the map on the right is biased towards party PSD. The objective value is 604 744 m (only a 0.35% decrease), showing how small changes in electoral maps can impact the electoral results (1 more win for PSD) without noticeable compactness problems.

For the region of Vila Real (number 17) the results are presented in Fig. 6. The most compact map (left) has the orange party winning 2 districts and a total compactness measure of 439 980 m. If we are also trying to maximize the wins of the pink party (center map) and compactness, it is possible to make PS win 2 out of the 3 districts. Still, the border length between territorial units in the same district goes down to 410 297 meters (a 6.75% decrease). On the other hand, the map on the right shows a redistricting option where PSD would win all the seats. Still, we can immediately notice that one of the districts has an extremely odd shape and is barely contiguous. This is confirmed by our compactness measure now being only 346 418 meters (a whopping 21.27% decrease in compactness to the most compact districting).

Fig. 7 shows the results for the region of Aveiro (number 01), which is one of the cases where redistricting at the municipality level is not possible. The map on the left presents the most compact distribution, where PS wins 5 districts while PSD wins the remaining 3. In the center map, we show the solution that maximizes the electoral district wins of PS while also maximizing compactness, and it is now possible to make PS win all but one district by cracking the opposition voters. Finally, the map on the right is the one that maximizes the wins of the PSD along with compactness, and the results show that it is possible to make PSD win 1 more electoral district bringing its total to 4 of the 8 single-member districts. In the electoral maps of Fig. 7, the changes are not so pronounced. Therefore the compactness is not significantly impacted.

Fig. 4. Complete electoral map for continental Portugal under a parallel voting system using the shortest-path formulation.

However, there is a 1.50% decrease in compactness when maximizing the results of PS and only 0.12% if we decide to maximize the results of PSD. Notwithstanding, it shows that using gerrymandering, it is possible to significantly change the number of seats in parliament by drawing the electoral districts in a certain way. We show that gerrymandering the area can help a party clearly win the elections in a particular district with 7 out of 8 seats (87.5%) or make it even between the two main parties.

In the 2019 elections, party PS received 36.65% of the national vote while PSD received 27.90%, meaning that, at the national level, PS is expected to be the party with the most parliament members under a first-past-the-post system. The complete electoral results confirm this prediction at the national level, which are presented in Table 7. The results show that by optimizing only for compactness, PS would win 86 out of 112 single-member districts (75%) and PSD the remaining 26 seats (25%). Using gerrymandering, in the most favorable distribution towards PS, it is possible to make it win 11 more seats, bringing its total to 97 (86.61%). Contrarily, under an unfavorable distribution, it can also lose 9 seats to the opposing party leaving it with 77 members in parliament out of the 112 (68.75%) awarded using the first-past-the-post voting system. Also presented in Table 7 is the consequent compactness decrease of gerrymandering an area. From these values, along with the previous Figures, we conclude that while sometimes gerrymandering is visually noticeable, subtle changes can be made that only slightly decrease the compactness but can have a significant impact on the electoral results

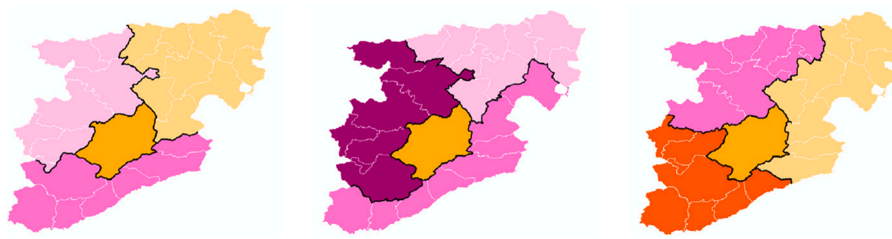

Fig. 5. Possible results for the region of Viseu (number 18) at the municipality level. In pink electoral districts won by party PS, in orange electoral districts won by party PSD. The left map is the distribution that only maximizes compactness. In the center, the most favorable districting towards PS and the one on the right the most biased towards PSD.

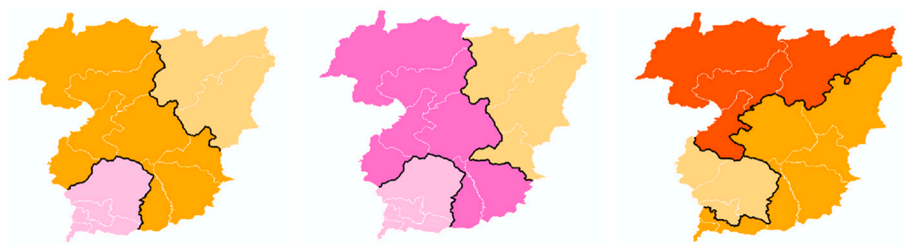

Fig. 6. Possible results for the region of Vila Real (number 17). In pink electoral districts won by party PS and in orange electoral districts won by party PSD. The left map is the distribution that only maximizes compactness. In the center, the most favorable districting towards PS and the one on the right the most biased towards PSD.

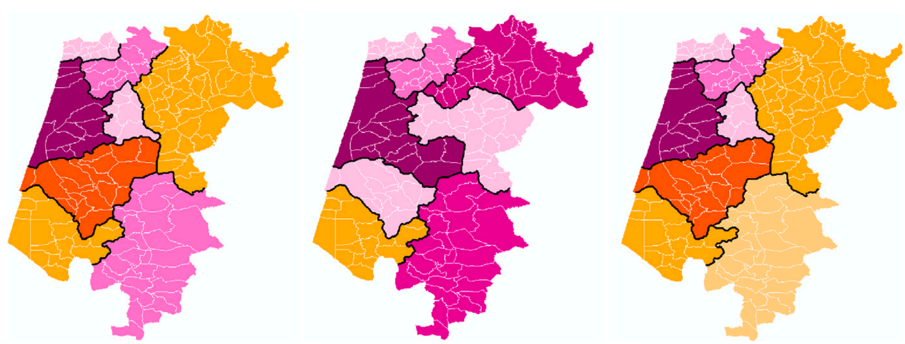

Fig. 7. Possible results for the region of Aveiro (number 01) at the parish level with a first division at the municipality level. In pink electoral districts won by party PS and in orange electoral districts won by party PSD. The map on the left only maximizes compactness. In the center, the most favorable districting towards PS and on the right the most biased towards PSD.

5. Conclusion and future work

Recent developments in hardware and algorithms to solve optimization problems increased the number of instances that can be solved using exact methods. In this work, novel and efficient Boolean formulations for electoral districting are proposed and applied to the Portuguese context. In particular, two different formulations to create contiguous districts are presented, an exact one (capable of finding global optimal solutions) and a heuristic but dramatically faster version. The heuristic version justifies its use in the most difficult to solve instances. Our experimental results on drawing electoral districts for the Portuguese parliament elections show that it can find good solutions in cases where complete methods sometimes fail even to produce a feasible solution using CPLEX. Moreover, we also propose optimizations to the base formulations that can also be adapted to classical approaches to the redistricting problem and manage to drastically improve its computational performance.

Following the propositions from the main parties to change the Portuguese electoral system to one using single-member districts, the formulations are tested in different scenarios using official geographical data. The results show that the proposed formulations can deliver compact and contiguous solutions within the population boundaries while conforming with current administrative divisions. Furthermore, an algorithm to redistrict large densely populated areas by making a first partition using existing administrative divisions is proposed, and that helps to find better solutions in several regions.

Additionally, our formulations can also be extended with gerrymandering techniques using official electoral results from previous parliament elections. Experimental results are shown for drawing electoral districts in Portugal, and the impact of gerrymandering is studied. In particular, we show that maximizing the number of electoral district wins of political parties requires a slight decrease in compactness in the Portuguese context.

The problem of redistricting electoral areas in the United States of America (USA) is where most research work has been conducted in the area. This occurs for several reasons: (i) there are more district-based elections than in other countries, (ii) the problem instances are usually larger (a major factor in the problem complexity), and (iii) several high-profile gerrymandering cases have increased public awareness to this issue. As future work, we propose to extend our experimental results by applying our formulations to this particular problem for both state and federal elections.

Furthermore, in the United States of America, where electoral maps were proven to have been gerrymandered (Fig. 1), we propose to use the gerrymandering capacities of our formulation and compare the results with those electoral maps. For future work, we will try to replicate or even further skew the results of the House of Representatives elections in several states.

Table 7
Complete electoral results for the districting in Section 4.3. The cells highlighted in yellow represent instances where the time limit was reached. Hence, the presented results are for the best solution found after 3 h (which might not be the global optimum). Whenever an instance can be gerrymandered the consequent compactness decrease is also presented.

Electoral results - Section 4.3								
Instance number	Max compact		Max PS			Max PSD		
	PS	PSD	PS	PSD	Compact. Decrease	PS	PSD	Compact. Decrease
1	2	0	2	0	0.00%	2	0	0.00%
2	1	1	2	0	23.70%	1	1	0.00%
3	1	1	1	1	0.00%	0	2	0.40%
4	1	1	2	0	2.67%	1	1	0.00%
5	2	0	2	0	0.00%	2	0	0.00%
6	1	2	2	1	1.19%	1	2	0.00%
7	3	0	3	0	0.00%	3	0	0.00%
8	1	3	2	2	3.47%	0	4	3.83%
9	0	2	0	2	0.00%	0	2	0.00%
10	2	0	2	0	0.00%	2	0	0.00%
11	2	0	2	0	0.00%	2	0	0.00%
12	3	0	3	0	0.00%	3	0	0.00%
13	2	0	2	0	0.00%	2	0	0.00%
14	5	0	5	0	0.00%	5	0	0.00%
15	2	0	2	0	0.00%	1	1	10.13%
16	1	1	1	1	0.00%	0	2	4.11%
17	2	1	3	0	13.24%	1	2	2.95%
18	7	1	7	1	0.00%	7	1	0.00%
19	7	1	8	0	0.00%	6	2	10.21%
20	8	0	8	0	0.00%	8	0	0.00%
21	1	0	1	0	0.00%	1	0	0.00%
22	3	2	3	2	0.00%	3	2	0.00%
23	5	0	5	0	0.00%	5	0	0.00%
24	3	2	4	1	0.30%	2	3	1.82%
25	4	1	5	0	0.90%	4	1	0.00%
26	4	1	4	1	0.00%	4	1	0.00%
27	4	0	4	0	0.00%	4	0	0.00%
28	5	0	5	0	0.00%	5	0	0.00%
29	1	2	2	1	18.27%	1	2	0.00%
30	1	2	2	1	6.75%	0	3	21.27%
31	2	2	3	1	14.05%	1	3	0.35%
TOTAL	86	26	97	15		77	35	

Finally, given the ability of our heuristic formulation to quickly generate contiguous electoral districts close to an optimal solution, we also propose to compare our shortest-path based formulation with other heuristic methods in several scenarios.

CRediT authorship contribution statement

Tiago Almeida: Data curation, Software, Visualization, Investigation, Writing – original draft, Writing – review & editing. **Vasco Manquinho:** Conceptualization, Methodology, Resources, Supervision, Project administration, Funding acquisition, Writing – original draft, Writing – review & editing.

Acknowledgments

This work was supported by Portuguese national funds through FCT, Fundação para a Ciência e a Tecnologia, under projects UIDB/50021/2020, PTDC/CCI-COM/31198/2017 and DSAIPA/AI/0044/2018.

References

Aloul, F.A., Sakallah, K.A., Markov, I.L., 2006. Efficient symmetry breaking for boolean satisfiability. IEEE Trans. Comput. 55 (5), 549–558. http://dx.doi.org/10.1109/TC.2006.75.

Altman, M., 1997. Is automation the answer: the computational complexity of automated redistricting. Rutgers Comput. Law Technol. J. 23, 81–142, URL http://maltman.hmdc.harvard.edu/pubpapers.shtml.

Apollonio, N., Becker, R.I., Lari, I., Ricca, F., Simeone, B., 2009. Bicolored graph partitioning, or: gerrymandering at its worst. Discrete Appl. Math. 157 (17), 3601–3614. http://dx.doi.org/10.1016/j.dam.2009.06.016.

Bação, F., Lobo, V.S., Painho, M., 2005. Applying genetic algorithms to zone design. Soft Comput. 9 (5), 341–348. http://dx.doi.org/10.1007/s00500-004-0413-4.

Balinski, M., 2008. Fair majority voting (or how to eliminate gerrymandering). Amer. Math. Monthly 115 (2), 97–113. http://dx.doi.org/10.1080/00029890.2008.11920503.

Bartholdi III, J.J., Tovey, C.A., Trick, M.A., 1992. How hard is it to control an election? Math. Comput. Modelling 16 (8–9), 27–40. http://dx.doi.org/10.1016/0895-7177(92)90085-Y.

Betzler, N., Uhlmann, J., 2009. Parameterized complexity of candidate control in elections and related digraph problems. Theoret. Comput. Sci. 410 (52), 5425–5442. http://dx.doi.org/10.1016/j.tcs.2009.05.029.

Bodin, L., 2006. A districting experiment with a clustering algorithm. Ann. New York Acad. Sci. 219, 209–214. http://dx.doi.org/10.1111/j.1749-6632.1973.tb41400.x.

Bottman, N., Essig, W., Whittle, S., 2007. Why weight? A cluster-theoretic approach to political districting. UMAP J. 28 (3).

Bozkaya, B., Erkut, E., Laporte, G., 2003. A tabu search heuristic and adaptive memory procedure for political districting. European J. Oper. Res. 144 (1), 12–26. http://dx.doi.org/10.1016/S0377-2217(01)00380-0.

Browdy, M.H., 1990. Simulated annealing: An improved computer model for political redistricting. Yale Law Policy Rev. 8 (1), 163–179, URL http://www.jstor.org/stable/40239326.

Carvajal, R., Constantino, M., Goycoolea, M., Vielma, J.P., Weintraub, A., 2013. Imposing connectivity constraints in forest planning models. Oper. Res. 61 (4), 824–836. http://dx.doi.org/10.1287/opre.2013.1183.

Constantino, M., Martins, I., Borges, J.G., 2008. A new mixed-integer programming model for harvest scheduling subject to maximum area restrictions. Oper. Res. 56 (3), 542–551. http://dx.doi.org/10.1287/opre.1070.0472.

Cormen, T.H., Leiserson, C.E., Rivest, R.L., Stein, C., 2001. Introduction to Algorithms, Second Edition. The MIT Press and McGraw-Hill Book Company.

Cox, G.W., 1997. Making Votes Count: Strategic Coordination in the World's Electoral Systems. Cambridge University Press.

D'Amico, S.J., Wang, S., Batta, R., Rump, C.M., 2002. A simulated annealing approach to police district design. Comput. OR 29 (6), 667–684. http://dx.doi.org/10.1016/S0305-0548(01)00056-9.

Datta, D., Figueira, J., Gourtani, A., Morton, A., 2013. Optimal administrative geographies: an algorithmic approach. Socio-Economic Plan. Sci. 47 (3), 247–257. http://dx.doi.org/10.1016/j.seps.2013.03.002.

Datta, D., Malczewski, J., Figueira, J.R., 2012. Spatial aggregation and compactness of census areas with a multiobjective genetic algorithm: a case study in Canada. Environ. Plan. B: Plann. Des. 39 (2), 376–392. http://dx.doi.org/10.1068/b38078.

DeFord, D.R., Duchin, M., Solomon, J., 2019. Recombination: A family of Markov chains for redistricting. CoRR, abs/1911.05725. URL arXiv:1911.05725.

D'Hondt, V., 1882. Système Pratique et Raisonné de Représentation Proportionnelle. librairie C. Muquardt, Merzbach et Falk, éditeurs, http://dx.doi.org/10.3931/e-rara-39876.

Duchin, M., Tenner, B.E., 2018. Discrete geometry for electoral geography. arXiv:1808.05860.

Dugosija, D., Savic, A., Maksimovic, Z., 2020. A new integer linear programming formulation for the problem of political districting. Ann. Oper. Res. 288 (1), 247–263. http://dx.doi.org/10.1007/s10479-020-03559-y.

Duque, J.C., Church, R.L., Middleton, R.S., 2011. The p-regions problem.. Geograph. Anal. 43 (1), 104–126.

Erdélyi, G., Fellows, M.R., Rothe, J., Schend, L., 2015a. Control complexity in bucklin and fallback voting: A theoretical analysis. J. Comput. System Sci. 81 (4), 632–660. http://dx.doi.org/10.1016/j.jcss.2014.11.002.

Erdélyi, G., Hemaspaandra, E., Hemaspaandra, L.A., 2015b. More natural models of electoral control by partition. In: Walsh, T. (Ed.), Proceedings of the 4th International Conference Algorithmic Decision Theory ADT. In: Lecture Notes in Computer Science, vol. 9346, Springer, pp. 396–413. http://dx.doi.org/10.1007/978-3-319-23114-3_24.

Faliszewski, P., Hemaspaandra, E., Hemaspaandra, L.A., Rothe, J., 2009. Llull and copeland voting computationally resist bribery and constructive control. J. Artificial Intelligence Res. 35, 275–341. http://dx.doi.org/10.1613/jair.2697.

Farrell, D.M., 2011. Electoral Systems: A Comparative Introduction. Macmillan International Higher Education.

Fifield, B., Higgins, M., Imai, K., Tarr, A., 2020. Automated redistricting simulation using Markov chain Monte Carlo. J. Comput. Graph. Statist. 29 (4), 715–728. http://dx.doi.org/10.1080/10618600.2020.1739532.

Garfinkel, R.S., Nemhauser, G.L., 1970. Optimal political districting by implicit enumeration techniques. Manage. Sci. 16 (8), B495–B508, URL http://www.jstor.org/stable/2628656.

Guest, O., Kanayet, F.J., Love, B.C., 2019. Gerrymandering and computational redistricting. J. Comput. Soc. Sci. 2 (2), 119–131. http://dx.doi.org/10.1007/s42001-019-00053-9.

Hayes, B., 1996. Machine politics. Am. Sci. 84 (6), 522–526.

Hemaspaandra, E., Hemaspaandra, L.A., Rothe, J., 2007. Anyone but him: The complexity of precluding an alternative. Artificial Intelligence 171 (5–6), 255–285. http://dx.doi.org/10.1016/j.artint.2007.01.005.

Herschlag, G., Kang, H.S., Luo, J., Graves, C.V., Bangia, S., Ravier, R., Mattingly, J.C., 2020. Quantifying gerrymandering in north carolina. Stat. Public Policy 7 (1), 30–38. http://dx.doi.org/10.1080/2330443X.2020.1796400.

Herschlag, G., Ravier, R., Mattingly, J.C., 2017. Evaluating partisan gerrymandering in wisconsin. arXiv:1709.01596.

Hess, S.W., Weaver, J.B., Siegfeldt, H.J., Whelan, J.N., Zitlau, P.A., 1965. Nonpartisan political redistricting by computer. Oper. Res. 13 (6), 998–1006, URL http://www.jstor.org/stable/167658.

Joshi, D., Soh, L., Samal, A., 2012. Redistricting using constrained polygonal clustering. IEEE Trans. Knowl. Data Eng. 24 (11), 2065–2079. http://dx.doi.org/10.1109/TKDE.2011.140.

Kotthoff, L., O'Sullivan, B., Ravi, S.S., Davidson, I., 2015. Complex clustering using constraint programming: Modelling electoral map creation. In: 14th International Workshop on Constraint Modelling and Reformulation.

Lewis, J.B., DeVine, B., Pitcher, L., Martis, K.C., Esri, 2014. United States congressional districts in North Carolina, 2013–2017, URL https://commons.wikimedia.org/wiki/File:United_States_Congressional_Districts_in_North_Carolina,_since_2013.tif.

Liu, Y.Y., Cho, W.K.T., Wang, S., 2016. PEAR: a massively parallel evolutionary computation approach for political redistricting optimization and analysis. Swarm Evol. Comput. 30, 78–92. http://dx.doi.org/10.1016/j.swevo.2016.04.004, URL https://doi.org/10.1016/j.swevo.2016.04.004.

Liu, H., Feng, H., Zhu, D., Luan, J., 2009. Parameterized computational complexity of control problems in voting systems. Theoret. Comput. Sci. 410 (27–29), 2746–2753. http://dx.doi.org/10.1016/j.tcs.2009.04.004.

Lublin, D., 1999. The Paradox of Representation: Racial Gerrymandering and Minority Interests in Congress. Princeton University Press, http://dx.doi.org/10.2307/j.ctv173f1kk.

Macmillan, W., 2001. Redistricting in a GIS environment: An optimisation algorithm using switching-points. J. Geogr. Syst. 3 (2), 167–180. http://dx.doi.org/10.1007/PL00011473.

Marques-Silva, J., Lynce, I., Manquinho, V., 2008. Symmetry breaking for maximum satisfiability. In: Cervesato, I., Veith, H., Voronkov, A. (Eds.), Logic for Programming, Artificial Intelligence, and Reasoning, 15th International Conference, LPAR 2008, Doha, Qatar, November 22-27, 2008. Proceedings. In: Lecture Notes in Computer Science, vol. 5330, Springer, pp. 1–15. http://dx.doi.org/10.1007/978-3-540-89439-1_1.

McDill, M.E., Rebain, S.A., Braze, J., 2002. Harvest scheduling with area-based adjacency constraints. Forest Sci. 48 (4), 631–642. http://dx.doi.org/10.1093/forestscience/48.4.631.

Mehrotra, A., Johnson, E.L., Nemhauser, G.L., 1998. An optimization based heuristic for political districting. Manage. Sci. 44 (8), 1100–1114. http://dx.doi.org/10.1287/mnsc.44.8.1100.

Murray, A.T., 1999. Spatial restrictions in harvest scheduling. Forest Sci. 45 (1), 45–52. http://dx.doi.org/10.1093/forestscience/45.1.45.

Nemoto, T., Hotta, K., 2003. Modelling and solution of the problem of optimal electoral districting. Commun. OR Soc. Japan 48, 300–306.

Niemi, R.G., Grofman, B., Carlucci, C., Hofeller, T., 1990. Measuring compactness and the role of a compactness standard in a test for partisan and racial gerrymandering. J. Politics 52 (4), 1155–1181. http://dx.doi.org/10.2307/2131686.

Oehrlein, J., Haunert, J.-H., 2017. A cutting-plane method for contiguity-constrained spatial aggregation. J. Spatial Inform. Sci. 2017, http://dx.doi.org/10.5311/JOSIS.2017.15.379.

Önal, H., Briers, R.A., 2006. Optimal selection of a connected reserve network. Oper. Res. 54 (2), 379–388. http://dx.doi.org/10.1287/opre.1060.0272.

Önal, H., Wang, Y., Dissanayake, S.T.M., Westervelt, J.D., 2016. Optimal design of compact and functionally contiguous conservation management areas. European J. Oper. Res. 251 (3), 957–968. http://dx.doi.org/10.1016/j.ejor.2015.12.005.

Orvalho, P., Terra-Neves, M., Ventura, M., Martins, R., Manquinho, V.M., 2019. Encodings for enumeration-based program synthesis. In: Schiex, T., de Givry, S. (Eds.), Principles and Practice of Constraint Programming - 25th International Conference, CP 2019, Stamford, CT, USA, September 30 - October 4, 2019, Proceedings. In: Lecture Notes in Computer Science, vol. 11802, Springer, pp. 583–599. http://dx.doi.org/10.1007/978-3-030-30048-7_34.

Polsby, D.D., Popper, R.D., 1991. The third criterion: Compactness as a procedural safeguard against partisan gerrymandering. Yale Law Policy Rev. 9 (2), 301–353, URL http://www.jstor.org/stable/40239359.

Reock, E.C., 1961. A note: Measuring compactness as a requirement of legislative apportionment. Midwest J. Polit. Sci. 5 (1), 70–74. http://dx.doi.org/10.2307/2109043.

Ricca, F., Scozzari, A., Simeone, B., 2008. Weighted voronoi region algorithms for political districting. Math. Comput. Model. 48 (9–10), 1468–1477. http://dx.doi.org/10.1016/j.mcm.2008.05.041.

Ricca, F., Scozzari, A., Simeone, B., 2013. Political districting: from classical models to recent approaches. Ann. Oper. Res. 204 (1), 271–299. http://dx.doi.org/10.1007/s10479-012-1267-2.

Ricca, F., Simeone, B., 1997. Political districting: Traps, criteria, algorithms and tradeoffs. Ricerca Oper. AIRO 27, 81–119.

Riker, W.H., 1982. The two-party system and duverger's law: An essay on the history of political science. Amer. Polit. Sci. Rev. 76 (4), 753–766. http://dx.doi.org/10.1017/S0003055400189580.

Sakallah, K.A., 2009. Symmetry and satisfiability. In: Biere, A., Heule, M., van Maaren, H., Walsh, T. (Eds.), Handbook of Satisfiability. In: Frontiers in Artificial Intelligence and Applications, vol. 185, IOS Press, pp. 289–338. http://dx.doi.org/10.3233/978-1-58603-929-5-289.

Schwartzberg, J.E., 1966. Reapportionment, gerrymanders, and the notion of compactness. Minnesota Law Rev. 50 (1701), 443, URL https://scholarship.law.umn.edu/mlr/1701/.

Shirabe, T., 2009. Districting modeling with exact contiguity constraints. Environ. Plan. B: Plann. Des. 36 (6), 1053–1066.

Svec, L., Burden, S., Dilley, A., 2007. Applying voronoi diagrams to the redistricting problem. UMAP J. 28 (3), 313–329.

Swamy, R., King, D., Jacobson, S., 2019. Multi-objective optimization for political districting: A scalable multilevel approach. Optim. Online.

Tam Cho, W.K., Liu, Y.Y., 2016. Toward a talismanic redistricting tool: A computational method for identifying extreme redistricting plans. Election Law J. 15 (4), 351–366.

Tavares-Pereira, F., Figueira, J.R., Mousseau, V., Roy, B., 2007. Multiple criteria districting problems. Ann. Oper. Res. 154 (1), 69–92. http://dx.doi.org/10.1007/s10479-007-0181-5.

Validi, H., Buchanan, A., Lykhovyd, E., 2020. Imposing contiguity constraints in political districting models. Optim. Online.

Vanneschi, L., Henriques, R., Castelli, M., 2017. Multi-objective genetic algorithm with variable neighbourhood search for the electoral redistricting problem. Swarm Evol. Comput. 36, 37–51. http://dx.doi.org/10.1016/j.swevo.2017.04.003.

Vickrey, W., 1961. On the prevention of gerrymandering. Polit. Sci. Q. 76 (1), 105–110, URL http://www.jstor.org/stable/2145973.

Young, H.P., 1988. Measuring the compactness of legislative districts. Legislative Stud. Q. 13 (1), 105–115. http://dx.doi.org/10.2307/439947.