CHAPTER 3

A Guided Tour of the Mathematics of Seat Allocation and Political Districting

Federica Ricca, Andrea Scozzari, and Paolo Serafini

3.1 Introduction

This chapter focuses on Seat Allocation and Political Districting, two of the main topics in the study of electoral systems. Models and algorithms from discrete mathematics and combinatorial optimization are used to formalize the problems and find solutions that meet some fairness requirements. The first problem concerns the assignment of seats to parties in political elections. In particular, we discuss the well-known Biproportional Apportionment Problem (BAP), that is, the problem of assigning the House seats in those countries that adopt a two-level proportional system. The problem is difficult also from a mathematical viewpoint, since it combines a matrix feasibility problem with the requirement of double proportionality. The second topic, Political Districting (PD), is a territorial problem in which electoral districts must be designed so that each voter is univocally assigned to one district. This is a relevant problem, since, given the same vote outcome of an election, depending on the district shape and size, the final seat allocation to parties could be drastically different. For this reason, PD procedures have been proposed to output district maps that meet a set of criteria aimed at avoiding district manipulation by parties.

Both BAP and PD are extensively studied in the literature, the first one starting from the seminal paper by Balinski and Demange (1989a,b), the second dating back to 1960's when the paper by Hess et al. (1965) formulated for the first time the problem as an optimization one. The chapter is organized in two parts, the first related to BAP, the second to PD.

3.2 Biproportional Apportionment Problem

3.2.1 Proportional Apportionments

Before describing the Biproportional Apportionment Problem it is necessary to briefly introduce the simpler Proportional Apportionment Problem in which the fixed number of seats of the House has to be divided among constituencies. A mathematically equivalent problem consists in dividing the seats of the House among parties. This second problem presents additional features of candidate selections which are clearly not present in the first problem. In this section we limit ourselves to outlining the main features of the first problem. We refer the reader to the monograph by Balinski and Young (2001) for a comprehensive review of apportionment problems.

Let H be the number of the seats of the House and let I be the set of constituencies, with $m = |I|$. Let p_i be the population of constituency i and let $P = \sum_{i \in I} p_i$ be the total population. In almost all nations the seats assigned to each constituency are required to be proportional to the populations, a notable exception being the European Parliament where the so-called degressive proportionality requirement is called for (see Grimmett (2012), Serafini (2012) and other papers in the same issue).

Ideally, exact proportionality would be obtained by assigning the number of seats $q_i := p_i H/P$ to constituency i, but q_i is in general a fractional number that must be rounded in some way. The question of how to round these numbers presents several subtle features and no univocal answer exists as the history of the US House of Representatives has shown (an interesting account can be found at the site `https://www.census.gov/history/www/reference/apportionment/`).

Perhaps the simplest method of rounding q_i is the Largest Remainder Rule, also known under the names of Hamilton, Vinton, Hare or Hare-Niemayer. First, to each constituency the number of seats $s_i := \lfloor q_i \rfloor$ is assigned. Then the remaining seats are assigned to those constituencies that have been most penalized by the rounding, namely the ones with largest remainders. It can be easily shown that this method finds the point in $\mathbb{R}^m$ with integral coordinates at minimum distance from the point $q \in \mathbb{R}^m$, where the distance can be measured with any norm.

In spite of the simplicity of the method and this important minimum norm property, the method is questionable for other reasons. First, it considers the absolute deviation while the relative deviation could be perceived more important. Second, it is prone to some anomalous behaviors, that are respectively known as the Alabama Paradox, the Population Paradox and the New State Paradox (Balinski and Young, 2001). For these reasons the method is avoided in many countries. In Italy the Largest Remainder Rule is stated in the Constitution.

The paradoxes are avoided by the divisor methods. A 'modern' way to present a divisor method is as follows. First, a *signpost* function is defined

$$\delta : \mathbb{Z} \to \mathbb{R}, \qquad \text{with} \qquad \delta(z) \in [z, z+1]$$

that assigns to each integer z a real number between z and $z+1$. The function $\delta(z)$ specifies how to round a real $a \in [z, z+1)$. The rounding, denoted as $[\![a]\!]$, is given by

$$[\![a]\!] = \begin{cases} \lfloor a \rfloor & \text{if } a \leq \delta(z) \\ \lceil a \rceil & \text{if } \delta(z) < a < z+1 \end{cases}$$

This definition implies $[\![a]\!] = \lfloor a \rfloor$ if $a = \delta(z)$. Actually, we have a tie since we might as well define $[\![a]\!] = \lceil a \rceil$ if $a = \delta(z)$. This ambiguity is exploited in the Tie-and-Transfer method for BAP as we shall see. In the Proportional Apportionment Problem the probability that $a = \delta(z)$ is almost negligible.

Then a multiplier λ is looked for such that the seats

$$s_i = [\![\lambda\, p_i]\!]$$

sum up to H. The crucial aspect of a divisor method is the choice of the signpost function. These are the choices that have been proposed and also implemented in some cases:

$\delta(z) = z$	Adams method
$\delta(z) = \dfrac{2}{\frac{1}{z} + \frac{1}{z+1}}$	Dean method
$\delta(z) = \sqrt{z\,(z+1)}$	Huntington-Hill method
$\delta(z) = z + 0.5$	Webster method
$\delta(z) = z + 1$	Jefferson or D'Hondt method

We just recall that the Adams method favors the small constituencies, while the opposite happens for the Jefferson method. The Huntington-Hill method is the one currently employed to apportion the seats of the US House of Representatives.

3.2.2 Biproportional Apportionment Problem: Introduction

A common feature of many parliaments is the presence of a house of representatives whose seats are not only *a priori* divided among constituencies but also, after the election, among the various competing lists. In these systems the vote assigned to a list is of primary importance and the choice of the actual representatives is done after having assigned the seats to the lists at national level. In other systems the seats assigned to a list are a consequence of the seats won by the candidates.

In this chapter we deal with the problem in which the seats allotted to each constituency are fixed, typically before the elections, the seats allotted to the lists are preliminarily computed on the basis of the votes received in the whole nation, and we have to compute the seats to assign to each list in each constituency. Clearly, we have to respect the previous seat assignments and try to have seats as much as possible proportional to the votes.

Formally, let m be the number of constituencies, H the total number of seats in the house, and R_i the seats allotted to constituency i (obviously $\sum_i R_i = H$). Let n be the number of lists. Let v_{ij} be the votes obtained by list j in constituency i. Let $V_j := \sum_i v_{ij}$ be the votes obtained by list j at national level and let $V := \sum_j V_j$ be the total number of votes. Let P_j be the total number of seats in the house assigned to list j (obviously $\sum_j P_j = H$). The computation of the numbers R_i (before the election) and the numbers P_j (after the election) is done by one of the methods seen in the previous section. Then we have to compute the seats s_{ij} to assign to list j in constituency i subject to:

1. $\sum_{i=1}^{m} s_{ij} = P_j$, for every list j;

2. $\sum_{j=1}^{n} s_{ij} = R_i$, for every constituency i;
3. If $v_{ij} = 0$ for some list j in some constituency i, then $s_{ij} = 0$;
4. The seats s_{ij} have to be "as proportional as possible" to the votes v_{ij}.

This is the so-called Biproportional Apportionment Problem. The first three requirements are clear. The crucial issue is the last requirement. Exact proportionality of the seats to both the lists and the constituencies cannot be achieved in general if we must satisfy requirements 1, 2 and 3. Therefore, we have to clearly define the goal we want to pursue. In addition we require integrality of the final outcome. The BAP is not a simple problem and one needs *ad hoc* mathematical tools to solve it.

Let us first note that the constraints 1, 2 and 3 are linear programming constraints whose underlying matrix is totally unimodular. Therefore, the feasible set of (where $E = \{(i, j) \in I \times J : v_{ij} > 0\}$)

$$\begin{aligned} \sum_{j:(i,j)\in E} x_{ij} &= R_i && i \in I \\ \sum_{i:(i,j)\in E} x_{ij} &= P_j && j \in J \\ x_{ij} &\geq 0 && (i,j) \in E \end{aligned} \tag{3.1}$$

is a polyhedron whose vertices have integral coordinates and therefore a seat apportionment can be found among its vertices. This is a fundamental property that allows to solve the BAP problem as a tractable linear programming problem. The property holds also if we bound each x_{ij} within an interval with integral extremes, i.e.,

$$l_{ij} \leq x_{ij} \leq u_{ij} \tag{3.2}$$

where l_{ij} and u_{ij} are integral. Hence the existence of a feasible fractional solution to (3.1) and (3.2) implies the existence of an integral solution to the same constraints.

Let us call *quotas* real numbers q_{ij} that would represent an 'ideal' seat apportionment if we were allowed to relax the integrality requirement and maybe also requirements 1 and 2. The definition of ideal is up to the lawmakers. For instance we might define as quotas the numbers $v_{ij} H/V$ that fully satisfy the proportionality requirement to both lists and constituencies, but they do not satisfy requirements 1 and 2.

In some nations (e.g., Italy and Belgium) the following quotas, called *regional quotas*, are used

$$q_{ij} = \frac{v_{ij}}{\sum_k v_{ik}} R_i$$

These quotas guarantee exact proportionality among lists within each constituency. By definition we have $\sum_j q_{ij} = R_i$, but in general $\sum_i q_{ij} = P_j$ does not hold and there is no proportionality among constituencies within each list.

It is possible to define quotas such that both requirements 1 and 2 are satisfied at the expense of losing exact proportionality. Such quotas are called *fair* *share quotas* and are defined in the next section. Let us note that a slight shift of votes in one constituency has the effect of propagating to the overall set of quotas if both sums must be satisfied. This may be not desirable if we want to preserve some form of autonomy among constituencies. For this reason the regional quotas, that are independent of each constituency, may be preferred. However, fair share quotas exhibit important mathematical properties and this is considered an important factor in favor of using the fair share quotas.

If the seats of an apportionment are obtained from the quotas by rounding each quota either up or down, we say that the apportionment *stays within the quotas*.

By and large there are two approaches to the BAP. In the first approach a set of axioms that every reasonable apportionment should satisfy is designed and then a method aimed at satisfying the axioms is looked for. Typically such a method is unique. This is the approach proposed by Balinski and Demange (1989a,b). The other approach consists in defining fractional ideal quotas and then finding a seat apportionment that minimizes some measure of deviation with respect to the ideal quotas (Ricca et al., 2012). A detailed comparison of the two approaches is discussed by Ricca et al. (2012) and we refer the reader to this paper for a more comprehensive understanding of the various issues.

3.2.3 Divisor Methods: Axioms

An apportionment method can be seen as a function S that maps the problem data, i.e., the vote matrix v_{ij} and the values H, R_i and P_j, into an integral non-negative matrix. It is convenient to denote the data as a pair (v, w) where v is the vote matrix and w is the set of numbers H, R_i and P_j. Then $S(v, w)$ is the particular matrix output by the apportionment method defined by the function S.

We may also relax the integrality requirement and consider fractional apportionments. In this case a fractional apportionment method can be seen as a function Q that maps (v, w) into a fractional non-negative matrix $Q(v, w)$.

Let us consider the following axioms that a fractional apportionment method Q for BAP should satisfy (Balinski and Demange, 1989b). Here $q = Q(v, w)$.

1. Exactness: if the v_{ij} satisfy $H \sum_j v_{ij} = R_i V$ and $H \sum_i v_{ij} = P_j V$, then $q = H\,v/V$.

2. Uniformity: let I be a subset of constituencies and J a subset of lists, and let v_{IJ} and q_{IJ} be the matrix restrictions to $I \times J$ of v and q, respectively. Moreover, let

$$\hat{R}_i := \sum_{j \in J} q_{ij}, \quad i \in I, \qquad \hat{P}_j := \sum_{i \in I} q_{ij}, \quad j \in J, \qquad \hat{H} = \sum_{i \in I} \sum_{j \in J} q_{ij}$$

These values define the data w_{IJ}. Then q_{IJ} must be an admissible apportionment output by Q if directly applied to the data (v_{IJ}, w_{IJ}).

3. Monotonicity: if v' and v are two vote matrices that are different only for one pair (h, k) where $v'_{hk} > v_{hk}$ and $q' = Q(v', w)$ then we must have $q'_{hk} \geq q_{hk}$.

4. Homogeneity: if two rows h and k of the vote matrix are proportional, i.e.,

$v_{hj} = \lambda\, v_{kj}$ for all j, and $R_h = R_k$, then the apportionment on the two rows must be the same, i.e., $q_{hj} = q_{kj}$ for all j. The same principle must hold for the columns.

We report here the axioms in a restricted framework with respect to Balinski and Demange (1989b), who consider R and P variable numbers within specified bounds. Another axiom (Relevance) is introduced by Balinski and Demange (1989b), that becomes void when R and P are fixed data.

It can be shown that Homogeneity and Uniformity imply together uniqueness of the apportionment. Uniqueness is clearly a necessary requirement for every apportionment method. We may invoke the same axioms also for an integral apportionment method. In addition a new axiom is introduced that calls for a 'continuity' property. To state this axiom we need to assume that the votes v_{ij} are real numbers. Then we require:

5. Completeness: let v^k be a sequence such that $v^k \to \bar{v}$ and let $s = S(v^k, w)$ for all k. Then $s = S(\bar{v}, w)$.

This axiom may be too restrictive if we allow zero votes for some pair (i, j). In this case it might happen that $v_{ij}^k > 0$, $v_{ij}^k \to 0$ and $s_{ij} = 1$. Since $S(\bar{v}, w)$ must output $s_{ij} = 0$, the axiom cannot be fulfilled. We may take the point of view that zero votes happen only because a certain list is not present in a particular constituency. In this case $v_{ij}^k = 0$ for any k.

The fundamental result by Balinski and Demange is that the unique fractional apportionment that satisfies the axioms 1–4 is a matrix F denoted *fair share* that can be expressed as

$$F_{ij} = \lambda_i\, v_{ij}\, \mu_j, \qquad i \in I,\ j \in J,$$

where $\lambda_i > 0$ and $\mu_j > 0$ are multipliers chosen to satisfy the constraints

$$\sum_{j \in J} F_{ij} = R_i, \quad i \in I, \qquad \sum_{i \in I} F_{ij} = P_j, \quad j \in J.$$

The existence of the fair share matrix is always granted if the vote matrix is strictly positive. If the vote matrix contains some zeros the fair share matrix might not exist as it happens in this simple example

$$v = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad R = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad P = (1 \quad 1)$$

Since $v_{ij} = 0$ implies $F_{ij} = 0$ the only matrix satisfying the sum constraint is the identity matrix. However, there are no positive multipliers such that $\lambda_1\, v_{12}\, \mu_2 = 0$.

An existence result even with some zero elements is provided by the following theorem (Bachem and Korte, 1979; Rothblum and Schneider, 1989; Kalantari et al., 2008).

Theorem 3.1. *A fair share matrix exists if and only if there exists a feasible solution to the constraints*

$$\sum_{j:(i,j) \in E} x_{ij} = R_i,\ i \in I, \qquad \sum_{i:(i,j) \in E} x_{ij} = P_j,\ j \in J, \quad x_{ij} \geq \frac{1}{|E|},\ (i,j) \in E. \tag{3.3}$$

The feasibility of (3.3) can be checked in polynomial time by standard network flow techniques. However, it is simpler to use the so-called RAS algorithm to compute F. This algorithm alternately scales rows and columns in order to satisfy in turn either row or column sum. Formally the following computation has to be carried out starting from the initial solution $F^0 = v$, $\lambda_i^0 = 1$ and $\mu_j^0 = 1$:

$$\begin{array}{lllll} \alpha_i := \dfrac{R_i}{\sum_j F_{ij}^k}, & \lambda_i^{k+1} = \alpha_i \, \lambda_i^k, & \bar{F}_{ij}^k = \alpha_i \, F_{ij}^k & j = 1, \dots, n, & i = 1, \dots m, \\ \beta_j := \dfrac{P_j}{\sum_i F_{ij}^k}, & \mu_j^{k+1} = \beta_j \, \mu_j^k, & F_{ij}^{k+1} = \beta_j \, \bar{F}_{ij}^k & i = 1, \dots, m, & j = 1, \dots n. \end{array}$$

One important property of the fair share matrix is that there always exists an apportionment obtained by rounding each matrix entry either down or up, i.e., it stays within the quotas. We have already observed that the existence of a feasible fractional solution to (3.1) and (3.2) implies the existence of an integral solution to the same constraints and F is feasible for (3.1) and (3.2) with $l_{ij} = \lfloor F_{ij} \rfloor$ and $u_{ij} = \lceil F_{ij} \rceil$.

3.2.4 Divisor Methods: TT and DAS Methods

Like in the Proportional Apportionment Problem, once a signpost function $\delta(z)$ is defined, we have to find multipliers λ_i and μ_j such that the seats obtained by

$$s_{ij} = [\![\lambda_i \, v_{ij} \, \mu_j]\!]$$

satisfy row and column sums. If we round the fair share matrix, it is unlikely that the sums are respected. Hence we have to find out other multipliers.

The Tie-and-Transfer method (TT) by Balinski and Demange (1989a) cleverly exploits the idea that if $\lambda_i \, v_{ij} \, \mu_j = \delta(\lfloor \lambda_i \, v_{ij} \, \mu_j \rfloor)$ then the matrix entry can be rounded either up or down because there is a tie. Hence the multipliers must be continuously updated in order to have a series of ties that allow a simultaneous transfer of seats in order to satisfy the sum constraints. Explaining in detail the TT method is beyond the scope of this short survey due to its many technical details. The reader is directed to the literature. The remarkable fact about the method is that it is polynomial and satisfies the axioms.

The Discrete Alternating Scaling Algorithm (DAS) by Pukelsheim (2004) is similar in the sense that it aims at finding multipliers λ_i and μ_j such that the rounding is consistent with the sum constraints. However, it differs in the way the multipliers are computed. Furthermore, the algorithm may stall, although with very low probability. However, its simplicity is an important pro toward a possible adoption and indeed it has been adopted in the Cantons of Zürich, Schaffhausen and Aargau (Switzerland) (Pukelsheim and Schuhmacher, 2004).

The DAS method works as the RAS algorithm for the computation of the fair share matrix. The only difference is that the sum constraint is enforced by using a divisor method applied to either the rows or to the columns in an alternate way.

In more detail let λ_i^k and μ_j^k be the multipliers obtained at the k-th step and let $q^k = \lambda_i^k \, v_{ij} \, \mu_j^k$. Starting with $\lambda_i^0 = 1$ and $\mu_j^0 = 1$ we iterate as

1. Let λ_i^k such that $s_{ij} = [\![\lambda_i^k \, q_{ij}^{k-1}]\!]$ and $\sum_j s_{ij} = R_i$. Compute $q_{ij}^k = \lambda_i^k \, q_{ij}^{k-1}$. If $\sum_i s_{ij} = P_j$ stop, otherwise $k := k+1$ and go to 2.

2. Let μ_j^k such that $s_{ij} = [\![q_{ij}^{k-1} \, \mu_j^k]\!]$ and $\sum_i s_{ij} = P_j$. Compute $q_{ij}^k = q_{ij}^{k-1} \, \mu_j^k$. If $\sum_j s_{ij} = R_i$ stop, otherwise $k := k+1$ and go to 1.

The final multipliers are given by

$$\lambda_i = \prod_k \lambda_i^k, \qquad \mu_j = \prod_k \mu_j^k.$$

3.2.5 Minimum Deviation Methods

Given ideal quotas the seats can be computed by finding those that minimize an appropriate measure of deviation from the quotas. Since there can be many different ways of measuring the deviation we may consequently define different apportionment methods. Which one to choose in practice is a decision of the lawmakers.

The important framework common to all methods is that the constraint matrix is (3.1) with the possible addition of (3.2) and therefore a linear objective function will always produce a seat apportionment. In particular, these problems can be cast as network flow problems for which fast and reliable algorithms are available.

One natural way of measuring the deviation of the computed seats s_{ij} from the ideal quotas q_{ij} considers an L_ρ-norm, so that the objective function is

$$\min_s \sum_{ij} |s_{ij} - q_{ij}|^\rho$$

The typical values for ρ are $\rho = 1$, $\rho = 2$ or $\rho = \infty$ (that corresponds to $\min_s \max_{ij} |s_{ij} - q_{ij}|$ and we speak of *minimax* solutions). In addition we may also require that the seats stay within the quotas. Other ways of measuring the deviation not directly linked to a norm may be also defined. For instance we may consider 'fair' a rounding of the quotas to the closest integer and 'unfair' to the second closest integer. If we want to find an apportionment within the quotas we necessarily round each entry in the table either fairly or unfairly. A possible objective could be the minimization of the number of unfair roundings. The apportionment found this way might be called a *Best Rounding apportionment*.

Since we want to model the problems as linear programming problems on the constraint set (3.1) with the possible addition of (3.2), the only modeling issue that remains to be solved is how to express the various minimizations as linear functions. For the L_1-norm we note that the function $f_{ij}(x) = |x - q_{ij}|$ is convex and piece-wise linear. The function

$$g_{ij}(x) = \begin{cases} q_{ij} - x & \text{if } x \le \lfloor q_{ij} \rfloor \\ (1 - 2\,\langle q_{ij} \rangle)\,(x - \lfloor q_{ij} \rfloor) + \langle q_{ij} \rangle & \text{if } \lfloor q_{ij} \rfloor \le x \le \lceil q_{ij} \rceil \\ x - q_{ij} & \text{if } x \ge \lceil q_{ij} \rceil. \end{cases} \tag{3.4}$$

(where $\langle a \rangle = a - \lfloor a \rfloor$ is the fractional part of a) is also convex and piece-wise linear with integral breakpoints. Furthermore, $g_{ij}(x) = f_{ij}(x)$ on the breakpoints of g.

Hence, if we minimize $\sum_{ij} g_{ij}(x_{ij})$ we obtain integral values for x. The function (3.4) can be turned into linear programming by expressing each x_{ij} as a sum of three additional variables

$$x_{ij} = \xi^1_{ij} + \xi^2_{ij} + \xi^3_{ij}$$

subject to $0 \le \xi^1_{ij} \le \lfloor q_{ij} \rfloor$, $0 \le \xi^2_{ij} \le 1$, $0 \le \xi^3_{ij}$ and having the following objective function

$$\min \sum_{(ij) \in E} q_{ij} - \xi^1_{ij} + (1 - 2\langle q_{ij} \rangle)\, \xi^2_{ij} + \langle q_{ij} \rangle + \xi^3_{ij} + 1 - \langle q_{ij} \rangle$$

that is equivalent up to a constant shift to

$$\min \sum_{(ij) \in E} -\xi^1_{ij} + (1 - 2\langle q_{ij} \rangle)\, \xi^2_{ij} + \xi^3_{ij}$$

The objective function coefficients are such that $\xi^2_{ij} > 0$ only if $\xi^1_{ij} = \lfloor q_{ij} \rfloor$ and $\xi^3_{ij} > 0$ only if $\xi^2_{ij} = 1$.

The same trick of substituting a piece-wise linear function with another one which has integral breakpoints and is equal to the first function on these breakpoints can work with any convex objective function. There is however a subtle theoretical issue that should not be neglected. For each breakpoint we have to introduce a new variable. Hence we should know if the number of breakpoints is polynomial. A trivial bound, based on the values R_i or P_j is only pseudo-polynomial.

Although in practice a naive implementation of this technique works well, because the vast majority of instances have optimal apportionments within the bounds $\lfloor q_{ij} \rfloor - 1$ and $\lceil q_{ij} \rceil + 1$, and therefore we do not need in practice more than five additional variables, yet we wonder whether exists a polynomial algorithm to solve the problem. The answer is affirmative thanks to a scaling procedure due to Minoux (1984).

For the L_2 norm we substitute the function $f_{ij}(x) = (x - q_{ij})^2$ with the function

$$g_{ij}(x) = (\lfloor x \rfloor - q_{ij})^2 + \langle x \rangle \; (1 + 2 (\lfloor x \rfloor - q_{ij}))$$

which can be linearized by introducing additional variables ξ^k_{ij} subject to

$$x_{ij} = \sum_k \xi^k_{ij}, \qquad 0 \le \xi^k_{ij} \le 1, \qquad k = 0, \ldots, \min\{P_j, R_i\}, \qquad (i,j) \in E$$

with objective function

$$\sum_{ij} \sum_k (1 + 2(k - q_{ij}))\, \xi^k_{ij}$$

Optimal apportionments for either norm L_1 or L_2 do not necessarily stay within the quotas. Counterexamples can be given (see, for instance, Ricca et al., 2012). If we want an apportionment within the quotas we simply add to (3.1) the constraints $\lfloor q_{ij} \rfloor \le x_{ij} \le \lceil q_{ij} \rceil$. In this case we can solve for any L_ρ-norm ($\rho < \infty$) by simply using the objective function (see Cox and Ernst, 1982)

$$\min \sum_{ij} \left((1 - \langle q_{ij} \rangle)^\rho - \langle q_{ij} \rangle^\rho \right) x_{ij}$$

The norms L_1 and L_2 tend to produce the same optimal apportionment. Indeed it is possible to prove the following result:

Theorem 3.2. *If an optimal apportionment with respect to the L_1-norm stays within the quotas, then the same apportionment is optimal with respect to the L_2-norm.*

If we have in mind a Best Rounding apportionment, we can count the unfair rounding and minimize this count by defining the sets

$$E^+ := \{(i,j) \in E : \langle q_{ij} \rangle < 0.5\}, \qquad E^- := \{(i,j) \in E : \langle q_{ij} \rangle > 0.5\},$$

and using the objective function

$$z = \min \sum_{(ij) \in E^+} x_{ij} - \sum_{(ij) \in E^-} x_{ij}$$

The actual count of unfair roundings is given by $z - \sum_{(ij) \in E^+} \lfloor q_{ij} \rfloor + \sum_{(ij) \in E^-} \lceil q_{ij} \rceil$. It is interesting to note that the TT and DAS methods obtain *a posteriori* quotas $\lambda_i \, v_{ij} \, \mu_j$ and an apportionment $[\![\lambda_i \, v_{ij} \, \mu_j]\!]$ that is necessarily a Best Rounding with respect to these quotas.

The approach for the L_∞-norm is different because it is not based on the direct solution of a linear programming minimization problem, rather on the solution of a sequence of feasibility problems. A thorough investigation of this approach can be found in Serafini and Simeone (2012a). Let us fix a deviation τ from the quotas. An apportionment that is also feasible for the constraints

$$q_{ij} - \tau \leq s_{ij} \leq q_{ij} + \tau$$

has maximum deviation not greater than τ. Since an apportionment must be integral, these constraints are equivalent to

$$\lceil q_{ij} - \tau \rceil \leq s_{ij} \leq \lfloor q_{ij} + \tau \rfloor \tag{3.5}$$

that, moreover, guarantee integrality of a feasible apportionment. Finding a feasible apportionment, or determining that the problem is infeasible, with respect to (3.1), (3.2) and (3.5) can be done via a Max Flow problem. We have to find the minimum value τ^* such that a feasible apportionment exists. This search can be done in a binary search fashion. The details of three different implementations can be found in Serafini and Simeone (2012a). We recall here that an optimal apportionment can be found in strongly polynomial time.

3.2.6 Other Issues

Non-uniqueness of the optimal apportionment is a serious issue and any method must be robust enough to prevent such circumstance. Uniqueness cannot be always guaranteed. One can construct examples in which the votes are so symmetrically distributed that there may be many equivalent apportionments. However, these circumstances may be considered extremely unlikely in a real election. There are other causes of non-uniqueness that some methods can exhibit that are inherent to the method itself and one has to find a way to fix them.

The L_∞-norm minimization has many equivalent optimal solutions because minimax solutions are insensitive to deviations for some pairs (i, j) which are less than the maximum deviation. A stronger form of L_∞-norm optimality to refine the choice among the optima is as follows: for a given apportionment x^* let $\tau^*_{hk} := |q_{hk} - x^*_{hk}|$ be the deviation for the pair (h, k). Let

$$L(h,k) := \left\{(i,j) = (h,k) : \tau^*_{ij} \le \tau^*_{hk}\right\}, \qquad U(h,k) := \left\{(i,j) : \tau^*_{ij} > \tau^*_{hk}\right\}.$$

$L(h, k)$ is the set of pairs with deviation not larger than τ^*_{hk} and $U(h, k)$ is the complement set, excluding (h, k) itself. Then we say that the apportionment x^* is *strongly optimal* if, for any pair (h, k), there is no apportionment with deviation $\tau_{hk} < \tau^*_{hk}$, $\tau_{ij} \le \tau^*_{hk}$ for $(i, j) \in L(h, k)$ and $\tau_{ij} \le \tau^*_{ij}$ for $(i, j) \in U(h, k)$.

Strongly optimal solutions are unique and a refinement of the previously stated binary search can be given that produces a strongly optimal solution (Serafini and Simeone, 2012a).

L_2-norm optimal solutions are robust in terms of uniqueness while L_1-norm optimal solutions can exhibit many equivalent solutions. It is shown in Ricca et al. (2012) that this undesirable circumstance is likely to happen if the apportionment does not stay within the quotas (compare with Theorem 3.2) and this in turn is a rare circumstance if fair share quotas are used.

We quote from Serafini and Simeone (2012b): “Electoral systems are usually quite complex and they are assembled out of many interacting components, ... it may happen that only mathematically sophisticated algorithms are available for solving a certain design problem. Are they “writable” as an actual law? Citizens rightly demand simple, easy to understand, voting systems. ... Which is better? To have simple, but unsound electoral laws, or sound, but complex ones?”

The way out from this dilemma is to “leave to a mathematically sophisticated algorithm the task of PRODUCING a sound solution, but attach to it a certificate of guarantee, that is, describe a simple procedure whereby ANYBODY CAN CHECK, through some elementary operations, that the solution output by the algorithm indeed satisfies all the requirements sought for.”

Since the minimization methods described are linear programming problems, strong duality holds for all of them and the certificate is indeed based on duality properties. Checking the claim that a solution is indeed optimal does not however require knowledge of mathematical programming theory. Only some elementary mathematical notions are needed. Describing the certificates in detail is out of the scope of this chapter and the reader is referred to Serafini and Simeone (2012b) and Serafini (2015).

3.3 Political Districting

Political Districting (PD) is particularly important in plurality systems with single-member districts. When only one seat is at stake in each district, the size and the shape of the districts may influence the outcome of the election, since even a single vote can produce the majority for one of the candidates. *Gerrymandering* is the name of the malpractice of designing biased electoral districts for favoring one preferred political party or candidate. But, even if gerrymandering is banned, the design of the districts remains a crucial technical issue in the definition of an electoral law, and it needs to be solved by using appropriate models and procedures. For this reason, many papers in the Operations Research (OR) literature studied this problem since the 1960's, providing different models and solution techniques (Grilli di Cortona et al., 1999, Ricca and Simeone, 1997 and Ricca et al., 2013).

3.3.1 Problem Definition

PD is a territorial partition problem which requires the discretization of the territory and imposes criteria related to spatial contiguity and population size. Here we assume that the territory is composed of a set of n elementary units, each identified by its geographical center and its population (*population units*).

Let $k < n$ be the total number of districts. We denote by p_i the size of the population of unit i, $i = 1, \ldots, n$, and by $P = \sum_{i=1}^{n} p_i$ the total population of the territory. The average district population is given by $\bar{P} = P/k$. A distance measure between units i and j is denoted by d_{ij}. The PD problem can be formulated as *finding a partition of the n units into k districts according to a specific set of criteria.* The main PD criteria are:

1. **Integrity:** each territorial unit cannot be split between two or more districts.

2. **Contiguity:** the units of each district should be geographically contiguous, that is, one can walk from any point in the district to any other without ever leaving the district.

3. **Population balance:** all districts should have the same portion of representation (*one person-one vote* principle); therefore single-member districts should have nearly the same populations.

4. **Compactness:** each district should be compact, that, according to the Oxford Dictionary, is, “closely and neatly packed together” (for example a round-shaped district).

An additional criterion frequently used in PD is the **respect of existing administrative subdivisions of the territory**. There are other PD criteria which are seldom used since there is no unanimous consensus on their legitimacy (e.g., *respect of natural boundaries*, *representation of ethnic minorities* and *respect of integrity of communities*). Broad discussions about political districting criteria can be found in Bozkaya et al. (2003), Grilli di Cortona et al. (1999), Kalcsics et al. (2005) and Ricca and Simeone (1997).

Traditionally, PD is formulated as an Integer Linear/Nonlinear Program (see, e.g., Hess et al. (1965), Garfinkel and Nemhauser (1970)), depending on the criterion selected for the objective function. From the seminal paper by Hess et al. (1965), works published in the 1960's and 1970's focused on location/allocation and transportation models and methods. Later, agglomerative techniques were mainly developed following Garfinkel and Nemhauser (1970) who proposed a set partitioning approach (Nygreen, 1988; Mehrotra et al., 1998).

Starting from the 1990's, local search methods became pervasive for PD (see

Bozkaya et al. (2003) and Ricca and Simeone (2008)), as well as techniques borrowed from the field of genetic and evolutionary algorithms. More recently, interesting approaches based on computational geometry were proposed (Kalcsics et al., 2005; Ricca et al., 2008).

In recent years, there was also a wide variety of papers basically describing the application of some known PD techniques (or slight variants) for the design of the electoral district map of a specific country. We note that PD can be seen as a particular case of the more general *territory design problem*, related to applications in public services like transportation districts, healthcare and school zoning, etc. On this topic, there is a rich and lively production of papers where PD is cited as one possible application, even if it is not the original motivation.

Many authors adopt a graph-theoretic model representing the territory as a connected n-node graph $G = (N, E)$ (*contiguity graph*, see Bodin (1973); Simeone (1978)), where the nodes correspond to the elementary territorial units and an edge between two nodes exists if and only if the two corresponding units are neighboring. To each node is assigned a weight representing its population.

In this case, the PD problem is formulated as follows: *find a compact partition of G into k connected components such that the weight of each component (sum of the weights of its nodes) is as close as possible to $\bar{P}$.*

It is well-known that the partition of a graph G into k connected components that minimizes population imbalance measured by an L_1-norm objective function is NP-hard even when $k = 2$ and G is a 2-spider, i.e., a tree with only one node with degree greater than 2 (De Simone et al., 1990). The problem remains NP-hard on spiders also for the L_ρ-norm with $2 \leqslant \rho < \infty$ and $k > 2$ (Schroeder, 2001).

In the following, we provide a brief overview of the PD mathematical models and methods of the last fifty years. Two main approaches emerge in our analysis of the literature, namely the *exact approach* (Section 3.3.2), and the *heuristic approach* (Section 3.3.3). The strength of the exact approach is that the problem is formulated by an algebraic optimization model. Therefore, in principle, any PD criterion can be modeled by a set of constraints, or it can be implemented in the objective function.

The drawback is that indicators adopted to measure the criteria may be highly non-linear (like for compactness), or even not computable by a formula, as it may happen for example for the respect of existing administrative subdivisions. In addition, the constraints in the PD model might be too many, i.e., their number may grow exponentially with the number of elementary units of the territory. This is the case of order constraints provided in Apollonio et al. (2008), by which contiguity of the districts is guaranteed, but at the cost of introducing an exponential number of constraints.

The power of the heuristic approach is that feasible solutions are characterized in a conceptually simple way so that they can evaluate a huge number of solutions in few seconds. As a counterpart, a loss in the quality of the solution must be accepted w.r.t. the exact approach. For both approaches, we discuss both the papers that are commonly considered milestones in this research field and also the ones that we deem to be the most representative, since, in our opinion, they produced innovative ideas or fixed drawbacks of previous works.

3.3.2 Exact Approach

In this section we review some classical mathematical models and solution techniques presented in the literature starting from the paper by Hess et al. (1965), which is generally considered the earliest OR paper in political districting. Here PD is formulated as a discrete location problem and the idea is to identify k units representing the centers of the k districts, so that each territorial unit must be assigned to exactly one center. The model has the following binary variables:

$$x_{ij} = \begin{cases} 1 & \text{if unit } i \text{ is assigned to center } j \\ 0 & \text{otherwise} \end{cases} \qquad i, j = 1, \ldots, n$$

and, in particular, $x_{jj} = 1$ if unit j is chosen as one of the centers and $x_{jj} = 0$ otherwise. The political districting problem is formulated as follows:

$$\begin{array}{lll} \min & \sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij}^2 \, p_i \, x_{ij} & \\ & \sum_{j=1}^{n} x_{ij} = 1 & i = 1, \ldots, n \\ & \sum_{j=1}^{n} x_{jj} = k & \\ & a \, \bar{P} \, x_{jj} \leqslant \sum_{i=1}^{n} p_i \, x_{ij} \leqslant b \, \bar{P} \, x_{jj} & j = 1, \ldots, n \\ & x_{ij} \in \{0, 1\} & i, j = 1, \ldots, n \end{array} \tag{3.6}$$

where a and b define the minimum and the maximum allowable district population, calculated as a percentage of the average district population $\bar{P}$ ($a < 1$, $b > 1$). By the first n constraints, each unit must belong to exactly one district. The next one imposes that the total number of districts is k. The $2n$ inequalities impose upper and lower bounds on the population of the districts. This type of constraint is frequently used to control population balance, since it is easy to read and understand also by non OR experts and lawmakers. The objective function measures compactness by the moment of inertia w.r.t. the district centers. The main drawback of the above integer programming model is that it does not take into account spatial contiguity of the districts at all. Therefore, an *a posteriori* revision may be necessary for assessing contiguity of the solution with an unavoidable loss in optimality.

In Garfinkel and Nemhauser (1970) a two-phase procedure based on a set partitioning approach is proposed. Phase I generates the set J of all possible feasible districts w.r.t. contiguity, population balance and compactness. In phase II the following set partitioning model is formulated that minimizes the overall deviation of district populations from $\bar{P}$.

$$\begin{array}{lll} \min & \sum_{j \in J} f_j \, x_j & \\ & \sum_{j \in J} a_{ij} \, x_j = 1 & i = 1, \ldots, n \\ & \sum_{j \in J} x_j = k & \\ & x_j \in \{0,1\} & j \in J \end{array} \tag{3.7}$$

where $f_j = (|P_j - \bar{P}|)/(\alpha \bar{P})$, and $\alpha \in [0, 1]$ is the tolerance on the percentage of deviation from $\bar{P}$ for the population of a district; $a_{ij} = 1$ if unit i is in district j and $a_{ij} = 0$ otherwise; $x_j = 1$ if district $j \in J$ is included in the partition and $x_j = 0$ otherwise. The same implicit enumeration strategy is followed to find all the feasible solutions in phase I and to find an optimal solution for (3.7) in phase II (for details the interested reader can refer to Geoffrion (1967)).

In this approach compactness is taken into account only in phase I, when districts are generated individually. It is measured on each single district separately with an index based on both the maximum distance between two territorial units in the district and the district area. A district is deemed compact if its index value is less than or equal to a fixed threshold. Then, the set partitioning problem in phase II does not consider any compactness measure for the whole district map, which, in fact, at the end, may result non-compact under different viewpoints. Note that, as suggested by Young (1988), there are many measures of compactness, and any good measure must apply both to the district map as a whole and to each district individually (for a classification of compactness measures see, e.g., Horn et al. (1993)).

The set partitioning approach in Garfinkel and Nemhauser (1970) was followed by other authors (Nygreen, 1988; Mehrotra et al., 1998) who suggested variants of model (3.7) aimed at improving the performance w.r.t. compactness. Both papers rely on a graph representation of the territory.

In Nygreen (1988) the innovative idea is that phase I is formulated in terms of spanning forests of G in which each subtree is rooted at some units playing the role of a district's center. Compactness is then controlled by imposing that the trees of the spanning forest have depth at most equal to two.

In Mehrotra et al. (1998) the problem is formulated as a constrained graph partitioning problem and a specialized branch-and-price solution methodology is developed. To take into account compactness properly, a cost function, based on distances computed between nodes in G, is defined on the set of possible districts, and the objective function of the set partitioning problem (*master problem*) is given by the sum of these costs. At each step, each new-generated district is priced with the same cost function used in the master problem, and this allows for controlling compactness of the whole map during the procedure.

The idea of formulating PD in terms of spanning forests is also exploited in Apollonio et al. (2008) and Lari et al. (2016), where the authors investigate *centered* graph partitioning problems, i.e., partitions of G into k of connected components, each including exactly one fixed center. They consider a class of objective functions based on unit-center costs that are independent of the topology of G (*flat costs*). For PD, population constraints are relaxed via a Lagrangean objective function, and flat costs correspond to the coefficients of such objective function. The problem becomes: *finding a spanning forest of G such that each tree in the forest contains exactly one center and the total cost is minimized.* The problem is shown to be NP-hard even on planar bipartite graphs (Apollonio et al., 2008; Lari et al., 2016), while it is polynomially solvable on trees. For this case, an interesting formulation is proposed where district contiguity is explicitly formulated by a set of order constraints. Unfortunately, these results cannot be directly exploited in PD applications, since the tree structure is too poorly connected to represent any real territory. In spite of this, the availability of efficient algorithms on trees leads to the idea of developing effective heuristics for finding a good district map on G through the (optimal) solution of a sequence of restrictions of the problem to spanning trees of G. It is worth noticing how contiguity is imposed in the above models by the use of order constraints. Since the district centers are fixed in advance, in a tree $T = (N, E)$ contiguity can be accomplished by imposing that if unit i is included in the district centered in s, then all units in G lying in the unique path $P_{i,s}$ from i to s must be included in the same district as i. The model has $O(n^2k)$ order constraints in total. This can be further improved to $O(n\,k)$ if one imposes order constraints on successive adjacent nodes in $P_{i,s}$ and exploits transitivity. Thus, the constraints of the PD model become:

$$\begin{aligned} &\sum_{s\in S} y_{is} = 1 && i \in U \\ &y_{is} \leqslant y_{j(i,s),s} && i \in U,\ s \in S,\ (i,s) \notin E \\ &y_{is} \in \{0,1\} && i \in U,\ s \in S \end{aligned} \tag{3.8}$$

where $S \subset N$, is the set of centers, with $|S| = k$, and $U = N \backslash S$. The binary variables y_{is} are defined as follows:

$$y_{is} = \begin{cases} 1 & \text{if unit } i \text{ belongs to the district centered in } s \\ 0 & \text{otherwise} \end{cases} \qquad i \in U,\ s \in S$$

Node $j(i, s)$ is the adjacent to i in the unique path from i to s. The feasible polytope described by (3.8) is integral. This could be exploited when the PD problem on a graph G is solved by the heuristic sketched above, that at each step can rely on linear programming for solving the problem on a spanning tree of G.

From the above discussion, two critical aspects emerge in the exact approach: i) guaranteeing contiguity, that needs to be formulated as a hard constraint; ii) measuring the other PD criteria, which may be a difficult task if an explicit analytic expression does not exist for some criteria. In this view, a heuristic approach may help, since the solution procedure is free from the rigid formulation of an algebraic model. In addition, the graph-theoretic model for the representation of the territory, that cannot be always fully exploited in a mathematical formulation, appears to be particularly fitting in a heuristic framework, as the papers reviewed in the following section show.

3.3.3 Heuristic Methods

In the last two decades, the use of heuristic techniques has taken a growing place in the study of PD problems. The main contributions in the literature are aimed at the evaluation of the performance of those meta-heuristics, like Tabu Search (TS), Simulated Annealing (SA), Threshold Algorithms (TA), Genetic Algorithms (GA), that have already shown to be successful for other difficult combinatorial optimization problems. Two extensive methodological works are provided in Ricca and Simeone (2008) and Bozkaya et al. (2003), both testing different versions of Local Search (LS) algorithms. LS is a powerful general purpose technique with a special capability of evaluating a huge number of different solutions in short times. The basic features of LS are: the *starting feasible solution*; the *(local) move*;

the *neighborhood* of a feasible solution. In PD the initial solution is generally easy to find since one can always rely on an already available administrative territorial division, or, in case of redistricting problems, even on the previous electoral district map which is going to be updated. A move operates a slight perturbation of a feasible solution. Given the current solution s, a neighboring solution of s is defined as any solution that can be obtained from s by performing a move. Although the variety of moves that can be thought of is wide, the principle of simplicity is in generally recommended, in order to avoid too sophisticated implementations which might slow down the computation. This principle is followed in both Ricca and Simeone (2008) and Bozkaya et al. (2003) where a move corresponds to the migration of one unit from a district to an adjacent one.

In Ricca and Simeone (2008) the aim is to investigate the intrinsic nature and potential of LS strategies like TS, SA, and TA. Therefore, streamlined versions of the algorithms are implemented. The authors rely on a graph-theoretic model to guarantee integrity and contiguity. PD is formulated as a multi-criteria optimization problem via a weighted objective function combining population balance, compactness, and conformity to administrative boundaries. In particular, good district maps are provided by *Old Bachelor Acceptance* (Hu et al., 1995), a threshold-based heuristic that is able to avoid premature stops in local optima by the use of a non-monotonic updating scheme.

In Bozkaya et al. (2003) a territory graph model is adopted and an enhanced LS procedure based on TS is developed within an adaptive memory search framework. During the procedure several 'good' district maps are generated, their districts are evaluated singularly by a performance function, and the best ones are recorded in order to be used again for restarting TS. This is, in fact, a mean for implementing both fitness selection, typical of GA, and a multi-start approach, that is generally recommended in LS. Beside the basic PD criteria, socio-economic homogeneity and integrity of communities are considered in a single weighted objective function.

A relatively new field of research on PD borrows notions and techniques from the computational geometry area. For the more general *territory design problem*, Kalcsics et al. (2005) propose an algorithm based on a continuous spatial model. The novelty is that discrete elementary territorial units are still considered but they are represented in the continuous space by the coordinates of their geographical centers. The algorithm repeatedly partition the territory into two half-spaces by drawing a straight line (*successive dichotomy strategy*). At each step, this generates two new subsets of territorial units. The benefit of the algorithm is that it is conceptually simple and easy to implement. This approach naturally satisfies contiguity, but which portion of territory must be divided next, and which straight line must be drawn, remain two substantial issues from which the performance of the algorithm strongly depends. In spite of this, in our opinion, the approach is worth to be investigated for further developments.

Ricca et al. (2008) apply to PD a heuristic approach based on Voronoi Regions (VR). The underlying idea is that VR are inherently compact, so that one may overcome the problem of choosing a measure of compactness. They refer to the graph representation of the territory and assign weights to the edges which represent distances between units. They introduce the notion of *weighted discrete*

Voronoi Regions that can be seen as the graph-theoretic counterpart of the ordinary VR in the continuous space. The authors propose algorithms that feature an iterative updating of the distances (according to different rules) in order to balance district populations as much as possible. In this model the authors exploit contiguity conditions formulated in Apollonio et al. (2008). Even if the graph is not a tree, these conditions can be used in the following way in the heuristic procedure. At each iteration contiguity of the districts is maintained thanks to the *geodesic consistency* property: if unit i belongs to district s and j lies on the geodesic between i and s, then j also belongs to district s, the geodesic being the shortest path between two nodes in G. By admitting a slight perturbation of the edge lengths, it can be assumed that there is a unique geodesic between any two nodes. Under this assumption, the authors prove that geodesic consistency implies contiguity. Actually, geodesic consistency can be seen as a way to formalize the order constraints in (3.8).

3.3.4 Practical and Application Issues

To conclude, we point out one main issue in the design of the electoral districts, that is: *if* and *how* the above discussed methods can be practically exploited in a law. It is generally difficult that formal models are accepted by lawmakers. However, differently from BAP, there is a general awareness that PD is a difficult problem. This could make computer based procedures more acceptable by lawmakers. Therefore, besides the study of new and more efficient methods, it is important to diffuse the already existing tools among the institutions. This would certainly help the administrative staff who has the (hard) task of executing all the procedures related to the political elections of a country. We believe that human contribution must not be excluded in the district definition process, but, when possible, it is recommended to take advantage from the power of mathematical modeling and automatic elaboration.

Bibliography

N. Apollonio, I. Lari, F. Ricca, B. Simeone, and J. Puerto. Polynomial algorithms for partitioning a tree into single-center subtrees to minimize flat service costs. *Networks*, 51:78–89, 2008.

A. Bachem and B. Korte. On the RAS- algorithm. *Computing*, 23:189–198, 1979.

M. L. Balinski and G. Demange. Algorithms for proportional matrices in reals and integers. *Mathematical Programming*, 45:193–210, 1989a.

M. L. Balinski and G. Demange. An axiomatic approach to proportionality between matrices. *Mathematics of Operations Research*, 14:700–219, 1989b.

M. L. Balinski and H. Young. *Fair Representation – Meeting the Ideal of One Man, One Vote.* Brookings Institution Press; second edition, 2001.

L. Bodin. A districting experiment with a clustering algorithm. *Annals of the New York Academy of Sciences*, 219(1):209–214, 1973.

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

L. Cox and L. Ernst. Controlled rounding. *INFOR—Information Systems and Operational Research*, 20:423–432, 1982.

C. De Simone, M. Lucertini, S. Pallottino, and B. Simeone. Fair dissections of spiders, worms, and caterpillars. *Networks*, 20(3):323–344, 1990.

R. Garfinkel and G. Nemhauser. Optimal political districting by implicit enumeration techniques. *Management Science*, 16(8):495–508, 1970.

A. M. Geoffrion. Integer programming by implicit enumeration and Balas' method. *SIAM Review*, 9:178–190, 1967.

P. Grilli di Cortona, C. Manzi, A. Pennisi, F. Ricca, and B. Simeone. *Evaluation and optimization of electoral systems.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1999.

G. R. Grimmett. European apportionment via the Cambridge compromise. *Mathematical Social Sciences*, 63(2):68–73, 2012.

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

D. L. Horn, C. R. Hampton, and A. J. Vandenberg. Practical application of district compactness. *Political Geography*, 12(2):103–120, 1993.

T. C. Hu, A. B. Kahng, and C. W. A. Tsao. Old bachelor acceptance: A new class of non-monotone threshold accepting methods. *ORSA Journal on Computing*, 7 (4):417–425, 1995.

B. Kalantari, I. Lari, F. Ricca, and B. Simeone. On the complexity of general matrix scaling and entropy minimization via the ras algorithm. *Mathematical Programming, Series A*, 112:371–401, 2008.

J. Kalcsics, S. Nickel, and M. Schröder. Towards a unified territorial design approach-applications, algorithms and gis integration. *Top*, 13(1):1–56, 2005.

I. Lari, F. Ricca, J. Puerto, and A. Scozzari. Partitioning a graph into connected components with fixed centers and optimizing cost-based objective functions or equipartition criteria. *Networks*, 67(1):69–81, 2016.

A. Mehrotra, E. Johnson, and G. Nemhauser. An optimization based heuristic for political districting. *Management Science*, 44(8):1100–1114, 1998.

M. Minoux. A polynomial algorithm for minimum quadratic cost flow problems. *European Journal of Operational Research*, 18:377–387, 1984.

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

F. Pukelsheim. Bazi – a Java program for proportional representation. Reports 1, Oberwolfach, 2004.

F. Pukelsheim and C. Schuhmacher. Das neue Zürcher Zuteilungsverfahren für Parlamentswahlen. *Aktuelle Juristische Praxis – Pratique Juridique Actuelle*, 13: 505–522, 2004.

F. Ricca and B. Simeone. Political redistricting: Traps, criteria, algorithms, and trade-offs. *Ricerca Operativa*, 27:81–119, 1997.

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

F. Ricca, A. Scozzari, and B. Simeone. Weighted Voronoi region algorithms for political districting. *Mathematical and Computer Modelling*, 48(9-10):1468–1477, 2008.

F. Ricca, A. Scozzari, P. Serafini, and B. Simeone. Error minimization methods in biproportional apportionment. *TOP*, 20:547–577, 2012.

F. Ricca, A. Scozzari, and B. Simeone. Political districting: from classical models to recent approaches. *Annals of Operations Research*, 204:271–299, 2013.

U. G. Rothblum and H. Schneider. Scaling of matrices which have prescribed row sums and column sums via optimization. *Linear Algebra Applications*, 114: 737–764, 1989.

M. Schroeder. *Gebiete optimal aufteilen - OR-Verfahren für die Gebietsaufteilung als Anwendungsfall gleichmäßiger Baumzerlegung*. PhD thesis, Universität Karlsruhe, 2001.

P. Serafini. Allocation of the EU parliament seats via integer linear programming and revised quotas. *Mathematical Social Sciences*, 63:107–113, 2012.

P. Serafini. Certificates of optimality for minimum norm biproportional apportionments. *Social Choice and Welfare*, 44:1–12, 2015.

P. Serafini and B. Simeone. Parametric maximum flow methods for minimax approximation of target quotas in biproportional apportionment. *Networks*, 59: 191–208, 2012a.

P. Serafini and B. Simeone. Certificates of optimality: the third way to biproportional apportionment. *Social Choice and Welfare*, 38:247–268, 2012b.

B. Simeone. Optimal graph partitioning. *Atti giornate di lavoro AIRO, Urbino*, pages 57–73, 1978.

H. Young. Measuring the compactness of legislative districts. *Legislative Studies Quarterly*, 13(1):105–115, 1988.