Network flow methods for electoral systems

Friedrich Pukelsheim* Federica Ricca⋆

Andrea Scozzari⋄ Paolo Serafini+ Bruno Simeone⋆ [1]

*Institut für Mathematik, Universität Augsburg
Universitätstrasse 14, 86159 Augsburg, Germany

⋆Dipartimento di Scienze Statistiche, Sapienza, Università di Roma
Piazzale Aldo Moro 5, 00185 Roma, Italy

⋄Facoltà di Economia, UNISU, Università Telematica delle Scienze Umane
Via Casalmonferrato 2/b, 00182 Roma, Italy.

+Dipartimento di Matematica e Informatica, Università di Udine
Via delle Scienze 206, 33100 Udine, Italy

Abstract

Researchers in the area of electoral systems have recently turned their attention to network flow techniques with the aim to resolve certain practically relevant problems arising in this area. The aim of the present paper is review some of this work, showing the applicability of these techniques even to problems of a very different nature.

Major emphasis will be placed on "biproportional apportionment", a problem that frequently arises in proportional electoral systems, but which in some countries is still ill-solved, or not dealt with rigorously, notwithstanding the availability of several sound solution procedures and their concrete application in some real-life elections. Beside biproportional apportionment, we shall discuss applications of network flows to problems such as vote transitions and political districting. Finally, we address the so-called "Give-up Problem", which arises in the current elections for the Italian Parliament. It is related to the possible assignment of seats to multiple winners of a given party. Based on the results and techniques presented in this paper, it is fair to state that network flow models and algorithms are indeed very flexible and effective tools for the analysis and the design of contemporary electoral systems.

Keywords: *network flows, electoral systems, biproportional seat apportionment, matrix scaling, political districting, closed lists, give-up.*

[1] We dedicate this work to our dear friend and senior coauthor Bruno Simeone, who passed away while this paper was being finished. He was always a constant source of inspiration and this paper owes greatly to his precious ideas.

1 Introduction

The use of network flow models and methods is widespread in operations research, with many applications in a large variety of areas [1]. In recent years, researchers in the field of electoral systems have turned their attention to network flow techniques in order to deal with biproportional apportionment and other electoral decision problems. We feel that the time has come to give an account of this research direction. For an optimization-oriented general introduction to electoral systems, the reader may refer to Grilli di Cortona et al. [28].

A transportation procedure appears in Hess et al. [30], the earliest operations research paper in political districting. Since then, network flow models have been proposed for the design of certain components of electoral systems or for the analysis of their behavior. In this survey we address some relevant problems in electoral systems, namely, the Biproportional Apportionment Problem (BAP), the computation of vote transitions, the design of political districts and the so-called "Give-Up Problem".

Biproportional seat apportionment, to which we devote Sections 2–8, is perhaps the main area of electoral systems where network flow techniques are brought to bear. The problem arises in situations where the entire electoral region, usually the whole nation, is subdivided into electoral districts. By constitutional or legal requirements, the electoral districts are to receive a share of seats proportional to their population counts. At the same time, political parties are to be allocated a number of seats that mirrors their nationwide electoral performance. Thus, BAP is a "matrix problem" for which we provide a formal definition in Section 2. This problem currently arises in the electoral laws of several countries, e.g., Italy, Mexico, Switzerland, Denmark, Faroe Islands, etc., but it may be of primary interest also in the European Parliament elections, where the districts correspond to member states of the European Union.

In the field of statistical applications, some authors analyzed the very structure of BAP as a transportation problem. For example, in [15, 16] the authors study two categories of problems: *controlled selection problems*, that is, controlling statistical disclosure in tables of frequency counts (in order to prevent small counts in such tables to be easily inferred); the more general statistical problem of replacing a table of noninteger counts by an integer one, matching the prescribed row- and column- sums, and minimizing a measure of the total distortion to the original table. A procedure to solve such problems that has a long standing already in statistics, is *Iterative Proportional Fitting* (IPFP) [17]. In computer science it is called *matrix scaling*, or the RAS method (after the matrix names used by Bacharach in his early paper [2]), whose complexity was recently analyzed by Kalantari, Lari, Ricca, and Simeone [35]. The reader may wish also to consult the classical monograph on this method by Bacharach [3].

In 1989, Balinski and Demange [5, 6] published two seminal papers where they characterized proportionality between integral matrices axiomatically, and proposed a procedure to find apportionments X proportional to V in the above axiomatic sense. Their procedure was implemented as the *Tie and Transfer* (TT) algorithm in the public domain software BAZI [38]. The results by Balinski and Demange are surveyed in Sections 3 and 4.

Looking for a simple algorithm for BAP, in 2004 Pukelsheim [45] proposed a RAS-like *Discrete Al-*

ternating Scaling (DAS) procedure which was actually applied in the elections of the Zurich, Aargau and Schaffhausen Cantons (see Section 5).

In 2008 Gaffke and Pukelsheim [21, 22] formulated BAP as a piecewise linear convex separable transportation model and showed that the TT method of Balinski and Demange may be viewed as an out-of-kilter algorithm for solving such model, while the Discrete Alternating Scaling method may be viewed as a cyclic coordinate ascent algorithm for solving its (Fenchel) dual (see Section 6).

These approaches are based on the idea of rounding certain fractional numbers, i.e., the "fair shares", which would be the ideal seat assignment if fractional seats were allowed. They focus more on how to round these quotas while keeping satisfied the row and column sums than on the distance of the final seat assignment to the quotas, taking for granted that the approximation is good because it can be obtained by rounding up or down the fair shares.

A different approach consists of minimizing an error measure of the actual seats with respect to "ideal" quotas. The quoted result (controlled rounding) by Cox and Ernst [16] can be viewed as a polynomial procedure to minimize the L_p-norm error with respect to ideal quotas, subject to the constraint that such quotas are rounded only to the up and down nearest integers. Network flow integrality is the basic property allowing for polynomiality of controlled rounding. It should be noted that this method does not minimize in general the L_p-norm, because there are instances where the minimum is obtained outside the above rounding interval. However, exploiting a result by Minoux [40, 41] one can refine the method into a polynomial algorithm for the general case. This will be discussed in Section 7.

Looking for the minimization of the L_∞-norm and following a different approach, Serafini and Simeone [52] formulate BAP as a minimax approximation of target shares, also providing a strongly polynomial parametric maximum flow algorithm to solve it (see Section 8).

In Section 9 we address a problem which is subject to very careful analysis after each election, i.e., understanding if and how electors have changed their vote with respect to the previous election. The votes migrating to different parties are generally referred to as *electoral flows* or *vote transitions* and, not surprisingly, transportation models arise in this context.

In Section 10 we show that network flow techniques can be applied also to the political districting problem. Beside the early method proposed by Hess et al. [30] some variants have been developed in the literature and actually applied to solve real-life political districting problems: a first example is related to the provincial electoral districts for the city of Saskatoon, Canada, in 1996 [31]; a second one refers to the definition of Parliamentary district boundaries in New Zealand in 1997 [26].

The last section (Section 11) is devoted to the presentation of the "Give-up Problem", that is, the problem of assigning seats to the winning candidates of a given party. The problem arises when the electoral system has closed lists in the districts and multiple winners are possible. The attention paid to this problem is motivated by the Italian case analysis, and justified by results that show that network flow techniques are appropriate for the solution of this problem.

2 Biproportional seat apportionment

A formal definition of BAP is as follows. Let H be the *house size*, that is, the total number of seats, of a parliament. Firstly, the H parliamentary seats are apportioned among m electoral districts proportionally to population counts, allocating r_i seats to district $i \in M = \{1, \dots, m\}$. Secondly, the H seats are apportioned among n lists of candidates of the contending parties, proportionally to the number of votes each party has received. Let c_j be the nationwide seats of party $j \in N = \{1, \dots, n\}$. Clearly, one has $\sum_{i \in M} r_i = \sum_{j \in N} c_j = H$. Both steps, of apportioning the H seats among the districts on the one hand, and among the parties on the other, form the *super-apportionment*. We assume that both steps have been carried out, so that the seats r_i and the seats c_j are known and available. Balinski and Young [7] is the ultimate comprehensive reference on proportional seat apportionment, its mathematical aspects, and its history.

Let v_{ij} be the number of votes in district i for party j. That is to say, the vote counts are the input data and form an $m \times n$ matrix V. Let $Z = \{(i, j) : v_{ij} = 0\}$ be the "zero-pattern" of V, that is, the set of the structural zeros of V.

The seat numbers x_{ij} form an integer nonnegative $m \times n$ matrix X, which is an *apportionment* if it satisfies the following constraints:

$$
\begin{array}{llll}
1) & \displaystyle\sum_{j \in N} x_{ij} = r_i, & i \in M & \text{(district sum);} \\
2) & \displaystyle\sum_{i \in M} x_{ij} = c_j, & j \in N & \text{(party sum);} \\
3) & \quad x_{ij} = 0, & (i, j) \in Z & \text{(zero-vote zero-seat).}
\end{array} \tag{1}
$$

We denote by $\mathcal{A}$ the set of apportionments and by $\hat{\mathcal{A}}$ the set of *fractional apportionments*, i.e., real matrices satisfying 1), 2) and 3). Constraints 1) and 2) mean that rows in X must sum to the prespecified row marginals r_i, $i \in M$, and column sums must be equal to the given column marginals c_j, $j \in N$. Condition 3) guarantees that a party j that does not receive votes in a district i is not awarded any seat in that district.

BAP can be formulated as follows: *given the vote matrix V and the vectors r and c, find an apportionment $X \in \mathcal{A}$ "as proportional as possible" to V.*

It is not obvious, and more of a challenge, to turn the proportionality requirement into an operational concept. The difficulty is twofold. On the one hand we have to find a definition of "ideal" proportionality, and, on the other hand, we have to make a compromise between the ideal proportionality and the integrality requirement for the seats.

The fact that the problem is not trivial is witnessed by the presence in the electoral law of some countries of unsound and self-contradictory procedures for solving BAP. For example, Balinski and Ramírez [9] discovered that the Mexican electoral law was not correct, with the result that the procedure was modified. However, in other countries - like Italy - this problem still persists.

In some countries, like Italy and Belgium, *regional quotas* are used as a template of ideal proportionality. For each region i and party j, they are given by

$$q_{ij} = r_i \frac{v_{ij}}{\sum_{h \in N} v_{ih}}, \qquad i \in M,\ j \in N.$$

Equivalently, q_{ij} is the (usually fractional) number of seats party j would receive in region i under the assumption of perfect proportionality between votes and seats in that region. Notice that regional quotas depend only on data associated with the given region. It is conceivable also to use party-wise (i.e., column-wise) quotas. However, both in the Belgian and Italian legislation, districts and parties are not dealt with in a symmetric fashion, and proportionality within districts is felt to be more important than proportionality within parties. This asymmetry is confirmed by the fact that in the Italian system seats are assigned to parties by an ordinary law and to districts by the very Constitution.

Regional quotas reflect proportionality within districts. Since they are usually not integers, one way to obtain an apportionment is to find a "suitable" rounding of the regional quotas that satisfies constraints (1). The Italian biproportional allocation procedure relies on the underlying assumption that one can always get an apportionment by rounding up or down the regional quotas. Unfortunately, realistic examples can be exhibited in which no up- or down-rounding of the regional quotas satisfies both the district- and the party-sum constraints. The Italian procedure tries to solve the biproportional apportionment problem in the wrong way [43]: the matrix of seats produced by such procedure may fail to satisfy the district-sum constraints, the party-sum constraints, or both. The result is that in the five last political elections for the Chamber of Deputies this has indeed happened three times (precisely in 1996, 2006, 2008). The unavoidable consequence is that *citizens living in different districts of the same country have different voting power*. For instance, in the 2006 political elections, the Trentino-Alto Adige district got 11 seats instead of the 10 granted by the Constitutional Law, while the Molise got 2 seats instead of 3. As a consequence, in Trentino-Alto Adige 85,456 votes were necessary to get one seat, while in Molise one needed 160,300 votes per seat. Therefore, it is legitimate to state that in the 2006 elections for Molise's citizens the motto "one-man-half-vote" applied! Similar results were obtained in the more recent Italian political elections of 2008 [44].

Actually, correct procedures for BAP do exist, as demonstrated by the many papers in the literature on this topic. In 1989, Michel Balinski and Gabrielle Demange [5, 6] published two seminal papers where they characterized proportionality between real and integral matrices axiomatically. Their results are surveyed in the next two sections.

3 Proportionality between two real matrices

In [5] Balinski and Demange characterize proportionality between two real matrices axiomatically. They introduce axioms of *Exactness*, *Relevance*, *Uniformity*, *Monotonicity*, and *Homogeneity*, describing reasonable properties that an apportionment should satisfy. The authors prove that, given V, r, c, there exists, under some necessary assumptions, a unique matrix F, called the *fair share matrix*, proportional to V (in the above axiomatic sense) with the same zero-pattern Z as V and fitting the row- and column-sums r, c. In order to find such a matrix Balinski and Demange follow the continuous approach of RAS. Thus,

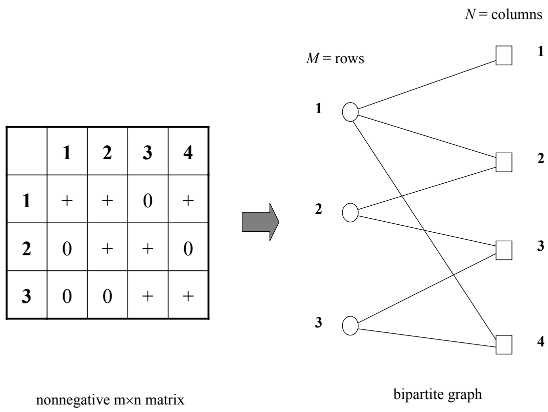

Figure 1: A nonnegative 3×4 matrix V and the corresponding bipartite graph.

matrix F has the form $[f_{ij}] = [\lambda_i \, v_{ij} \, \mu_j]$ for suitable positive row multipliers λ_i and column multipliers μ_j.

The RAS algorithm can be briefly summarized in the following main steps. Starting from V, all rows are scaled to fit their prespecified row-sums, thus generating a row-wise rescaled matrix $V(1)$. In $V(1)$, all columns are scaled to fit their prespecified column-sums, giving rise to a column-wise rescaled matrix $V(2)$. Continue by alternately scaling all rows at each odd step, then all columns at the subsequent even step. The procedure yields rescaled matrices $V(t)$ that are usually convergent to the fair share matrix sought for, $\lim_{t\to\infty} V(t) = F$.

Unfortunately, F is a fractional apportionment, while BAP requires an integral one. Thus, a suitable rounding of F must be performed in order to get a solution for BAP. Balinski and Demange observe that from the Integrality Theorem of Flows [1] the following fundamental *Rounding Property* holds: one can always obtain an apportionment by rounding either up or down the entries of the fair share matrix F.

It is well known (see, for example, [48]) that a nonnegative $m \times n$ matrix A can be represented by a bipartite graph $(M, N; E)$, where the node sets M and N correspond to the set of rows and the set of columns of A, respectively, and there exists an edge in E if and only if $a_{ij} > 0$ (Fig. 1). Furthermore, one may direct each edge in E from M to N and assign source values r_i to the nodes in M and sink values c_j to the nodes in N. Let us denote this network by $\mathcal{G}$. Thus, the problem of finding a nonnegative integer matrix satisfying constraints 1), 2) and 3) of (1) can always be formulated as finding a feasible solution to a transportation problem on $\mathcal{G}$, where constraints 1) and 2) correspond to supply and demand constraints, while 3) defines forbidden routes in $\mathcal{G}$.

The following theorem summarizes some equivalent conditions for the existence of a fair share matrix (see, [4, 35, 46, 49, 52]), where $N(S)$ denotes the neighborhood of $S \subset M$, i.e., $N(S) = \{j \in N : (i,j) \in E$ for some $i \in S\}$.

Theorem 1 *For a nonnegative $m \times n$ matrix V and nonnegative vectors r and c, the following statements are equivalent:*

(i) there exists a fair share matrix for (V, r, c);

(ii) there exists a matrix X satisfying the following system of linear constraints:

$$\begin{array}{rcll} \sum_{j:(i,j)\in E} x_{ij} & = & r_i & i \in M \\ \sum_{i:(i,j)\in E} x_{ij} & = & c_j & j \in N \\ x_{ij} & \geq & \dfrac{1}{|E|} & (i,j) \in E \end{array} \qquad (2)$$

(iii) $\sum_{i\in S} r_i \leq \sum_{j\in N(S)} c_j$ *for each* $\emptyset \neq S \subset M$.

∎

Condition *(ii)* can be checked via the solution of a max-flow problem on a suitably modified network. Condition *(iii)* follows from the well-known Marriage Theorem [34] if the transportation problem on $(M, N; E)$ is suitably reformulated on a bipartite graph with all supplies and all demands equal to 1 .

To network optimization people, interest in the above theorem is twofold. On the one hand, one can check condition *(ii)* by solving a suitable feasible flow problem which, in turn, is well known to be reducible to a maximum flow one. So, one gets yet another application of maximum flows, namely, checking the convergence of the RAS algorithm. On the other hand, things work also in the reverse way. Given a transportation problem with supplies r, demands c, and with forbidden routes, one can find a feasible transportation plan, if any, as follows: start from an arbitrary nonnegative $m \times n$ matrix V such that $v_{ij} = 0$ for any pair i, j corresponding to a forbidden route. Run RAS on V: if the algorithm converges, then the unique limit matrix provides a feasible solution; otherwise, no such solution may exist.

Actually, this result provides an alternative method, computationally effective in practice, for finding a feasible solution to a transportation problem. The result looks quite surprising if one considers that the above procedure works no matter what the starting matrix V is: V can be chosen arbitrarily, provided that its zero entries correspond to the forbidden routes in the transportation problem. It must be understood that this invariance result is related to the existence of a feasible solution and not to the feasible solution itself (if any). Indeed, starting from different matrices V one obtains different limit matrices.

4 Proportionality between two integral matrices: the Tie and Transfer method

The Tie and Transfer (TT) method of Balinski and Demange is a divisor-based algorithm to find an apportionment $X \in \mathcal{A}$. In order to understand divisor-based methods for BAP, we refer to the simpler case of *vector apportionment*, that is, apportionment in one dimension. Given a nonnegative real n-vector $v = (v_1, \ldots, v_n)$ and a positive integer H, one wants to find a nonnegative integral n-vector $x = (x_1, \ldots, x_n)$ (the apportionment) with sum of components equal to H and such that the x_i's are "as proportional as possible" to the v_i's.

Any divisor method is characterized by a *signpost sequence* given by a signpost function $s(z)$ mapping each integer z into a real number in the interval $[z-1, z]$; that is,

$$z - 1 \leq s(z) \leq z, \qquad z = 1, 2, \ldots$$

Let $[\![t]\!]$ denote the rounding of $t \in [z-1, z]$. Then, if $z - 1 \leq t \leq s(z)$, we have $[\![t]\!] = z - 1$, while we have $[\![t]\!] = z$ if $s(z) \leq t \leq z$. When $t = s(z)$ one has either $[\![t]\!] = z - 1$ or $[\![t]\!] = z$. For example, standard

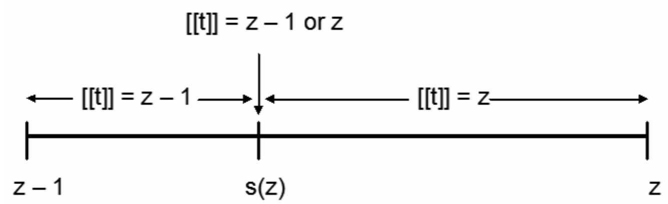

Figure 2: Signpost function in divisor-based methods.

rounding, that is, rounding to the closest integer, is defined by the signposts $s(z) = z - 1/2$ (see Fig. 2). The divisor method corresponding to a signpost sequence $\{s(z)\}$ consists of the choice of a multiplier λ such that, letting $x_j = [\![\lambda\, v_j]\!]$ for each j (or, equivalently, $s(x_j) \le \lambda\, v_j \le s(x_j + 1)$), the resulting vector x has the sum of components equal to H.

It transpires that the multiplier λ plays a crucial role. Given λ, we easily calculate the seat numbers from the formula $x_j = [\![\lambda\, v_j]\!]$. Conversely, asssume that we are given seat numbers x_j that sum up to the desired house size H. Then the apportionment $(x_1, \ldots, x_n)$ originates from the divisor method with signpost sequence $s(z)$ if and only if

$$\max_{j:v_j>0} \frac{s(x_j)}{v_j} \le \min_{j:v_j>0} \frac{s(x_j+1)}{v_j}$$

and in this case every number λ in the *multiplier interval* [22]

$$[\lambda^-, \lambda^+] = \Big[\ \max_{j:v_j>0} \frac{s(x_j)}{v_j}\ ,\ \min_{j:v_j>0} \frac{s(x_j+1)}{v_j}\ \Big]$$

may serve as a viable multiplier for the apportionment under consideration.

As in the fractional case, Balinski and Demange show in [5] that divisor-based matrix apportionment methods satisfy some theoretical properties that guarantee proportionality of the resulting apportionment from an axiomatic viewpoint. They characterize proportionality between two integral matrices by a system of six axioms, five of which are the integer counterparts of the previous ones, while the additional one is an axiom of *Completeness* [5]. Then, they introduce the TT algorithm, a procedure whose basic strategy is *Scale and Round*. Given a matrix V, vectors r, c, and a divisor method with signpost sequence $\{s(z)\}$ $(s(1) > 0)$, there exists, under the same assumptions as in the real case, an apportionment X proportional to V w.r.t. the introduced axioms.

To obtain the seat numbers x_{ij}, the TT algorithm computes row multipliers $\lambda_i > 0$ and column multipliers $\mu_j > 0$ such that

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

and the multipliers satisfy the following *rounding inequalities*

$$s(x_{ij}) \le \lambda_i\, v_{ij}\, \mu_j \le s(x_{ij} + 1). \tag{3}$$

When $\lambda_i\, v_{ij}\, \mu_j = s(\lceil \lambda_i\, v_{ij}\, \mu_j \rceil)$ occurs, $\lambda_i\, v_{ij}\, \mu_j$ cannot be rounded unequivocally. In this case we have a *tie*. If we resolve the tie by assigning $x_{ij} = \lfloor \lambda_i\, v_{ij}\, \mu_j \rfloor$ we refer to an *upper tie*, whereas we refer to a *lower tie* if we set $x_{ij} = \lceil \lambda_i \, v_{ij} \, \mu_j \rceil$. We observe that sometimes the other constraints allow for only one way to resolve the tie in order to obtain a feasible apportionment [55].

It must be noticed that any apportionment X', obtained from X after replacement of x_{ij} by $x'_{ij} = x_{ij}+1$, for some upper ties (i,j), and by $x'_{ij} = x_{ij}-1$, for some lower ties (i,j), also satisfies the rounding inequalities relative to the same multipliers.

In [8] it is proved that divisor-based methods provide an apportionment X (if it exists) unique up to ties.

The TT algorithm searches for $X \in \mathcal{A}$ that minimizes the following L_1-error:

$$\frac{1}{2}\sum_{i\in M} |x_{iN} - r_i| + \frac{1}{2}\sum_{j\in N} |x_{Mj} - c_j| \tag{4}$$

where $x_{iN} = \sum_{j\in N} x_{ij}$ and $x_{Mj} = \sum_{i\in M} x_{ij}$.

The objective function (4) is iteratively minimized by a procedure relying on "transfer" operations. W.l.o.g, initially all the row multipliers are set to 1. The column multipliers μ_j, $j \in N$, are then computed such that $v_{ij}\,\mu_j$ satisfies the column-sum c_j, that is, $\mu_j = c_j / \sum_{i\in M} v_{ij}$, $j \in N$. Starting from V, for each column j the vector apportionment is solved w.r.t. $v_{1j}, \ldots, v_{mj}$ and c_j, thus obtaining an integer matrix $X^{(0)}$ (current solution) where column-sums match c, but row-sums are generally not satisfied. Hence, the error in (4) reduces to

$$\frac{1}{2}\sum_{i\in M} |x_{iN} - r_i|. \tag{5}$$

Starting from $X^{(0)}$ the algorithm proceeds by decreasing the error while maintaining the column-sums constraints satisfied. Since in $X^{(0)}$ column-sums are satisfied, and we also have $\sum_{j\in N} c_j = \sum_{i\in M} r_i = H$, then, if the row-sums are not satisfied, there must exists at least one *underbalanced* row i for which $\sum_{j\in N} x_{ij} < r_i$, and at least one *overbalanced* row k for which $\sum_{j\in N} x_{kj} > r_k$. In order to decrease the error, one can "transfer" seats from overbalanced rows to underbalanced ones, leaving the balanced rows unchanged. The only condition that must be satisfied is that only the cells corresponding to ties can be modified. In other words, if x_{ij} corresponds to an upper tie in a underbalanced row it can be increased by 1, and if x_{ij} corresponds to lower tie in an overbalanced row it can be decreased by 1, while all other cells must remain the same (see Fig. 3 where the grey cells corresponds to ties, and a +1 denotes an upper tie, while -1 stands for a lower tie).

It is easy to recognize that at each iteration t a directed bipartite graph can always be associated to each matrix $X^{(t)}$ [55] with the set of nodes corresponding to those rows and columns in $X^{(t)}$ that are involved in some ties (sometimes this graph is referred to as the *row/column graph*). For $i \in M$ and $j \in N$ and $x_{ij} = [\![\lambda_i \, v_{ij} \, \mu_j]\!]$, an arc (i,j) exists if $\lambda_i \, v_{ij} \, \mu_j = s(x_{ij}+1)$ (upper tie), while an arc (j,i) exists if $\lambda_i v_{ij} \mu_j = s(x_{ij})$ (lower tie). Then, the transfer of a seat from an overbalanced row (district) to an underbalanced one corresponds to a flow along a simple (even) path P connecting such rows in the row/column graph. Assume that $X^{(t)}$ is the current matrix, then, if path P can be found, the corresponding transfer can be performed producing a decrease of 1 in the error (5) (*Primal Step* of the TT algorithm). Suppose that $P = \{(i_1, j_1), (j_1, i_2), \ldots, (j_{g-1}, i_g)\}$. Then, the transfer along P corresponds

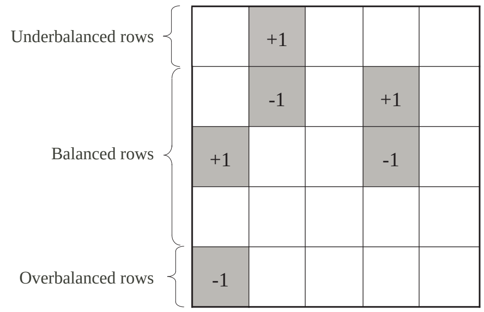

Figure 3: The structure of the current matrix $X^{(t)}$ in the TT algorithm.

to rounding up $x^{(t)}_{i_1 j_1}$, rounding down $x^{(t)}_{i_2 j_1}$ and so on, up to rounding down the last element $x^{(t)}_{i_g j_{g-1}}$. The global result will be that the i_1-th row-sum is increased by 1, the i_g-th row-sum is decreased by 1, while the other row-sums and all the column-sums are not affected by the transfer. Consequently, the error (5) decreases by 1. In order to perform a transfer, for a given $L \subset M \cup N$, TT applies a breadth-first-search to identify all vertices that are reachable from a vertex in L through a simple path in the row/column graph. At the beginning of iteration t, L corresponds to the indices of the underbalanced rows in $X^{(t)}$; then, all the reachable vertices are added to L. If one of the overbalanced rows is reached, path P has been found; otherwise, matrix $X^{(t)}$ must be updated in order to produce additional ties that may help in reaching an overbalanced row in a following (primal) step.

The updating of $X^{(t)}$ is performed during a *Dual Step* in which row and column multipliers are suitably modified. At this stage of the algorithm the current set L may include indices from both rows and columns of $X^{(t)}$, that are considered as *labeled*, that is, $L = M_L \cup N_L$, where M_L and N_L are subsets of labeled rows' and column's indices, respectively.

The multipliers are updated through a factor $\delta > 0$ such that, when multiplying all rows $i \in M_L$ by δ and all columns $j \in N_L$ by $1/\delta$, the current solution remains feasible. Thus, δ is computed as a bottleneck value, i.e., it is the maximum value that guarantees that all the rounding inequalities are still valid, but at least one is satisfied with equality. This produces at least one additional tie in the new feasible apportionment $X^{(t+1)}$ (see, [5, 55]).

The underlying idea is that the addition of new ties may help in finding a path P from an overbalanced row to an underbalanced one in the updated row/column graph corresponding to $X^{(t+1)}$ (that means performing an additional transfer).

The TT procedure either produces an apportionment $X \in \mathcal{A}$, or halts (after a dual step) reporting that no solution exists. A detailed description (a pseudo-code) of the algorithm can be found in [37] and in [55] where the author also provides a polynomial time implementation of it.

5 Discrete alternating scaling procedure

For the sake of completeness, we briefly recall the Discrete Alternating Scaling (DAS) procedure for BAP proposed in [45]. The procedure is very simple and performs the following basic steps: starting from V, alternately scale each row i of the current (unrounded) matrix so that the sum of its rounded entries matches r_i and then scale each column j of the current (unrounded) matrix so that the sum of its rounded entries matches c_j.

In order to do this, each row multiplier λ_i must be chosen in a feasible interval $[\lambda_i^-, \lambda_i^+]$ defined as before; similarly for column multipliers.

Usually the algorithm enjoys finite termination, providing the required apportionment. However, there are rare cases when the algorithm stalls at a nonoptimal pair of row and column multipliers (see [22]).

Discrete Alternating Scaling was implemented in BAZI (see Maier and Pukelsheim [38]), with the provision that, if stalling is produced the algorithm automatically switches to the TT algorithm for which termination is guaranteed. In the following we report the pseudo-code of DAS where we set $v_{ij}(0) = v_{ij}$, $t = 1$ and increase t by one after each step, until x does not change from one step to the other.

Odd Step: Find row multipliers $\lambda_i(t)$, $i \in M$, such that $v_{ij}(t) = [\lambda_i(t)\, v_{ij}(t-1)]$, and $x_{ij}(t) = [\![\lambda_i(t)\, v_{ij}(t-1)]\!]$ satisfy the conditions $x_{iN}(t) = r_i$, $i \in M$.

Even Step: Find column multipliers $\mu_j(t)$, $j \in N$, such that $v_{ij}(t) = [v_{ij}(t-1)\, \mu_j(t)]$, and $x_{ij}(t) = [\![v_{ij}(t-1)\, \mu_j(t)]\!]$ satisfy the conditions $x_{Mj}(t) = c_j$, $j \in N$.

If the procedure terminates successfully at step $\bar{t}$, it outputs an apportionment X given by $x_{ij} = [\![\lambda_i\, v_{ij}\, \mu_j]\!]$, where $\lambda_i = \lambda_i(1)\, \lambda_i(2) \cdots \lambda_i(\bar{t})$ and $\mu_j = \mu_j(1)\, \mu_j(2) \cdots \mu_j(\bar{t})$.

6 Convex separable formulation of the biproportional apportionment problem

It is tempting, of course, to try to embed the Balinski and Demange procedure into an optimization approach. Following Carnal [14] and Helgason, Jörnsten, and Migdalas [29], Gaffke and Pukelsheim [21, 22] propose a problem formulation that is not restricted to standard rounding, but admits more general rounding rules. Any such rule equips an integer interval $[z-1, z]$ with a signpost $s(z)$. Gaffke and Pukelsheim analyze both the vector and matrix apportionment problems.

Generalizing the notion of proportionality in vector apportionment, the authors provide a definition of proportionality between a feasible apportionment matrix X and the corresponding vote matrix V that is based on some "critical inequalities". Let $\bar{\mathcal{A}}$ denote the set of $m \times n$ integer matrices X inheriting all zeros that appear in the vote matrix V, and let $\mathrm{supp}(V) = \{(i,j) |\ v_{ij} \neq 0\}$ be the support set of the vote matrix V. Note that $\mathrm{supp}(V)$ coincides with the set E of the edges in the graph $\mathcal{G}$ defined in Section 3. A *cycle* on $\mathrm{supp}(V)$ is a sequence of positive entries of V where two consecutive entries sharing the same row and two consecutive entries sharing the same column alternate, and the first entry in the sequence equals the last one.

Example 1 [22] Consider the following 3×3 matrix V

$$\begin{pmatrix} 0 & v_{12} & v_{13} \\ v_{21} & 0 & v_{23} \\ v_{31} & v_{32} & 0 \end{pmatrix}$$

A cycle is given by the following succession of positive entries: v_{12}, v_{13}, v_{23}, v_{21}, v_{31}, v_{32} v_{12}.

Given a matrix of votes V and a cycle on $\mathrm{supp}(V)$, the corresponding *critical inequality* is given by

$$\prod_{f\leq g} \frac{s(x_{i_f j_g})}{v_{i_f j_g}} \leq \prod_{f\leq g} \frac{s(x_{i_f j^*_g}+1)}{v_{i_f j^*_g}} \tag{6}$$

where the cycle is determined by the vectors of row and column indices $i_{(g)} = (i_1, i_2, \ldots, i_{g-1}, i_g)$, $j_{(g)} = (j_1, j_2, \ldots, j_{g-1}, j_g)$, $g \geq 2$, and the vector $j^*_{(g)} = (j_2, \ldots, j_{g-1}, j_g, j_1)$, i.e., the *cyclic permutation* of vector $j_{(g)}$.

The following theorem in [22] provides necessary and sufficient conditions for the existence of row and column multipliers for BAP.

Theorem 2 *Let $X \in \mathcal{A}$ be a feasible apportionment for V, r and c. Then X obeys the set of critical inequalities (6) for all cycles on* $\mathrm{supp}(V)$ *if and only if there exist row multipliers $\lambda_1, \ldots, \lambda_m > 0$ and column multipliers $\mu_1, \ldots, \mu_n > 0$ satisfying*

$$\frac{s(x_{ij})}{v_{ij}} \leq \lambda_i\, \mu_j \leq \frac{s(x_{ij}+1)}{v_{ij}} \qquad (i,j) \in \mathrm{supp}(V). \tag{7}$$

Then, the formulation of BAP as a convex integer minimization problem with linear constraints follows:

$$\min \quad \left\{ \prod_{(i,j):v_{ij}>0} \quad \prod_{z\leq x_{ij}:s(z)>0} \frac{s(z)}{v_{ij}} \ :\ x \in \mathcal{A} \right\} \tag{8}$$

The authors also provide the following optimality results.

Theorem 3 *Let $X \in \mathcal{A}$ be an apportionment for V, r and c. Then the following three statements are equivalent:*

1. *for all rows i and for all columns j there exist multipliers λ_i and μ_j such that $x_{ij} \in [\![\lambda_i\, a_{ij}\, \mu_j]\!]$;*
2. *X satisfies the critical inequalities for all cycles on* $\mathrm{supp}(V)$*;*
3. *X is an optimal solution of problem (8).*

Corollary 1 *(Multiple solutions) For every optimal apportionment matrix of problem (8), $X \in \mathcal{A}$, the following statements are equivalent:*

1. *there exists a matrix $Y \in \mathcal{A}$ such that $Y \neq X$;*
2. *there exists a cycle on* $\mathrm{supp}(V)$ *for which the critical inequality holds with equality.*

Corollary 2 *(Uniqueness) For every optimal apportionment matrix of problem (8), $X \in \mathcal{A}$, the following statements are equivalent:*

1. *the set $\mathcal{A}$ is a singleton, that is, $\mathcal{A} = \{X\}$;*
2. *for every cycle on* $\mathrm{supp}(V)$ *the critical inequality is strict.*

In [21] Gaffke and Pukelsheim take the logarithm of the objective function (8)

$$\min \quad \left\{ \sum_{(i,j):v_{ij}>0} \quad \sum_{z\leq x_{ij}:s(z)>0} \log \frac{s(z)}{v_{ij}} \ :\ x \in \mathcal{A} \right\} \tag{9}$$

and then they treat the problem as a piecewise linear separable transportation model.

Actually, the last formulation corresponds to a minimum cost flow problem defined over the bipartite graph $\mathcal{G}$ modified to have a set of parallel arcs $(i,j)^z$, $z = 1, 2, \ldots, H$, replacing the single arc (i, j). The capacity of each new arc is 1 and the cost coefficients are $c^z_{ij} = \log(z/v_{ij})$ (see, for example, [48]).

Alternatively, the problem can be formulated as a standard minimum cost flow (without parallel arcs) on a suitably modified graph G'. In order to get G' from $\mathcal{G}$ it suffices to introduce H copies of each node $i \in M$ and assign to each arc (ℓ, j) capacity 1 and cost $c^z_{ij} = \log(z/v_{ij})$, where ℓ is the z-th copy of node i in G'. W.r.t. the previous model, in this formulation the number of variables increases to $H \cdot |M| \cdot |N|$. In any case, the problem can be efficiently solved by standard min-cost flow algorithms, such as the successive path or cycle cancelling algorithms [1, 48].

In [21] problem (9) is analyzed in the more general framework of separable convex integer minimization problems under a set of linear equality restrictions with a totally unimodular matrix of coefficients [51].

Let $x = [x_e]_{e\in E}$ be a vector whose components are labeled by the elements e of a finite set E, and let $f_e(\cdot)$, $e \in E$, be real functions. Let A be a totally unimodular matrix with α rows and $\beta = |E|$ columns, $b \in \mathbb{Z}^\alpha$, and $\rho \in \mathbb{Z}^\beta$ a positive vector. Each function $f_e(\cdot)$, $e \in E$, is assumed to be convex in the interval $0 \le x_e \le \rho_e$. Then, the following separable convex integer minimization problem is formulated as follows

$$\begin{aligned} \min \quad & F(x) = \sum_{e\in E} f_e(x_e) \\ & A\,x = b \\ & 0 \le x \le \rho \\ & x \in \mathbb{Z}^\beta \end{aligned} \tag{10}$$

The authors assume that $f_e(\cdot)$, $e \in E$, are piecewise linear.

It is easy to check that BAP formulation given by (9) is a particular case of (10) where one has $\alpha = m + n$, $\beta = m \times n$.

Under the above assumptions, if the linear system

$$A\,x = b, \quad 0 \le x \le \rho$$

has a solution $x \in \mathbb{R}^\beta$, then it also has a solution $x \in \mathbb{Z}^\beta$. Hence, the authors observe that Fenchel duality can be applied to (10) despite the integer restrictions on the variables.

The authors provide a primal augmentation algorithm for (10) and also a dual algorithm. They show that the Balinski and Demange TT algorithm corresponds to such a dual algorithm when one has to solve problem (9). Actually TT may be viewed as an out-of-kilter algorithm [20, 42] for solving (9). Moreover, Gaffke and Pukelsheim formulate the (Fenchel) dual problem of (9) and discuss how the DAS procedure can be viewed as a cyclic coordinate ascent algorithm for solving this dual formulation. The dual variables of such a problem correspond to the row and column multipliers w.r.t. BAP. However, since the objective function of the dual problem is nondifferentiable, DAS might not converge to a maximizer and may stall at a nonoptimal solution. This situation is illustrated by the authors who provide some small examples to analyze the structure of stalling instances.

Practical instances point toward a peculiarity of the solution matrix X, which may feature *discordant seat assignments*, when held against the input vote matrix V. When comparing two cells (i, j) and (k, ℓ), that is, party j in district i and party ℓ in district k, it may happen that fewer votes go along with more seats, i.e., $v_{ij} < v_{k\ell}$ and $x_{ij} > x_{k\ell}$. Discordant seat assignments represent local adjustments that are unavoidable in order to achieve global biproportionality, as already observed by Gassner [25].

A particular irritation occurs when a single-seat district is struck by a discordant seat assignment, so that the one and only seat does not go to the district candidate who performed best, but to someone else who did less well. To overcome this obstacle, Maier [37] proposes a Winner-Take-One (WTO) amendment stipulating that in each district the strongest party is allocated at least one seat. The BAP formulations presented above are clearly powerful enough to support the additional district-wise WTO amendment.

7 Controlled rounding procedure

If one adopts the point of view that the fair share matrix would be the ideal seat assignment if only seats were allowed to be fractional, then it is natural to consider the actual integral seat assignment as an "error" with respect to the ideal fractional assignment. Then it makes sense to find an assignment that minimizes a certain measure of the error. This section and the next one describe approaches to BAP which explicitly exploit the idea of minimizing some given error.

The procedure devised by Cox and Ernst [16] is meant to round a matrix of rational numbers so that the row sums and the column sums of the rounded matrix are equal to pre-specified integer numbers. Although the authors investigate statistical problems, the rounding problem has some of the features of a BAP, as observed by Gassner [24]. The matrix to be rounded can be viewed as a matrix of quotas, like the fair share matrix or the regional quotas. We notice that in this case it is taken for granted that the seats are obtained only by rounding up or down the quotas.

Cox and Ernst formulate the following *Controlled Rounding Problem.* Given a real $m \times n$ matrix A such that

$$\sum_{j \in N} a_{ij} = \tilde{r}_i, \quad i \in M, \qquad \sum_{i \in M} a_{ij} = \tilde{c}_j, \quad j \in N$$

where $\tilde{r}_i$ and $\tilde{c}_j$ are not necessarily integers, a controlled rounding of A is a matrix X satisfying the following conditions:

$$\begin{array}{lll} (1) & \text{either } x_{ij} = \lfloor a_{ij} \rfloor \text{ or } x_{ij} = \lfloor a_{ij} \rfloor + 1 & i \in M,\ j \in N \\ (2) & \sum_{j \in N} x_{ij} \in \{\lfloor \tilde{r}_i \rfloor, \lfloor \tilde{r}_i \rfloor + 1\} & i \in M \\ & \sum_{i \in M} x_{ij} \in \{\lfloor \tilde{c}_j \rfloor, \lfloor \tilde{c}_j \rfloor + 1\} & j \in N \end{array} \tag{11}$$

The authors show that the above problem can be equivalently formulated as a nonlinear transportation problem even if $n+m+1$ additional variables must be introduced. In order to describe how the transformation works, we may consider w.l.o.g. the simplified version of the optimal controlled rounding problem under the condition

$$0 \le a_{ij} < 1, \quad i \in M, j \in N. \tag{12}$$

Consider the constraints (11-2) which can be rewritten as:

$$\begin{aligned}
\sum_{j=1}^{n} x_{ij} &= \lfloor \tilde{r}_i \rfloor + y_i, \qquad && i \in M \\
\sum_{i=1}^{m} x_{ij} &= \lfloor \tilde{c}_j \rfloor + z_j, \qquad && j \in N
\end{aligned}$$

where y_i and z_j are binary variables. The above system of linear equations can be rewritten as the set of constraints of a (balanced) capacitated transportation problem with $m+1$ origins and $n+1$ destinations, where $x_{i,n+1} := 1 - y_i$ and $x_{m+1,j} := 1 - z_j$:

$$\begin{aligned}
&\sum_{j=1}^{n+1} x_{ij} = \lfloor \tilde{r}_i \rfloor + 1, && i = 1, \ldots, m \\
&\sum_{j=1}^{n+1} x_{m+1,j} = \sum_{j=1}^{n} (\lfloor \tilde{c}_j \rfloor + 1) - \left\lfloor \textstyle\sum_j \tilde{c}_j \right\rfloor \\
&\sum_{i=1}^{m+1} x_{ij} = \lfloor \tilde{c}_j \rfloor + 1, && j = 1, \ldots, n \\
&\sum_{i=1}^{m+1} x_{i,n+1} = \sum_{i=1}^{m} (\lfloor \tilde{r}_i \rfloor + 1) - \lfloor \textstyle\sum_i \tilde{r})i \rfloor \\
&0 \leq x_{ij} \leq 1, && i = 1, \ldots, m+1; \quad j = 1, \ldots, n+1.
\end{aligned} \tag{13}$$

The authors show that a feasible solution always exists and, in view of the Integrality Theorem of Network Flows, also a binary feasible solution always exists. Furthermore they extend their results to the case where some entries of the matrix X must have a fixed integer value. This is particularly relevant for BAP, because of the presence of the zero vote set Z for which the corresponding element in X must be 0.

Since the solution of (13) is not unique in general, one can search for an *optimal* controlled rounding of A by minimizing either the L_p-norm

$$\sum_{i \in M} \sum_{j \in N} \left(|x_{ij} - a_{ij}|^p \right)^{\frac{1}{p}}, \qquad 1 \leq p < \infty,$$

or, the L_∞-norm

$$\max \{ |x_{ij} - a_{ij}| : i \in M, j \in N \}.$$

For the L_p norm they adopt a standard device for functions of binary variables, which consists of linearly interpolating the function values at 0 and at 1, thus obtaining a linear function on $[0, 1]$. Again, the Integrality Theorem of Network Flows guarantees integrality of the linear optimum. For the L_∞ norm a more complex linearization is suggested.

It is important to remark that limiting the seat values to either rounding down or up the fair shares introduces a constraint which may cut off the true solution minimizing either the L_1-norm or the L_2-norm over all possible apportionments. Consider the following example with A an $(n+2) \times (n+2)$ matrix of fair share quotas

$$A = \begin{pmatrix} \frac{n-1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} & \frac{1}{n} \\ \frac{1}{n} & \frac{n-1}{n} & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{1}{n} & 0 & \cdots & \frac{n-1}{n} & 0 \\ \frac{1}{n} & 0 & \cdots & 0 & \frac{n-1}{n} \end{pmatrix}, \qquad r = (\,2 \quad 1 \quad \cdots \quad 1\,), \qquad p = (\,2 \quad 1 \quad \cdots \quad 1\,).$$

There are essentially three apportionments up to permutation of the indices $\{2, \ldots, n+2\}$, namely

$$X^1 = \begin{pmatrix} 2 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{pmatrix}, \quad X^2 = \begin{pmatrix} 1 & 1 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{pmatrix}, \quad X^3 = \begin{pmatrix} 0 & 1 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

for which we have

$$\|X^1 - A\|_1 = 4 + \frac{4}{n}, \qquad \|X^2 - A\|_1 = 6 - \frac{2}{n}, \qquad \|X^3 - A\|_1 = 10 - \frac{10}{n}$$

and

$$\|X^1 - A\|_2^2 = 1 + \frac{5}{n} + \frac{4}{n^2}, \qquad \|X^2 - A\|_2^2 = 3 - \frac{3}{n} + \frac{4}{n^2}, \qquad \|X^3 - A\|_2^2 = 7 - \frac{11}{n} + \frac{4}{n^2}$$

Hence for $n > 4$ the optimal apportionment (both for the L_1 and the L_2-norm) is X^1, with x^1_{11} outside the range $\{0, 1\}$. If we restrict the apportionments to $\{0, 1\}$, then the optimal apportionment is X^2.

We may therefore wonder whether it is possible to solve efficiently the L_1 and L_2 minimization without the restriction of finding seats within $\{\lfloor a_{ij} \rfloor, \lceil a_{ij} \rceil\}$. The answer is affirmative thanks to the properties of network flows. As for the minimization of the L_2 norm we direct the reader to a result by Minoux [40].

For the L_1 minimization we sketch here a simple procedure. The main idea is to replace (if a_{ij} is not integral) each function $f_{ij}(x) := |x - a_{ij}|$, which is convex piecewise linear but has a breakpoint at the fractional value a_{ij}, with the convex piecewise linear function

$$f'_{ij}(x) = \begin{cases} a_{ij} - x & \text{if } x \le \lfloor a_{ij} \rfloor \\ (1 - 2 \; <a_{ij}>)\,(x - \lfloor a_{ij} \rfloor) + <a_{ij}> & \text{if } \lfloor a_{ij} \rfloor \le x \le \lceil a_{ij} \rceil \\ x - a_{ij} & \text{if } x \ge \lceil a_{ij} \rceil \end{cases}$$

where $<a> := a - \lfloor a \rfloor$ is the fractional part of a. The functions $f(x)$ and $f'(x)$ concide at all integral points x and therefore we may replace $f(x)$ with $f'(x)$ since we are interested only in integral values of x. But now $f'(x)$ has breakpoints at integral values and network flow techniques can be easily applied to produce integral values.

The same technique might be applied to any convex objective function, by sampling the function at the integral points and building an equivalent (on the integral points) convex piecewise linear function. However, a question arises. The number of breakpoints might grow in a nonpolynomial way (different from the L_1 case). The trick devised by Minoux [40] just overcomes this difficulty.

8 Minimax approximation of target quotas

Serafini and Simeone [52] approach BAP by focusing on the minimization of the maximum error. They do not make any assumption on the quotas q_{ij} to which the apportionment should be as proportional as possible, apart from the obvious requirement that $\sum_{ij} q_{ij} = H$ and $v_{ij} = 0$ implies $q_{ij} = 0$. These "target" quotas could be the fair shares or the regional quotas or any other type of quotas defined by the electoral system.

They define the error w.r.t. the target quotas in assigning the actual seats in two alternative ways. The *absolute error* τ and the *relative error* σ are defined as

$$\tau := \max_{ij} |x_{ij} - q_{ij}|, \qquad \sigma := \max\{ \max_{(ij)\notin Z} \frac{x_{ij} - q_{ij}}{q_{ij}} \, ; \max_{(ij)\notin Z, q_{ij}\geq 1} \frac{q_{ij} - x_{ij}}{q_{ij}} \}$$

The approach proposed in [52] calls for finding a feasible apportionment minimizing either the absolute error or the relative error. If the absolute error is minimized, the best approximation problem is formulated as follows:

$$\begin{aligned} \min \quad & \tau \\ & q_{ij} - \tau \leq x_{ij} \leq q_{ij} + \tau \qquad i \in M,\, j \in N,\, (ij) \notin Z \\ & x \in \mathcal{A} \end{aligned} \tag{14}$$

while, when the relative error is considered, the formulation is:

$$\begin{aligned} \min \quad & \sigma \\ 0 \ \leq\ & x_{ij} \leq (1+\sigma)\, q_{ij} \qquad i \in M,\, j \in N:\ q_{ij} < 1,\, (ij) \notin Z \\ (1-\sigma)\, q_{ij} \ \leq\ & x_{ij} \leq (1+\sigma)\, q_{ij} \qquad i \in M,\, j \in N:\ q_{ij} \geq 1,\, (ij) \notin Z \\ & x \in \mathcal{A} \end{aligned} \tag{15}$$

W.r.t. formulations (14) and (15), the authors note that, given a bound $\tau > 0$ on the absolute error or a bound $\sigma > 0$ on the relative error, both problems can be modeled as a feasible flow problem with lower and upper arc capacities on the network $\mathcal{G}$. If the absolute error is adopted, each arc (i, j) has a capacity interval given by

$$[c^-_{ij} \ , \ c^+_{ij}] := [\lceil q_{ij} - \tau \rceil^+ \ , \ \lfloor q_{ij} + \tau \rfloor],$$

where, by definition $a^+ := \max\{a, 0\}$. If one measures the relative error, each arc (i, j), $i \in M$, $j \in N$, has a capacity interval

$$[c^-_{ij} \ , \ c^+_{ij}] := \begin{cases} [\lceil (1-\sigma)\, q_{ij} \rceil^+ , \lfloor (1+\sigma)\, q_{ij} \rfloor] & \text{if } q_{ij} \geq 1 \\ [0, \lfloor (1+\sigma)\, q_{ij} \rfloor] & \text{if } q_{ij} < 1. \end{cases}$$

A feasible flow x_{ij} satisfies $c^-_{ij} \leq x_{ij} \leq c^+_{ij}$ and, by flow properties, if there is a feasible flow x there is also an integral flow since the capacity values are integers. Hence one wants to find the minimum value for τ or for σ such that a feasible flow exists. The existence of a feasible flow can be easily established through a max-flow problem.

	A	B	C	D	E	F
1	0.992	0.870	0.170	0.994	0.988	0.986
2	0.460	0.580	0.991	0.993	0.989	0.987
3	0.001	0.001	0.001	0.986	0.001	0.010
4	0.001	0.001	0.001	0.440	0.001	0.556
5	0.001	0.001	0.001	0.140	0.856	0.001

Table 1: An example of a matrix Q of regional quotas.

By the integrality of the x_{ij}'s only a finite number of values for τ are relevant to the solution, namely, those for which either $q_{ij} - \tau$ or $q_{ij} + \tau$ is integral for some (i, j). Similarly, one can define the relevant values for the relative error minimization. The number of relevant values is at most $(H + |M|) \cdot |N|$.

For the absolute and relative errors the authors provide some useful error bounds, some of which are exploited in the design of algorithms for the minimization of the absolute error.

Clearly if the target quotas are the fair shares, the optimal absolute error τ^* is bounded above by 1. In this case the number of relevant errors to be checked is at most $|M| \cdot |N|$, i.e., a polynomial bound. However, if other quotas are used, like the regional quotas, there are no "natural" bounds. The authors show that if the seats c_j are assigned by the rule of Largest Remainders applied to the vector $\sum_i v_{ij}$ (like in the Italian system) and $v_{ij} > 0$ for all i, j, then $\tau^* < 2$. The authors also provide the following example that shows that there are instances with $\tau^* > 1$ under the same assumptions.

Consider the matrix q given in Table 1 (parties A–F, regions 1–5) with $r = (5\ 5\ 1\ 1\ 1)$ and $c = (1\ 1\ 1\ 4\ 3\ 3)$. Computing the column-wise sums of q_{ij}, one has

$$\sum_{i \in M} q_{ij} = (\, 1.455 \quad 1.453 \quad 1.164 \quad 3.553 \quad 2.835 \quad 2.540 \,)$$

thus implying $c = (1\ 1\ 1\ 4\ 3\ 3)$ by the method of Largest Remainders.

Rounding down the regional quotas in the above matrix, one always gets 0 and rounding them up one always gets 1. One can check that there is no way to assign 0 or 1 seats to each pair (i, j). Indeed the parties D, E and F would receive at most 6 seats altogether in the regions 1 and 2. Hence the parties A, B and C would receive at least $10 - 6 = 4$ seats in the same regions 1 and 2. But these parties are allotted 3 seats in total! So at least one party among D, E and F should receive two seats either in region 1 or 2. For a minimax solution this has to be for party D in region 1 with optimal error equal to 1.006. The seat assignment in Table 2 corresponds to one of these solutions.

Serafini and Simeone design three algorithms to find a solution minimizing the maximum absolute error taking into account both the computational complexity and the simplicity of implementation. Noting that the size of problems to be solved is never large, they point out that speed of computation can be reasonably exchanged in favor of simplicity of the description in the law and implementation itself.

As already pointed out, only a finite number of errors are relevant to the solution and this number is at most $(H + |M|) \cdot |N|$. At first sight it might seem that it is enough to carry out a binary search over this set in order to find the smallest relevant error such that the network flow problem admits a

	A	B	C	D	E	F
1	1	0	0	2	1	1
2	0	1	1	1	1	1
3	0	0	0	1	0	0
4	0	0	0	0	0	1
5	0	0	0	0	1	0

Table 2: A minimax solution to the best approximation of matrix Q of Table 1.

solution. However a naive implementation of the binary search requiring the sheer calculation (without taking into account sorting) of all relevant errors calls for an execution time linearly dependent on H and thus pseudopolynomial. Therefore binary search can be used but with some caution.

We refer the reader to [52] for details on the three algorithms. We limit ourselves here to say that the first algorithm is simple but runs in pseudopolynomial time; the second algorithm is more complex but runs in polynomial time and the third algorithm is strongly polynomial at the expense of being a complex three-stage algorithm. The authors also provide a (weakly) polynomial algorithm for the minimization of the maximum relative error. They avoid the trap of the pseudopolynomial number of relevant errors by using binary search on the relevant errors without the need of computing all of them. See [52] for details.

The authors are also concerned with two other practical and important issues: uniqueness of the optimal apportionment and possibility of providing the layman with a certificate of optimality.

As for the first problem it is clear that any sound seat assignment method that takes as input the votes, must output a unique apportionment. On the other hand, optimization problems usually admit multiple optimal solutions. Therefore it is crucial to develop a method that outputs a unique apportionment.

One way to overcome the difficulty of nonunique solutions consists of finding unordered lexicominima, as defined in Schrage [50]. For details, the reader is referred to [52]. In this case the vectors to be ordered consist of the absolute errors for all pairs (ij). In order to find unordered lexicominima, once a minimax solution has been found with relevant error τ_k for the *blocking pair* $(ij)_k$ (k is the index of the ordered relevant τ's), a solution minimizing

$$\max_{(ij)\neq(ij)_k} |x_{ij} - q_{ij}|$$

is found. This can be done as before with the only difference that the capacity for the pair $(ij)_k$ is no longer changed. Once a second solution with error τ_h ($h < k$) and blocking pair $(ij)_h$ has been found, one proceeds recursively by fixing the capacities of the blocking pairs one at a time. If for the current relevant τ, $\tau < 1/2$ holds, one simply fixes the capacity interval for the arc (i, j) to $[\bar{q}_{ij}, \bar{q}_{ij}]$ with $\bar{q}_{ij}$ equal to q_{ij} rounded to the nearest integer and the computation is finished because there cannot be any better error.

As for the second problem, it can be argued that sound assignment procedures available in the literature are generally too complex to be fully understood by the general public. A voting system cannot be based on the simple trust that the persons involved in the computations are honest and do not make mistakes. Therefore, a way to check the election outcome which does not call for difficult mathematical concepts should be provided. Serafini and Simeone [52, 53] point out that it is possible to exploit the Max Flow-Min Cut Theorem [1] to get a certificate of optimality whereby anybody can check through simple elementary calculations that no solution can be better than the given one. As a hint of how such a certificate can work we refer to the previous discussion showing the impossibility of rounding down or up the quotas in Table 1.

At the end of the sections devoted to BAP, it may be useful to compare the different approaches to BAP. In particular we compare (i) Balinski and Demange's Tie-and-Transfer with rounding to the closest integer (TT), (ii) Pukelsheim's Discrete Alternating Scaling (DAS), (iii) Gaffke and Pukelsheim's algorithm, in the Rote and Zachariasen's minimum cost flow implementation (GFRZ), (iv) Cox and Ernst's Controlled Rounding (CR), and finally, (v) Serafini and Simeone's minimax approximation of target shares (MM). Note that the first three algorithms yield, for a given rounding method, the same apportionment, namely, the unique one that satisfies all the six proportionality axioms.

We have taken into consideration the following criteria.

1. Finiteness (the algorithm stops after a finite number of steps): there are rare pathological instances where DAS does not converge;
2. Feasibility (the output seat assignment always yields an apportionment): CR does not guarantee feasibility for general quotas;
3. Soundness (satisfaction of Balinski and Demange's six integral proportionality axioms): only TT, DAS and GFRZ satisfy the axioms;
4. Uniqueness (uniqueness of the seat assignment output by the method): the optimal solution of CR may not be unique; MM exhibits a unique solution with the lexicomin refinement;
5. Theoretical Complexity (worst-case rate of growth of the number of elementary operations as the instance size increases): TT is pseudopolynomial in general and weakly polynomial in Zachariasen's [55] implementation; the complexity of DAS is unknown; GFRZ is weakly polynomial; MM and CR are strongly polynomial;
6. Generality (range of applications besides Biproportional Apportionment): MM can be applied to other problems (see [52]);
7. Flexibility (dependence on other parameters besides input data): the freedom of choice of the rounding method in TT, DAS, and PGRZ is counter-balanced by the freedom of choice of the target quotas in CR and MM;
8. Ease of implementation (no need to write sophisticated ad hoc computer codes): DAS has perhaps the simplest implementation and TT the most sophisticated;
9. Transparency (possibility of translating the procedure into a simple, easy-to-understand, electoral law): the simplest version of MM can be stated very easily;
10. Certifiability (easy verifiability by a layman, through simple elementary operations, that the seat assignment output by the method satisfies the claimed requirements, like feasibility, optimality): MM seems to be the only method with this feature (see [53]).

9 Vote transitions

Published election data include the marginal distributions of votes cast at two successive elections. There is a strong political interest in the estimation, on the basis of these data and possibly of local surveys or exit polls, of vote transitions between parties at the two elections. To this purpose, a large variety of statistical methods are available: among them, what is known as Goodman's ecological regression [27, 36]; the compound multinomial model of Brown and Payne [13] and its reformulation as a multivariate generalized linear model in [19]; the quadratic programming estimation model of [54]; entropy maximization models [32, 33]. Here, for the purposes of this survey and for the sake of illustration, we restrict ourselves to present the nonlinear transportation model for vote transition estimation discussed in [10]. Let I be the set of parties in the first election and J the set of parties in the subsequent election. The sets I and J do not necessarily coincide. Moreover, one can keep track of new voters, lost voters, and abstentionism by the introduction of dummy parties in I or in J. Let a_i be the number of votes received by party $i \in I$ in the first election at national level and b_j the number of votes received by party $j \in J$ in the second election, again at the national level. The unknown data to be inferred are the values x_{ij}, defined as the number of voters who in the first election voted for party i and in the next election voted for party j. Clearly these values must satisfy the following transportation constraints:

$$\begin{array}{lll} \sum_{i\in I} x_{ij} = b_j, & j \in J, & \\ \sum_{j\in J} x_{ij} = a_i, & i \in I, & \\ x_{ij} \geq 0, & i \in I, & j \in J \end{array} \tag{16}$$

Let y_{ij} be the probability that somebody votes for party i in the first election and for party j in the next election. Such probabilities are estimated through a loglinear regression model from a sample of empirical frequencies y_{ij}^k in each electoral district $k \in K$ - where K is the set of all electoral districts - along with the values $z_{1k}, \ldots, z_{dk}$ of certain socio-economic variables observed in district k.

Then one obtains the desired vote transitions x_{ij} by running the RAS algorithm (see Section 3) on the starting matrix y_{ij}, so as to fit the marginals a_i and b_j. According to a well-known result of [12], the matrix x_{ij} is the unique optimal solution to the entropy maximization problem

$$\max\ -\sum_{i\in I}\sum_{j\in J} x_{ij}\ \log\frac{x_{ij}}{y_{ij}}, \tag{17}$$

subject to (16).

Hence x_{ij} is the optimal solution to a transportation problem with concave separable objective function, to be maximized.

The choice of the objective function in (17) has the effect that the optimal solution is the "most likely" vote transition matrix fitting the marginals a_i and b_j, conditional on the probabilities y_{ij}.

Johnston and Hay ([32], see also [33]) argue that, for the purposes of post-electoral analysis, the estimate of the vote transition matrix x_{ij} at the national level should be supplemented by detailed information about the disaggregated matrices x_{ij}^k, similarly defined for each district k. For the actual computation of the 3-dimensional array x_{ij}^k they propose, as a natural extension of (17) with constraints

(16), a 3-dimensional transportation model whose objective function, to be maximized, has again the meaning of an entropy. Out of the three marginals b_{ik}, a_{jk}, x_{ij} (corresponding to summations over the index i, j, k, respectively) the first two are typically known from election records, while the third one, x_{ij}, must be estimated, e.g. through the solution of (17) with constraints (16). Although such model brings us outside the realm of network flows, it is worth pointing out that one can still find an optimal solution via a straightforward 3-dimensional generalization of the RAS algorithm (see [11]).

10 Political districting

A transportation procedure appears in Hess et al. [30], the earliest operations research paper in political districting. First the districting problem is formalized as a discrete location problem. Each territorial unit must be assigned to exactly one center and all units assigned to the same center form a district. Let n be the total number of territorial units and k be the number of districts. The political districting model is the following:

$$
\begin{aligned}
\min \quad & \sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij}^2\, p_i\, x_{ij} & \\
& \sum_{j=1}^{n} x_{ij} = 1 & i = 1, \dots, n \\
& \sum_{j=1}^{n} x_{jj} = k & \\
& a\,\bar{P}\,x_{jj} \le \sum_{i=1}^{n} p_i\, x_{ij} \le b\,\bar{P}\,x_{jj} & j = 1, \dots, n \\
& x_{ij} \in \{0, 1\}, & i, j = 1, \dots, n
\end{aligned}
$$

where x_{ij} is a binary variable equal to 1 when unit i is assigned to center j, p_i is the population of unit i, d_{ij} is the distance between unit i and center j, and a and b are the minimum and the maximum allowable district population fractions, calculated as a percentage of the average district population $\bar{P}$ (total population divided by k). Moreover, the variable x_{jj} takes the value 1 whenever unit j is chosen as one of the centers. The first n constraints mean that each unit must belong exactly to one district. The next one imposes that the total number of districts is exactly k. The other two groups of n constraints represent the conditions on the maximum and the minimum allowable district population, with respect to chosen parameters $a < 1$ and $b > 1$ (population equality). Finally, the objective function (total inertia) is a measure of compactness.

Due to the computational difficulty in the solution of the above model, in [30] an iterative heuristic procedure is proposed as an alternative solution approach. Essentially, the generic iteration of the algorithm consists of five steps: (1) guess the district centers; (2) solve a transportation problem to assign population equally to these centers at minimum cost (defined in terms of distances between units and centers of the districts); (3) adjust the solution of the transportation problem so that each territorial unit is entirely within one district; (4) compute centroids of the current districts and use them to update the district centers; (5) repeat from step (1) until the procedure converges (i.e., the centers do not change in two successive iterations).

The main step of the above procedure is step (2) in which a transportation problem must be formulated and solved. The formulation of the problem is the following. The set of origins in the transportation graph represents the current centers, all with supplies equal to $\bar{P}$. The set of destinations represents the territorial units, with demands equal to their population. Each edge (i, j) of the graph has a weight equal to d_{ij}^2.

In the above iterative procedure it may happen that in the solution of the transportation problem a territorial unit i is split between two or more districts; in this case, in step (3) i is entirely assigned to the district to which the largest quota of its population was assigned. The convergence of the procedure is not guaranteed in theory. However, the authors report that, in real-life applications, the heuristic converges to a local minimum in fewer than ten iterations (that is, ten transportation problems must be solved).

Following the approach of Hess et al., other authors developed political districting methods related to network flows. The procedure proposed in [31] differs from the previous one in the first and in the third step. Instead of adopting an iterative strategy based on successive adjustments of the centers, Hojati locates them only once at the beginning of the procedure and this choice is permanent. To solve this problem, the author introduces a (mixed integer) warehouse location model, similar to the one in [30], but based on two different sets of variables, namely, x_{ij}, $i, j = 1, \ldots, n$, representing the proportion of population of unit i assigned to district j, and indicator variables y_j, $j = 1, \ldots, n$, such that $y_j = 1$ if unit j is chosen as the center of a district and $y_j = 0$ otherwise. A Lagrangian relaxation of the resulting model is derived and is solved by a subgradient optimization algorithm.

After step (2), when in the solution of the transportation problem there are split territorial units (i.e., units fractionally assigned to more than one center) Hojati introduces the *Split Resolution Problem* (SRP) which is formulated as a graph-theoretic model. Actually, he takes into consideration the subgraph of the transportation graph whose vertices are given, on the one hand, by the split units and, on the other hand, by those centers to which some split units have been (partially) assigned. The author shows that SRP can be solved by a sequence of capacitated transportation problems defined over a suitable modified network (see [31]).

The procedure proposed in [26] basically follows the iterative location/allocation approach pioneered by Hess et al., but with the main difference that a new method for assigning territorial units to districts is adopted. For this step, the authors introduce a minimum cost network flow problem defined on the following network. The nodes of the network are the territorial units, the district centers and an additional sink node t. Each unit-node i has a supply equal to p_i, while the sink-node has a demand equal to $P = \sum_{i=1}^{n} p_i$. Beside the arcs (i, j) corresponding to all the unit-center pairs, $i = 1, \ldots, n$, $j = 1, \ldots, n$, there exists an arc (j, t) in the network for each district center j.

The authors introduce different cost functions to define the costs associated with the arcs of the network (see, [26], page 20, Table 1) with the aim of modelling additional political districting issues that were not taken into account in the original model of Hess et al. Their iterative procedure stops when the difference between the optimal value of two successive solutions of the minimum cost flow problem is sufficiently small. The authors notice that split units can still arise and for this problem they suggest following the same rule adopted by Hess et al. [30].

11 The Give-up Problem

Ricca, Scozzari and Simeone [47] discuss network flow techniques for the decision problem related to seat "give-ups" by multiple winners in a system with closed (blocked) lists. This problem is of particular interest in Italy where multiple winners may arise both in the election of the Chamber of Deputies and of the Senate. In fact, the current Italian electoral law requires that each party presents, in each district, an *ordered* list of candidates. It also allows for the same candidate to be present in more than one list. Voters can cast their ballots for parties, but not for candidates. If a party receives w seats in a district, the winners of that party will be exactly the first w of its list in that district, but if a candidate is a winner in more than one district, he or she must give up all the seats won but one. The decisions about give-ups are usually centralized. Clearly, central decisions must be based on inter-district comparisons of preferences.

Then, for a given party, the *Give-up Problem* can be formulated as finding a set of give-ups consistent with the inter-district system of preferences of that party. To this purpose, the authors introduce two classes of models, i.e., "utility" and "ordinal" ones, and show that for both of them some natural formulations of the above Give-up Problem can be efficiently solved by network flow techniques.

A strict linear order $\succ$ is defined over the set of candidates. Ordinal models rely exclusively on order relations between candidates w.r.t. $\succ$. In utility models for each district k and each candidate i in the list of that district, a *disutility* or *cost* c_{ki} of letting i win in district k is defined and the objective is the minimization of a cost function (or equivalently the maximization of the total utility of a party).

An instance of the Give-up Problem refers to a single political party and it is defined by three integers n, m and S, with $S \leq n$, and by a four tuple $(C, R, list, seat)$, where:

- C is a set of n candidates;
- R is a set of m regions;
- $list = \{L_1, \dots, L_k, \dots, L_m\}$ is a set of m regional ordered lists of candidates, $L_k \subseteq C$, $\forall k = 1, \dots, m$, and $|L_k|$ is the length of the list L_k;
- $seat = (s_1, \dots, s_k, \dots, s_m)$ is a vector of integers, where, for all k, s_k $(1 \leq s_k \leq |L_k|)$ denotes the number of seats obtained by the party in the k-th region, with $S = s_1 + s_2 + \cdots + s_m$.

The authors assume the following *hypothesis of consistency*: if $i \succ j$ then i precedes j in all the lists where both i and j compete. Similarly, a cost matrix $[c_{ki}]$, $k \in R$, $i \in C$ is said to be consistent if $i \succ j$ implies $c_{ki} < c_{kj}$, for all $k \in R$.

A *feasible seat assignment* x is an assignment of seats to candidates such that (i) each candidate gets at most one seat; (ii) the number of candidates who win a seat in district k is exactly s_k. Then, a feasible seat assignment can be described by a binary matrix $[x_{ki}]$ such that (i) $\sum_{k \in R} x_{ki} \leq 1$, $i \in C$; (ii) $\sum_{i \in C} x_{ki} = s_k$, $k \in R$.

In general, one would like the final set of winners to be concentrated in the top part of the ranking given by $\succ$. This broad goal can be formalized in several ways. The simplest formulation is to find a feasible seat assignment whose winners are precisely the first S candidates in the linear order $\succ$. A feasible assignment satisfying this property will be called *perfect*. However, such assignment may not exist as the following example shows.

Example 2 Suppose that in a party there are six candidates, $C = \{1, 2, 3, 4, 5, 6\}$, and that $S = |R| = 3$, $s_k = 1$, for $k = 1, 2, 3$. The party presents the following three lists in three districts: $L_1 = \{1, 2, 3\}$, $L_2 = \{1, 2, 6\}$ and $L_3 = \{4, 5, 6\}$. It is easy to check that no *perfect* set of winners exists. Here, there is no way to assign the $S = 3$ seats to the first 3 candidates, since all these candidates can only receive a seat either in district $k = 1$, or in district $k = 2$.

The first result in [47] is the characterization of a perfect seat assignment as a feasible flow in a suitable network.

Consider the bipartite graph (R, C, E), where the two sides correspond to the regions R and the complete set of candidates C, respectively, and an arc (k, i) exists if and only if candidate i is included in list L_k. The edge-set is denoted by E. Now direct all the edges in E from R to C; add a source s and a sink t; then connect the source s to each region-node $k = 1, \dots, m$, and each candidate-node $i = 1, \dots, n$ to the sink t. Let $\mathcal{N} = (V, E')$ be the resulting network, with $V = \{s\} \cup R \cup C \cup \{t\}$ and $E' = E \cup \{(s, k) : k \in R\} \cup \{(i, t) : i \in C\}$. Assign to each arc (s, k), $k = 1, \dots, m$, both an upper and a lower capacity equal to s_k; assign to all the other arcs in $\mathcal{N}$ a lower capacity and an upper capacity equal to 0 and 1, respectively.

Consider the subnetwork $\mathcal{M}$ of $\mathcal{N}$ induced by the subset of nodes $\{s\} \cup R \cup J_S \cup \{t\}$, where J_S is the set of the first S candidates in C.

Proposition 1 A perfect seat assignment exists if and only if there exists a feasible flow in the network $\mathcal{M}$ with the above defined lower and upper capacities.

It is well known (see, e.g. [1]) that a feasible flow in a network with lower and upper capacities can be found, if it exists, through the solution of a maximum flow problem. Indeed, this is the main technical difficulty in solving the existence problem for a perfect assignment.

In order to manage the problem when a perfect assignment does not exist, as a first possibility, the authors introduce the formulation of an ordinal model.

For any given instance of the Give-up Problem, and for a given ranking $\succ$, the *height* of a feasible assignment is defined as the smallest positive integer h such that all the candidates after the h-th in $\succ$ get no seats. One then looks for a feasible assignment minimizing the height.

Another option is to define a notion of *lexicographically best* assignment. For a given feasible assignment x, let $\mathcal{I}(x) = \{i_1(x), \dots, i_n(x)\}$ be a binary indicator vector such that $|\mathcal{I}(x)| = n$ and for $\nu = 1, \dots, n$, $i_\nu(x) = \sum_{k \in R} x_{k\nu}$. Let x and y be two feasible assignments, $x \neq y$, and let $\mathcal{I}(x)$ and $\mathcal{I}(y)$ be the associated binary indicator vectors. One says that x is *lexicographically better* than (or *dominates*) y if $\mathcal{I}(x)$ is lexicographically greater than $\mathcal{I}(y)$. A feasible assignment x is called *lexicographically best* if it is not dominated by any other feasible assignment.

The authors show that one can find a feasible assignment minimizing the height by solving a *bottleneck transportation* problem. As a consequence, one can solve the height minimization problem in strongly polynomial time by $O(\log(n))$ max-flow computations. They also show that a lexicographically best assignment is a minimum height assignment.

As an alternative, one can find an optimal solution to a utility model. Let $[c_{ki}]$ be a cost matrix, where $c_{ki} \geq 0$ is the cost of assigning a seat in region k to candidate i. Then, the *cost* of a feasible seat assignment x is given by:

$$c(x) = \sum_{k \in R, i \in C} c_{ki} x_{ki} \tag{18}$$

A cost matrix is defined to be *consistent* if $i \succ j$ implies $c_{ki} < c_{kj}$, for all $k \in R$, and to be *uniform* if, for every candidate $i \in C$, $c_{ki} = c_{ri}$ for all $k \neq r \in R$. A utility model with consistent and uniform cost matrix is called a *score function model.* In particular, the uniformity assumption on the cost matrix implies that, for any given $i \in C$, one has $c_{ki} = \gamma_i$ for all $k \in R$. Let Z be the sum of the γ_i's over the first S candidates $i \in C$ according to $\succ$.

In order to formulate a utility model, the authors consider the network $\mathcal{N} = (V, E')$, and introduce a nonnegative cost function $c : E' \to R_0^+$ that assigns a cost to each arc in E'. The costs on the arcs $\{(s,k) : k \in R\} \cup \{(i,t) : i \in C\}$ are all equal to zero, while, for all $k \in R$ and $i \in C$, the cost on the arc (k,i) is equal to the corresponding c_{ki}.

A best seat assignment w.r.t. a utility model can be found by solving a minimum cost flow problem on the network $\mathcal{N}$ [1].

The following proposition provides an alternative characterization of a perfect seat assignment as an optimal solution of a minimum cost flow problem on $\mathcal{N}$.

Proposition 2 A perfect seat assignment exists if and only if, for any score function model, there exists an optimal solution x^* to the corresponding minimum cost flow problem on $\mathcal{N}$ whose total cost is $c(x^*) = Z$.

The central result in [47] establishes strong relations between the optimal solution of a score function model, a minimum height assignment, and a lexicographically best one. The result exploits the notion of "illegitimate path". Given an instance of the Give-up Problem and the corresponding bipartite graph (R, C, E), for any given feasible assignment x an *illegitimate* path w.r.t. x is an even path from a non-winner candidate i to a winner candidate j, with $i \succ j$, and formed alternately by edges with flow 0 and flow 1 in x.

Theorem 4 *Given an instance of the Give-up Problem and a feasible seat assignment x, the following four statements are equivalent:*

1) x is an optimal assignment for every score function model;

2) x is an optimal assignment for some score function model;

3) there is no illegitimate path w.r.t. x;

4) x is lexicographically best.

The notion of illegitimate path can be also used to develop an algorithm for finding a best assignment w.r.t. a score function model. Actually, starting from a feasible flow on $\mathcal{N}$, at each step the algorithm either finds an illegitimate path, or it stops with an optimal solution. Let y be any feasible seat assignment. Then, since for a given non-winner i in y, in (R, C, E) there exist at most $|E'|$ different illegitimate paths starting from i, a best assignment w.r.t. a score function model can be found in time $O(|E'|n)$ by the above algorithm.

The authors note that in many situations imposing a strict linear order on the set of candidates might be too restrictive. Thus, they also take into account the possibility that, given any two candidates i and j, neither of them is better than the other, since i and j could be regarded by the party as "indifferent". In this case, one has a "ranking with ties" of the candidates: this concept is captured by the formal notion of *weak order*, i.e., a complete and transitive relation $\succeq$. The authors show that the main results already obtained for strict linear orders can be generalized to weak orders, provided that the basic definitions and constructions are suitably modified (see, [47]).

References

[1] R.K. Ahuja, T.L. Magnanti, and J.B. Orlin, Network flows: Theory, algorithms and applications, Prentice Hall, New Jersey, 1993.

[2] M. Bacharach, Estimating nonnegative matrices from marginal data, International Economic Review 6 (1965), 294–310.

[3] M. Bacharach, Biproportional matrices & input-output change, Cambridge University Press, Cambridge, UK, 1970.

[4] A. Bachem and B. Korte, On the RAS algorithm, Computing 23 (1979), 189–198.

[5] M.L. Balinski and G. Demange, An axiomatic approach to proportionality between matrices, Mathematics of Operations Research 14 (1989), 700–719.

[6] M.L. Balinski and G. Demange, Algorithms for proportional matrices in reals and integers, Mathematical Programming 45 (1989), 193–210.

[7] M.L. Balinski and H.P. Young, Fair representation—Meeting the ideal of one man, one vote, Second Edition, Brookings Institution Press, Washington, DC, 2001.

[8] M.L. Balinski and S.T. Rachev, Rounding proportions: Methods of rounding, Mathematical Scientist 22 (1997), 1–26.

[9] M.L. Balinski and V. Ramírez, Mexican electoral law: 1996 version, Electoral Studies 16 (1997), 329–349.

[10] S. Berg, Spatial influence on voter transitions in Swedish elections: An application of Johnston's maximum entropy method, Electoral Studies 7 (1988), 233–250.

[11] Y.M.M. Bishop, S.E. Fienberg, and P.W. Holland, Discrete multivariate analysis, Massachusetts Institute of Technology Press, Cambridge, 1975.

[12] L.M. Bregman, Proof of convergence of Sheleikhovskii's method for a problem with transportation constraints, USSR Comput. Math. Math. Phys 7 (1967), 191–204.

[13] P.J. Brown and C.D. Payne, Aggregate data, ecological regression, and voting transitions, Journal of the American Statistical Association 81 (1986), 452–460.

[14] H. Carnal, Mathématiques et politique, Elemente der Mathematik 48 (1993), 27–32.

[15] B.D. Causey, L.H. Cox, and L.R. Ernst, Application of transportation theory to statistical problems, Journal of the American Statistical Association 80 (1985), 903–909.

[16] L.H. Cox and L.R. Ernst, Controlled rounding, INFOR—Information Systems and Operational Research 20 (1982), 423–432.

[17] W.E. Deming and F.F. Stephan, On a least squares adjustment of a sampled frequency table when the expected marginal totals are known, Annals of Mathematical Statistics 11 (1940), 427–444.

[18] B. Fleishmann and J.N. Paraschis, Solving a large scale districting problem: A case report, Computers and Operations Research 15 (1988), 521–533.

[19] A. Forcina and G.M. Marchetti, Modelling transition probabilities in the analysis of aggregate data, Lecture Notes in Statistics 57 (1989), 157–164.

[20] D.R. Fulkerson, An out-of-kilter method for minimal cost flow problems, SIAM Journal on Mathematics 9 (1961), 18–27.

[21] N. Gaffke and F. Pukelsheim, Vector and matrix apportionment problems and separable convex integer optimization, Mathematical Methods of Operations Research 67 (2008), 133–159.

[22] N. Gaffke and F. Pukelsheim, Divisor methods for proportional representation systems: An optimization approach to vector and matrix problems, Mathematical Social Sciences 56 (2008), 166–184.

[23] M.B. Gassner, Two-dimensional rounding for a quasi-proportional representation, European Journal of Political Economy 4 (1988), 529–538.

[24] M.B. Gassner, An impossibility theorem for fair bidimensional representation: Towards a biproportional solution, Mathematical Psychology in Progress, E.E. Roskam (Editor), Berlin, 1989, pp. 345–365.

[25] M.B. Gassner, Représentations parlementaires—Méthodes mathématiques biproportionnelles de Répartition des Sièges, Édition de l'Université de Bruxelles, Bruxelles, 2000.

[26] J.A. George, B.W. Lamar, and C.A. Wallace, Political district determination using large-scale network optimization, Socio-Economic Planning Science 31 (1997), 11–28.

[27] L.A. Goodman, Ecological regression and the behaviour of individuals, American Sociological Review 18 (1953), 351–367.

[28] P. Grilli di Cortona, C. Manzi, A. Pennisi, F. Ricca, and B. Simeone, Evaluation and optimization of electoral systems, SIAM Monographs on Discrete Mathematics and Applications, Philadelphia, SIAM, Society for Industrial and Applied Mathematics, 1999.

[29] T. Helgason, K. Jörnsten, and A. Migdalas, Maxmin formulation of the apportionments of seats to a parliament, Minimax and Applications, D. Du and P.M. Pardalos (Editors), Dordrecht, 1995, pp. 109–118.

[30] W. Hess, J.B. Weaver, H.J. Siegfeldt, J.N. Whelan, and P.A. Zitlau, Nonpartisan political redistricting by computer, Operations Research 13 (1965), 998–1006.

[31] M. Hojati, Optimal political districting, Computers and Operations Research 23 (1996), 1147–1161.

[32] R.J. Johnston and A.M. Hay, On the parameters of uniform swing in single-member constituency electoral systems, Environment and Planning A 14 (1982), 61–74.

[33] R.J. Johnston and C. Pattie, Ecological inference and entropy-maximizing: An alternative estimation procedure for split-ticket voting, Political Analysis 8 (2000), 333–345.

[34] D. Jungnickel, Graphs, networks and algorithms, Springer, Berlin, 1999.

[35] 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 (2008), 371–401.

[36] G. King, O. Rosen, and M.A. Tanner, Beta-binomial hierarchical models for ecological inference, Sociological Methods & Research 28 (2004), 61–90.

[37] S. Maier, Biproportional apportionment methods: Constraints, algorithms, and simulation, Dr. Hut Verlag, München, 2009.

[38] S. Maier and F. Pukelsheim, BAZI: A free computer program for proportional representation apportionment, Preprint Nr. 042/2007. Institut für Mathematik, Universität Augsburg, 2007. Internet: www.opus-bayern.de/uni-augsburg/volltexte/2007/711/, 2007.

[39] S. Maier, P. Zachariassen, and M. Zachariasen, Divisor-based biproportional apportionment in electoral systems: A real-life benchmark study, Management Science 56 (2010), 373–387.

[40] M. Minoux, A polynomial algorithm for minimum quadratic cost flow problems, European Journal of Operational Research 18 (1984), 377-387.

[41] M. Minoux, Solving integer minimum cost flows with separable convex cost objective polynomially, Mathematical Programming Studies 26 (1986), 237-239.

[42] G.J. Minty, Solving steady-state nonlinear networks of “monotone” elements, IRE Transactions on Circuit Theory 8 (1961), 99–104.

[43] A. Pennisi, The Italian bug: A flawed procedure for bi-proportional seat allocation, Mathematics and Democracy. Recent Advances in Voting Systems and Collective Choice, Studies in Choice and Welfare, Springer, Berlin, 2006, pp. 151–165.

[44] A. Pennisi, F. Ricca, P. Serafini, and B. Simeone, Ottimizzazione combinatoria per la progettazione di sistemi elettorali, Scienza delle decisioni in Italia: Applicazioni della Ricerca Operativa ai problemi aziendali, ECIG, Genova, 2008, pp. 75–87.

[45] F. Pukelsheim, BAZI : A Java program for proportional representation, Oberwolfach Reports 1, pp. 735-737, Internet: http://www.math.uni-augsburg.de/stochastik/pukelsheim/2004d.pdf , 2004.

[46] F. Pukelsheim and B. Simeone, On the iterative proportional fitting procedure: Structure of accumulation points and L1-error analysis, Metrika, forthcoming.

[47] F. Ricca, A. Scozzari, and B. Simeone, The Give-up Problem for closed regional lists with multi-winners, Technical Report 16-2008, Dipartimento di Statistica, Probabilità e Statistiche Applicate, Università di Roma "La Sapienza", 2008.

[48] G. Rote and M. Zachariasen, Matrix scaling by network flow, Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, Proceedings in Applied Mathematics, 125, Philadelphia, PA, 2007, pp. 848–854.

[49] U.G. Rothblum and H. Schneider, Scaling of matrices which have prescribed row sums and column sums via optimization, Linear Algebra and Its Applications 114/115 (1989), 737–764.

[50] L. Schrage, LINDO: An optimization modeling system, Fourth Edition, The Scientific Press, San Francisco, CA, 1991.

[51] A. Schrijver, Theory of linear and integer programming, Wiley, New York, 1999.

[52] P. Serafini and B. Simeone, Parametric maximum flow methods for minimax approximation of target quotas in biproportional apportionment, Networks, forthcoming.

[53] P. Serafini and B. Simeone, Certificates of optimality: The third way to biproportional apportionment, Technical Report 17-2010, Dipartimento di Statistica, Probabilità e Statistiche Applicate, Università di Roma "La Sapienza", http://www.dspsa.uniroma1.it/on-line/Home/Ricerca/Pubblicazioni/RapportiTecnicidal2009/articolo1003553.html, 2010.

[54] G.J.G. Upton, A note on the estimation of voter transition probabilities, Journal of the Royal Statistical Society 141 (1978), 507–512.

[55] M. Zachariasen, Algorithmic aspects of divisor-based biproportional rounding, Technical Report 06/05, Department of Computer Science, University of Copenhagen, 2006.