Phragmén's Voting Methods and Justified Representation

Markus Brill
University of Oxford
mbrill@cs.ox.ac.uk

Rupert Freeman
Duke University
rupert@cs.duke.edu

Svante Janson
Uppsala University
svante.janson@math.uu.se

Martin Lackner
University of Oxford
martin.lackner@cs.ox.ac.uk

Abstract

In the late 19th century, Lars Edvard Phragmén proposed a load-balancing approach for selecting committees based on approval ballots. We consider three committee voting rules resulting from this approach: two optimization variants—one minimizing the maximal load and one minimizing the variance of loads—and a sequential variant. We study Phragmén's methods from an axiomatic point of view, focussing on *justified representation* and related properties that have recently been introduced by Aziz et al. (2015a) and Sánchez-Fernández et al. (2017). We show that the sequential variant satisfies *proportional justified representation*, making it the first known polynomial-time computable method with this property. Moreover, we show that the optimization variants satisfy *perfect representation*. We also analyze the computational complexity of Phragmén's methods and provide mixed-integer programming based algorithms for computing them.

1 Introduction

An important part of multiagent systems research concerns the study of preference aggregation mechanisms (e.g., Conitzer 2010). Recent years have witnessed an increasing interest in *committee voting rules* (e.g., Elkind et al. 2014; Skowron, Faliszewski, and Lang 2015; Aziz et al. 2015a; Caragiannis et al. 2016). In this setting, a fixed-size subset of alternatives has to be selected based on the preferences of a group of agents. In this paper, we assume that the preferences of individual agents are given by *approval ballots*, specifying which alternatives are "approved" by the agents.

An important issue in group decision making is *(proportional) representation*. Informally, an outcome of a decision making process is representative if it reflects the preferences of the members of the group. In the context of approval-based committee elections, reasoning about representation is non-trivial. Since approval sets may overlap arbitrarily, there are many different ways in which the set of agents can be split into more or less "cohesive" subgroups. Whether a given subgroup has a justified claim to be represented in the committee depends on the size of the subgroup as well as on its level of cohesiveness.

Aziz et al. (2015a) and Sánchez-Fernández et al. (2016; 2017) have identified axiomatic properties capturing the intuitive notion that subgroups that are "large enough" and "cohesive enough" deserve to be represented in the committee: *justified representation (JR)*, *proportional justified representation (PJR)*, and *extended justified representation (EJR)*. While a number of standard committee voting rules have been shown to satisfy the basic requirement of JR, it turns out that the more demanding properties PJR and EJR are much harder to satisfy. Essentially, the only rule that is known to satisfy PJR and EJR is *Proportional Approval Voting (PAV)*, which was proposed by Danish polymath Thorvald N. Thiele in the late 19th century (Thiele 1895; Janson 2016). Unfortunately, PAV is NP-hard to compute. It has therefore remained an open question whether computationally tractable rules satisfying the more demanding representation properties exist.

In this paper, we consider committee voting rules that are due to Swedish mathematician Lars Edvard Phragmén (1894; 1895; 1896; 1899). Although Phragmén's methods were proposed in the same era as PAV, they have received considerably less attention. Variants of both Phragmén's methods and PAV have been used in Swedish parliamentary elections (for distribution of seats within parties), and a version of one of Phragmén's methods is still part of the election law, although in a minor role (Janson 2016).

Phragmén phrases committee elections as load balancing problems: Adding a candidate to the committee incurs some *load*, and this load should be shared among the agents approving this candidate. Phragmén suggests choosing committees in such a way that the corresponding load distributions are as *balanced* as possible, and different ways of measuring balancedness result in different optimization objectives. This approach yields two *optimization* variants, one minimizing the maximal load and one minimizing the variance of loads, and one *sequential* variant, which proceeds by greedily selecting candidates so as to keep the maximal load as small as possible.

After briefly reviewing related work in Section 2 and introducing some basic notation in Section 3, we formally define Phragmén's rules in Section 4. In Section 5, we analyze the computational complexity of Phragmén's rules and we provide algorithms for computing them. The algorithms for the optimization variants are based on mixed-

integer linear and quadratic programming. In Section 6, we consider the representation axioms mentioned above. We show that the sequential variant satisfies PJR, making it the first known polynomial-time computable method with this property.[1] Moreover, we show that the optimization variants satisfy *perfect representation (PR)*, a further representation axiom introduced by Sánchez-Fernández et al. (2017). The latter result provides a contrast to PAV, which is known to violate PR. Omitted proofs can be found in the full version of this paper.

2 Related Work

Proportional representation is an important issue in committee voting (see the influential paper by Monroe, 1995, and the references therein) and methods ensuring representation often lead to interesting computational problems (Potthof and Brams 1998; Procaccia, Rosenschein, and Zohar 2008; Lu and Boutilier 2011; Betzler, Slinko, and Uhlmann 2013). In the setting of *approval-based* committee voting (Kilgour 2010), Aziz et al. (2015a) proposed two representation axioms: *justified representation (JR)* and its strengthening *extended justified representation (EJR)*. Later, Sánchez-Fernández et al. (2017) observed that EJR is not compatible with what they call *perfect representation (PR)* and proposed an axiomatic property, *proportional justified representation (PJR)*, that is. EJR implies PJR, which in turn implies JR.

Aziz et al. (2015a) and Sánchez-Fernández et al. (2017) showed that most common committee voting rules fail EJR and PJR. A notable exception is Thiele's PAV, which satsfies EJR (and thus PJR). Interestingly, variants of PAV based on different weight vectors fail both EJR and PJR. The same is true for a greedy approximation algorithm for PAV known as *sequential PAV* or *reweighted approval voting*.

Computing the outcome of PAV is NP-hard (Skowron, Faliszewski, and Lang 2015; Aziz et al. 2015b) and thus not feasible in polynomial time unless P = NP. Therefore, it has remained an open question whether there exist polynomial-time computable rules satisfying EJR or PJR. Sánchez-Fernández et al. (2017) have shown that the polynomial-time computable *Greedy Monroe* rule satisfies PJR in the special case where the committee size divides the number of voters (but fails PJR in the general case).

3 Preliminaries

We consider a social choice setting with a finite set $N = \{1, \ldots, n\}$ of *voters* and a finite set C of *candidates*. Throughout the paper we let $m = |C|$ denote the number of candidates and $n = |N|$ the number of voters. The preferences of each voter $i \in N$ are given by a subset $A_i \subseteq C$, representing the subset of candidates that the voter approves of. We refer to the list $A = (A_1, \ldots, A_n)$ as the *preference profile*. For a candidate $c \in C$, we let N_c denote the set of voters approving c, i.e., $N_c = \{i \in N : c \in A_i\}$. To avoid trivialities, we assume that $N_c \neq \emptyset$ for all $c \in C$.

We want to select a subset consisting of exactly k candidates, for a given natural number $k \leq m$. An *approval-based multi-winner voting rule* (henceforth simply *rule*) maps an instance (A, k) to a subset $S \subseteq C$ of size k, the *committee*. In general, there may be ties, and we then allow the rule to yield several choices, so formally the rule is a map from instances to non-empty sets of committees.

Finally, for a tuple of real numbers $z = (z_1, \ldots, z_n)$, we let $z_{(\ell)}$ denote the ℓ-th largest element in z.

4 Phragmén's Methods

The main idea behind Phragmén's methods is to identify committees whose "support" is distributed as evenly as possible among the electorate. Phragmén used different formulations for explaining his methods; we refer the reader to the survey by Janson (2016) for an overview and more details. In this paper, we adopt the formulation from the 1899 paper (Phragmén 1899). In this formulation, every candidate in the committee is thought of as incurring one unit of "*load*," and the load incurred by candidate c needs to be distributed among the voters in N_c. The goal is to find a committee of size k for which the corresponding load distribution is as balanced as possible.

Formally, a *load distribution* is a two-dimensional array $x = (x_{i,c})_{i \in N, c \in C}$ satisfying the following four conditions:

$$0 \leq x_{i,c} \leq 1 \qquad \text{for all } i \in N \text{ and } c \in C \quad (1)$$

$$x_{i,c} = 0 \qquad \text{if } c \notin A_i \quad (2)$$

$$\sum_{i \in N} \sum_{c \in C} x_{i,c} = k \qquad (3)$$

$$\sum_{i \in N} x_{i,c} \in \{0, 1\} \qquad \text{for all } c \in C \quad (4)$$

Here, $x_{i,c}$ corresponds to the load that voter i receives from candidate c. Condition (2) ensures that the load incurred by candidate c is distributed among voters in N_c only, and Conditions (3) and (4) ensure that x corresponds to a size-k committee $\{c \in C : \sum_{i \in N} x_{i,c} = 1\}$.

For a load distribution x, we let $\bar{x}_i$ denote the total load of voter $i \in N$, i.e., $\bar{x}_i = \sum_{c \in C} x_{i,c}$, and we refer to $(\bar{x}_1, \ldots, \bar{x}_n)$ as the vector of *voter loads*. Using this notation, Condition (3) reads $\sum_{i \in N} \bar{x}_i = k$. Note that Condition (3) implies that the *average* voter load is $\frac{k}{n}$.

There are different ways to measure how *balanced* a given load distribution is, each giving rise to a different optimization objective. One such objective is to minimize the maximal load assigned to a voter, i.e., $\min_x \max_{i \in N} \bar{x}_i$. (This is equivalent to minimizing the maximal difference between a voter load and the average voter load.) Obviously, the average voter load $\frac{k}{n}$ is a lower bound on the maximal voter load, and we call a load distribution x *perfect* if $\bar{x}_i = \frac{k}{n}$ for all $i \in N$. Another objective is to minimize the *variance* of voter loads, i.e., the sum of squared distances from the average voter load. Again, a perfect load distribution is optimal for this objective.

We further distinguish between "optimization" methods, where we solve a global optimization problem to find a load

distribution optimizing the objective, and "sequential" methods, where we iteratively construct a load distribution, in each round greedily choosing a candidate optimizing the objective at that iteration.

In this paper, we consider three rules: the optimization methods *max-Phragmén* and *var-Phragmén*—minimizing the maximal voter load and the variance of voter loads, respectively—and the sequential method *seq-Phragmén*, which greedily minimizes the maximal voter load. The method seq-Phragmén was introduced by Phragmén (1894; 1895; 1896; 1899), and it is the variant that he proposed to be used in actual elections. Phragmén defined this method as a generalization of D'Hondt's apportionment method to the case without party lists: for every instance of the party-list setting, seq-Phragmén and D'Hondt's method coincide (Phragmén 1895; Janson 2016; Brill, Laslier, and Skowron 2017). Optimization variants and the objective of minimizing the variance are discussed in the 1896 paper (Phragmén 1896). Despite their intuitive appeal, Phragmén's methods have received little attention in the social choice literature.[2]

4.1 Optimization Variants

We first define the optimization variants.

max-Phragmén: The rule *max-Phragmén* selects committees minimizing the maximal voter load. In case that two or more committees have the same (minimal) maximal load, we employ a specific tie-breaking. This is because it might be the case that for two load distributions x and y, although $\max_{i \in N} \bar{x}_i = \max_{i \in N} \bar{y}_i$, one load distribution is clearly preferable to the other.

Example 1. *Consider $A = (\{a\}, \{a\}, \{b\}, \{c\})$ and $k = 2$. Any committee of size 2 contains either b or c, which are approved by only one voter each, so the maximum load is 1 for all committees. Thus, all subsets of size 2 minimize the maximal voter load, although arguably the committees containing a are preferable to the committee $\{b, c\}$.*

Thus, to refine the set of winning committees, we compare two committees in the following way.

Definition 1. *For $y = (y_1, \ldots, y_n)$ and $z = (z_1, \ldots, z_n)$, y is* leximax-smaller *than z, denoted $y \overset{.}{<} z$, if there exists $j \leq n$ such that $y_{(j)} < z_{(j)}$ and $y_{(i)} = z_{(i)}$ for all $i \leq j - 1$.*

max-Phragmén selects all committees corresponding to load distributions $(\bar{x}_1, \ldots, \bar{x}_n)$ that are leximax-optimal, i.e., minimal with respect to $\overset{.}{<}$. As we will see in Section 6.3, this tie-breaking is necessary in order to guarantee strong representation properties.

Rather than minimizing the maximum load, one could also aim to (lexicographically) *maximize the minimal voter load*. This variant of Phragmén's method would select committees minimizing the number of unrepresented voters, even in the face of large cohesive groups of voters. Therefore, this method will not do well in terms of the representation axioms considered in Section 6. For this reason, we do not consider it further in this paper.

var-Phragmén: The rule var-Phragmén selects committees corresponding to load distributions minimizing $\sum_{i \in N} \bar{x}_i^2$. Minimizing the sum of squares of $(\bar{x}_1, \ldots, \bar{x}_n)$ indeed minimizes the variance of $(\bar{x}_1, \ldots, \bar{x}_n)$, as is well-known.

The following example demonstrates that the maximal voter load under var-Phragmén may indeed be greater than under max-Phragmén.

Example 2. *Let $C = \{a, b, c, d\}$, $k = 3$, and consider the profile $A = (\{a\}, \{b\}, \{b, c\}, \{a, b, c\}, \{d\})$. For this instance, max-Phragmén selects the committee $\{a, b, c\}$ and var-Phragmén selects the committee $\{a, b, d\}$. Optimal load distributions corresponding to these committees are illustrated in Figure 1. Load distributions minimizing the maximal voter load (like the one illustrated by the first diagram in Figure 1) satisfy $\max_{i \in N} \bar{x}_i = \frac{3}{4}$ and $\sum_{i \in N} \bar{x}_i^2 = 4(\frac{3}{4})^2 = \frac{9}{4}$, and the load distribution minimizing the variance of voter loads (illustrated by the second diagram in Figure 1) satisfies $\max_{i \in N} \bar{x}_i = 1$ and $\sum_{i \in N} \bar{x}_i^2 = 4(\frac{1}{2})^2 + 1^2 = 2$.*

4.2 Sequential Method

We now introduce the sequential method.

seq-Phragmén: The rule seq-Phragmén starts with an empty committee and iteratively adds candidates, always choosing the candidate that minimizes the (new) maximal voter load. Let $\bar{x}_i^{(j)}$ denote the voter loads after round j. At first, all voters have a load of 0, i.e., $\bar{x}_i^{(0)} = 0$ for all $i \in N$. As a first candidate we select one that is supported by most voters as it is the one that increases the maximal load the least. In the next round, we again choose a candidate that induces a (new) maximal voter load that is as small as possible, but now we have to take into account that some voters already have a non-zero load. The new maximal load if c is chosen as the $(j + 1)$-st committee member is calculated as

$$s_c^{(j+1)} = \frac{1 + \sum_{i \in N_c} \bar{x}_i^{(j)}}{|N_c|}. \quad (5)$$

This is because we distribute the load of 1 among all voters in N_c in such a way that all these voters have the same voter load afterwards. Let c be the candidate that minimizes $s_c^{(j+1)}$ among those that are not yet in the committee.[3] Then we add c to the committee and set

$$\bar{x}_i^{(j+1)} = \begin{cases} s_c^{(j+1)} & \text{if } i \in N_c \\ \bar{x}_i^{(j)} & \text{otherwise.} \end{cases} \quad (6)$$

It follows that $\sum_{i \in N} \bar{x}_i^{(j+1)} = j + 1$. After k iterations, we have obtained a load distribution and a committee.

The definitions ensure that voter loads never decrease, i.e., $\bar{x}_i^{(j+1)} \geq \bar{x}_i^{(j)}$ for all $i \in N$ and all $j < k$. This is because a candidate minimizing the new maximal load is selected in each round. If the selection of candidate c in round $j + 1$ led to a load distribution $x^{(j+1)}$ with $s_c^{(j+1)} = \bar{x}_i^{(j+1)} < \bar{x}_i^{(j)}$ for some $i \in N_c$, then candidate c would have been selected in an earlier round, a contradiction. (See also Lemma 5*(i)*.)

---

[2]Notable exceptions are a survey by Janson (2012) (in Swedish) and a paper by Mora and Oliver (2015) (in Catalan).

[3]If there are several candidates minimizing $s_c^{(j+1)}$, we use a fixed tie-breaking rule to decide which candidate to add.
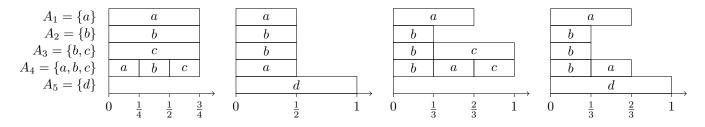
Figure 1: Illustration of Examples 2 and 3. From left to right: The first diagram illustrates a load distribution minimizing the maximal voter load $\max_{i \in N} \bar{x}_i$; the second diagram illustrates the unique load distribution minimizing $\sum_{i \in N} \bar{x}_i^2$; the third and fourth diagrams illustrate the load distributions obtained by seq-Phragmén with ties broken in favor of c or d, respectively.

Phragmén (1899) illustrates his sequential method by imagining the different groups of ballots as represented by cylindrical vessels, with base area proportional to the number of ballots in each group. The already elected candidates are represented by a liquid that is fixed in the vessels, and the additional unit of load incurred by adding another candidate to the committee is represented by pouring 1 unit of a liquid into the vessels representing voters approving this candidate. The liquid then distributes among these vessels such that the height of the liquid is the same in all vessels. This is to be tried for each candidate; the candidate that requires the smallest height is elected, and the corresponding amounts of liquid are added to the vessels and fixed there.

The sequential method seq-Phragmén can be seen as a (polynomial-time computable) heuristic to approximate the optimization method max-Phragmén. Unsurprisingly, the load distribution constructed by seq-Phragmén might not be optimally balanced.

Example 3. *Consider again the instance from Example 2. We have $s_b^{(1)} = \frac{1}{3}$, $s_a^{(1)} = s_c^{(1)} = \frac{1}{2}$, and $s_d^{(1)} = 1$. Therefore, candidate b is chosen in the first round. In the second round, we have $s_a^{(2)} = \frac{2}{3}$, $s_c^{(2)} = \frac{5}{6}$ and $s_d^{(2)} = 1$, so candidate a is chosen. In the third round, there is a tie between c and d because $s_c^{(3)} = s_d^{(3)} = 1$. Thus, the final committee is either $\{a, b, c\}$ or $\{a, b, d\}$, depending on which tie-breaking method is used. Figure 1 illustrates the resulting load distributions, both of which are suboptimal for the optimization problems corresponding to max-Phragmén and var-Phragmén.*

One can also define a sequential version of var-Phragmén, by in each iteration selecting a candidate minimizing the variance of the resulting load distribution (Mora 2016). This variant does not fare well in terms of the representation axioms considered in Section 6, and we therefore do not consider it any further.

5 Computational Aspects

In this section, we study the computational complexity of Phragmén's methods, and we provide algorithms for finding winning committees. Sánchez-Fernández et al. (2017) have shown that every rule satisfying *perfect representation* (see Section 6) is NP-hard; this essentially follows from earlier work by Procaccia, Rosenschein, and Zohar (2008). Since we show that max-Phragmén and var-Phragmén both satisfy

this condition (Theorems 8 and 11), it follows that there do not exist polynomial-time algorithms for computing a committee for either of these rules, unless P = NP.

We complement these hardness results by considering two basic decision problems. MAX-PHRAGMÉN asks whether an instance allows a load distribution x such that $(\bar{x}_1, \ldots, \bar{x}_n) \mathrel{\dot{<}} (y_1, \ldots, y_n)$ for some given n-tuple $(y_1, \ldots, y_n)$. And VAR-PHRAGMÉN asks whether an instance allows a load distribution x such that $\sum_{i \in N} \bar{x}_i^2 < \alpha$ for some given threshold value $\alpha > 0$. Both problems can be interpreted as asking whether a given load distribution is optimal. We show that both problems are NP-complete even for rather restricted instances. For a preference profile A, let $s(A)$ denote the maximum number of candidates a voter approves, and let $d(A)$ denote the maximum number of voters that approve a candidate.

Theorem 1. MAX-PHRAGMÉN, *and* VAR-PHRAGMÉN *are NP-complete, even restricted to instances with $s(A) = 2$ and $d(A) = 3$.*

We now turn to algorithms for computing Phragmén's rules. First, we show how the outcome of max-Phragmén can be computed with the help of mixed-integer linear programs (MILPs). We start by formulating an MILP that solves the decision problem MAX-PHRAGMÉN. We use the variables $x_{i,c}$ (for $i \in N$, $c \in C$), $e_{i,j}$ (for $i, j \in N$), s_i (for $i \in N$), t_j (for $j \in N$), and ϵ. For a given n-tuple $\mathbf{y} = (y_1, \ldots, y_n)$ of real numbers, let $\mathsf{P}(\mathbf{y})$ be the MILP that maximizes ϵ under the constraints (1)–(4) and (7)–(14).

$$e_{i,j} \in \{0, 1\} \qquad \text{for all } i, j \in N \quad (7)$$

$$s_i \in \{0, 1\} \qquad \text{for all } i \in N \quad (8)$$

$$t_j \in \{0, 1\} \qquad \text{for all } j \in N \quad (9)$$

$$s_i + \sum_{j \in N} e_{i,j} = 1 \qquad \text{for all } i \in N \quad (10)$$

$$t_j + \sum_{i \in N} e_{i,j} \leq 1 \qquad \text{for all } j \in N \quad (11)$$

$$\sum_{j \in N} t_j = 1 \qquad \text{for all } i \in N \quad (12)$$

$$\bar{x}_i - k(1 - e_{i,j}) \leq y_j \qquad \text{for all } i, j \in N \quad (13)$$

$$\bar{x}_i - k(2 - s_i - t_j) \leq y_j - \epsilon \qquad \text{for all } i, j \in N \quad (14)$$

The main idea of this MILP is as follows: The variables $e_{i,j}$ encode a partial bijection π from a subset of N to a subset of N; the variables s_i encode the subset $S \subseteq N$ where π is not defined; and the variables t_j encode $t \in N$, an index of an element in $\{y_j : j \notin range(\pi)\}$. Constraint (10) encodes the relation between π and S: for every $i \in N$, either $s_i = 1$ or $e_{i,j} = 1$ for some $j \in N$. In a similar fashion, constraint (11) encodes the relation between π and t: for every $i \in N$, $t_i = 1$ only if $e_{i,j} = 0$ for all $j \in N$. Together with constraint (12), we enforce that there exists exactly one $j \in N$ such that $t_j = 1$. Hence at least one voter has a load strictly smaller than y_t and $(\bar{x}_1, \ldots, \bar{x}_n) \mathrel{\dot{<}} (y_1, \ldots, y_n)$.

The final two constraints ensure that indeed $(\bar{x}_1, \ldots, \bar{x}_n) \mathrel{\dot{<}} (y_1, \ldots, y_n)$. From constraint (13) it follows that $\bar{x}_i \leq y_j$ whenever $\pi(i) = j$. This is because if $e_{i,j} = 0$ (i.e., $\pi(i) \neq j$), constraint (13) reduces to $\bar{x}_i - k \leq y_j$, which is trivially satisfied because every load distribution x satisfies $\bar{x}_i \leq k$ for all $i \in N$. If $e_{i,j} = 1$ (i.e., $\pi(i) = j$), however, constraint (13) reads $\bar{x}_i \leq y_j$. Similarly, constraint (14) enforces that $x_i \leq y_t - \epsilon \leq \max_{j \in N \setminus range(\pi)} y_j - \epsilon$ for $i \in S$. As we maximize ϵ, we look for a solution where $x_i < \max_{j \in N \setminus range(\pi)} y_j$. We conclude that a feasible solution with objective function value $\epsilon > 0$ encodes a load distribution x with $(\bar{x}_1, \ldots, \bar{x}_n) \mathrel{\dot{<}} (y_1, \ldots, y_n)$. Observe that $\mathsf{P}(\mathbf{y})$ solves the MAX-PHRAGMÉN decision problem: given voter loads $\mathbf{y}$, $\mathsf{P}(\mathbf{y})$ returns $\epsilon > 0$ if and only if MAX-PHRAGMÉN with input $\mathbf{y}$ is a Yes-instance.

We now present an MILP-based algorithm that computes the outcome of max-Phragmén. Our algorithm solves a sequence of at most $2n$ instantiations of the MILP P, using the optimal solutions of previously solved instances as constraints for subsequent calls. We assume that P returns the load distribution x and the objective function value ϵ. For an overview of the procedure, see Algorithm 1.

Algorithm 1: Computing max-Phragmén

$\mathbf{y} \leftarrow (k, 0, \ldots, 0)$
for $\ell = 1 \ldots n$ **do**
    $x, \epsilon \leftarrow \mathsf{P}(\mathbf{y})$
    $\bar{x} \leftarrow (\bar{x}_1, \ldots, \bar{x}_n)$ `// ` $\bar{x}_{(1)}, \ldots, \bar{x}_{(\ell)}$ `optimal`
    if $\epsilon = 0$ **then**   `// no improvement`
        $x', \epsilon' \leftarrow \mathsf{P}(\bar{x})$
        if $\epsilon' = 0$ **then**   `// ` $\bar{x}$ `optimal`
            return $\{c \in C : \sum_{i \in N} x_{i,c} = 1\}$
    $\mathbf{y} \leftarrow (\bar{x}_{(1)}, \ldots, \bar{x}_{(\ell+1)}, 0, \ldots, 0)$
return $\{c \in C : \sum_{i \in N} x_{i,c} = 1\}$

We start with $\mathbf{y} = (k, 0, \ldots, 0)$, an n-tuple consisting of one k and $n - 1$ zeroes. We employ P to find a strictly better solution. The only entry of $\mathbf{y}$ that can be improved is $\mathbf{y}_{(1)} = k$ and hence the solution x returned by P minimizes the largest load; let $\bar{x}_{(1)}$ be the largest load and $\bar{x}_{(2)}$ the second-largest. We repeat this procedure with

$\mathbf{y} = (\bar{x}_{(1)}, \bar{x}_{(2)}, 0, \ldots, 0)$. We already know that $\bar{x}_{(1)}$ is optimal and cannot be further decreased (and 0 cannot be improved), hence the next P instance minimizes the second-largest load. We iterate this process and in step ℓ guarantee that the ℓ-th largest load is optimal. If at some point P returns $\epsilon = 0$, we verify whether the current solution is optimal: if $\mathsf{P}(\bar{x})$ also returns $\epsilon = 0$, the load distribution x is indeed optimal and the algorithm terminates. In any case Algorithm 1 returns $\{c \in C : \sum_{i \in N} x_{i,c} = 1\}$, the committee corresponding to the load distribution x.

We have therefore proven the following result.

Theorem 2. *max-Phragmén can be computed by solving at most $2n$ mixed-integer linear programs with $\mathcal{O}(nm + n^2)$ variables.*

To compute var-Phragmén, we solve a mixed-integer quadratic program, i.e., a program consisting of linear constraints and a quadratic optimization statement.

Theorem 3. *var-Phragmén can be computed by solving one mixed-integer quadratic program with $\mathcal{O}(nm)$ variables.*

Finally, we study the runtime for computing seq-Phragmén. A naive estimate is that seq-Phragmén can be computed in $\mathcal{O}(kmn)$ time. This estimate ignores the cost of computing the quantities $s_c^{(j)}$, i.e., numerical operations are assumed to require constant time. While this is a sensible assumption in many cases, here it is questionable since computing $s_c^{(j)}$ exactly requires fractions with large numerators and denominators. Indeed, the denominator of $s_c^{(j)}$ can grow exponentially with j. Hence, the following theorem also takes the complexity of these operations into account.

Theorem 4. *The output of seq-Phragmén can be computed in $\mathcal{O}(k^3 mn(\log n)^2)$ time.*

6 Phragmén's Methods and Representation

In this section, we study which representation axioms are satisfied by Phragmén's methods. Our results are summarized in Table 1. Particularly noteworthy are the results that seq-Phragmén satisfies PJR and that max-Phragmén and var-Phragmén satisfy PR.

6.1 Justified Representation Axioms

We start by restating the definitions of Aziz et al. (2015a) and Sánchez-Fernández et al. (2017).

Definition 2. *A committee $S \subseteq C$ with $|S| = k$ provides*

- *justified representation (JR) if there does not exist a set $N^* \subseteq N$ of voters with $|N^*| \geq \frac{n}{k}$, $|\bigcap_{i \in N^*} A_i| \geq 1$ and $|S \cap A_i| = 0$ for all $i \in N^*$.*

- *proportional justified representation (PJR) if there does not exist an integer $\ell > 0$ and a set $N^* \subseteq N$ of voters with $|N^*| \geq \ell \frac{n}{k}$, $|\bigcap_{i \in N^*} A_i| \geq \ell$ and $|S \cap (\bigcup_{i \in N^*} A_i)| < \ell$.*

- *extended justified representation (EJR) if there does not exist an integer $\ell > 0$ and a set $N^* \subseteq N$ of voters with $|N^*| \geq \ell \frac{n}{k}$, $|\bigcap_{i \in N^*} A_i| \geq \ell$ and $|S \cap A_i| < \ell$ for all $i \in N^*$.*

	JR	PJR	EJR	PR
seq-Phr.	✓(Cor. 7)	✓(Th. 6)	– (Ex. 5)	– (Ex. 4)
max-Phr.	✓(Cor. 10)	✓(Th. 9)	– (Ex. 4)	✓(Th. 8)
var-Phr.	✓(Th. 12)	– (Ex. 6)	– (Ex. 4)	✓(Th. 11)

Table 1: Phragmén's methods and representation axioms

A rule f satisfies *JR (respectively, PJR or EJR) if, for every instance (A, k), every committee $S \in f(A, k)$ provides JR (respectively, PJR or EJR).*

It follows immediately from the definitions that a rule satisfying EJR also satisfies PJR, and that a rule satisfying PJR also satisfies JR.

The following definition is due to Sánchez-Fernández et al. (2017).

Definition 3. *Consider an instance (A, k) so that k divides $n = |N|$. A committee $S = \{c_1, \ldots, c_k\} \subseteq C$ provides* perfect representation *if there exists a partition of the set N of voters into k pairwise disjoint subsets $N_1, \ldots, N_k$ such that, for all $j \in \{1, \ldots, k\}$, $|N_j| = \frac{n}{k}$ and $c_j \in \bigcap_{i \in N_j} A_i$. Let $PR(A, k)$ denote the set of all committees providing perfect representation for the instance (A, k). A rule f satisfies perfect representation (PR) if, for every instance (A, k) where k divides n and $PR(A, k) \neq \emptyset$, we have $f(A, k) \subseteq PR(A, k)$.*

The following example, which also appears in the papers by Aziz et al. (2015a) and Sánchez-Fernández et al. (2017), illustrates the requirements of the different axioms.

Example 4. *Let $C = \{a, b, c, d, e, f\}$ and consider the 8-voter preference profile given by $A_1 = \{a\}$, $A_2 = \{b\}$, $A_3 = \{c\}$, $A_4 = \{d\}$, $A_5 = \{a, e, f\}$, $A_6 = \{b, e, f\}$, $A_7 = \{c, e, f\}$, $A_8 = \{d, e, f\}$. Let $k = 4$ and assume that ties are broken alphabetically. Then, seq-Phragmén chooses e, f, a, and b (in this order). The final loads are $(\bar{x}_1, \ldots, \bar{x}_8) = (\frac{3}{4}, \frac{3}{4}, 0, 0, \frac{3}{4}, \frac{3}{4}, \frac{1}{2}, \frac{1}{2})$. This is indeed not optimal as there is a perfect load distribution y with $\bar{y}_i = \frac{1}{2}$ for all $i \in N$. The corresponding committee $\{a, b, c, d\}$ is selected by both max-Phragmén and var-Phragmén.*

Consider the group of voters $N^* = \{5, 6, 7, 8\}$, of size $\ell \frac{n}{k} = 2\frac{8}{4} = 4$, where $\ell = 2$. Since the voters all approve candidates e and f, a set of size $\ell = 2$, the conditions for JR, PJR, and EJR all bind. JR requires that at least one candidate approved by at least one voter in N^* is chosen. PJR requires that at least 2 candidates are chosen that are each supported by at least one voter from N^*, while EJR requires that some voter from N^* is represented twice. Thus, EJR dictates that either e or f is chosen. On the other hand, the only committee providing PR is $\{a, b, c, d\}$. As a consequence, no rule can satisfy both PR and EJR. Note that max-Phragmén and var-Phragmén both violate EJR in this example, and that seq-Phragmén violates PR.

The incompatibility of PR and EJR was first observed by Sánchez-Fernández et al. (2017).

6.2 Results for seq-Phragmén

In this section we establish our main result: seq-Phragmén satisfies proportional justified representation.

We need the following notation. For the committee S that is selected by seq-Phragmén (using a fixed tie-breaking rule), we can relabel the candidates such that $S = \{c_1 \ldots, c_k\}$ and candidate c_j was chosen in round j. Then, we have $c_j = \arg\min_{c \in C \setminus \{c_1, \ldots, c_{j-1}\}} s_c^{(j)}$. Using this convention, we define $s^{(j)} = s_{c_j}^{(j)}$. That is, $s^{(j)}$ is the new load of all voters in N_{c_j} after candidate c_j is added to the committee in round j. We call $(s^{(1)}, \ldots, s^{(k)})$ the *max-load sequence*. (Note that different tie-breaking rules can lead to different max-load sequences.)

The following lemma has two parts. The first part states that the max-load sequence is monotonically increasing. The second part states that, when computing the optimal distribution of the load of a candidate c among its voters, it never helps to restrict attention to a subset $N' \subset N_c$.

Lemma 5. *Fix an instance (A, k).*

(i) The max-load sequence satisfies $s^{(1)} \leq \ldots \leq s^{(k)}$.

(ii) For a candidate $c \in C$, a subset $N' \subseteq N_c$, and $j \leq k$, let $s_c^{(j)}[N']$ denote the maximal voter load after optimally distributing the load of c among all voters in N'. Then, $s_c^{(j)}[N_c] \leq s_c^{(j)}[N']$ for all $N' \subseteq N_c$.

We are now ready to prove our main theorem.

Theorem 6. *seq-Phragmén satisfies PJR.*

Proof. PJR requires that $|S \cap (\bigcup_{i \in N^*} A_i)| \geq \ell$ for all groups $N^* \subseteq N$ of voters satisfying $|N^*| \geq \ell \frac{n}{k}$ and $|\bigcap_{i \in N^*} A_i| \geq \ell$ for some integer $\ell > 0$. We show that seq-Phragmén satisfies a strictly stronger property by weakening the constraint $|N^*| \geq \ell \frac{n}{k}$ to $|N^*| > \ell \frac{n}{k+1}$.

Consider an instance (A, k) and let S be the committee selected by seq-Phragmén. Assume for contradiction that there exists a group $N^* \subseteq N$ of voters and an integer $\ell > 0$ with $|N^*| > \ell \frac{n}{k+1}$ such that $|\bigcap_{i \in N^*} A_i| \geq \ell$ and $|S \cap (\bigcup_{i \in N^*} A_i)| \leq \ell - 1$.

Let $c \in (\bigcap_{i \in N^*} A_i) \setminus S$ and consider round k (the last round) of the seq-Phragmén procedure. Adding candidate c to the committee would have caused a maximal voter load of

$$s_c^{(k)} = \frac{1 + \sum_{i \in N_c} \bar{x}_i^{(k-1)}}{|N_c|} \leq \frac{1 + \sum_{i \in N^*} \bar{x}_i^{(k-1)}}{|N^*|}$$
$$\leq \frac{1 + (\ell - 1)}{|N^*|} = \frac{\ell}{|N^*|} < \frac{k+1}{n}.$$

Here, the first inequality follows from part (ii) of Lemma 5 (observe that $N^* \subseteq N_c$), the second inequality follows from $|S \cap (\bigcup_{i \in N^*} A_i)| \leq \ell - 1$, and the strict inequality follows from $|N^*| > \ell \frac{n}{k+1}$.

Let c_k be the candidate that was chosen in round k. Since candidate c was *not* chosen, we have $c \neq c_k$ and $s_{c_k}^{(k)} \leq s_c^{(k)}$. Using part (i) of Lemma 5, we have $s^{(1)} \leq \ldots \leq s^{(k)} = s_{c_k}^{(k)} \leq s_c^{(k)} < \frac{k+1}{n}$. In particular, this implies that at the end of round k, every voter $i \in N$ has a load $\bar{x}_i^{(k)}$ that is strictly

less than $\frac{k+1}{n}$. Summing the loads over all voters, we get

$$\sum_{i \in N} \bar{x}_i^{(k)} = \sum_{i \in N^*} \bar{x}_i^{(k)} + \sum_{i \in N \setminus N^*} \bar{x}_i^{(k)}$$
$$\leq (\ell - 1) + |N \setminus N^*| \cdot s^{(k)}$$
$$< \ell - 1 + \frac{n}{k+1}(k+1-\ell)\frac{k+1}{n} = k,$$

where we have used the fact that $|N \setminus N^*| \leq \frac{n}{k+1}(k+1-\ell)$. But $\sum_{i \in N} \bar{x}_i^{(k)} < k$ is a contradiction, because the sum of all voter loads (at the end of the seq-Phragmén procedure) must equal k. This completes the proof. □

A consequence of Theorem 6 is that a committee providing PJR can be computed in polynomial time. We note that the proof of Theorem 6 shows that seq-Phragmén satisfies a property that is strictly stronger than PJR, because the constraint on the size of group N^* has been relaxed.

As an immediate corollary of Theorem 6, we obtain that seq-Phragmén satisfies JR.

Corollary 7. *seq-Phragmén satisfies JR.*

However, seq-Phragmén violates EJR, as the following example demonstrates.

Example 5. *Consider the following instance with $n = 24$, $k = 12$, and $C = \{a, b, c_1, c_2, \ldots, c_{12}\}$.*

$2 \times \{a, b, c_1\}$	$6 \times \{c_1, c_2, \ldots, c_{12}\}$
$2 \times \{a, b, c_2\}$	$5 \times \{c_2, c_3, \ldots, c_{12}\}$
	$9 \times \{c_3, c_4, \ldots, c_{12}\}$

seq-Phragmén selects $S = \{c_1, c_2, \ldots, c_{12}\}$. To see that S does not provide EJR, consider the group N^* consisting of the four voters on the left. We have $|N^*| = 4 = 2\frac{n}{k}$ and $|\bigcap_{i \in N^*} A_i| = |\{a, b\}| = 2$. Therefore, EJR requires that at least one voter in N^* approves at least 2 candidates in S, which is not the case.

Therefore, it remains an open problem whether an committee providing EJR can be computed in polynomial time.

Note that seq-Phragmén also fails PR (see Example 4). This is not surprising, considering that PR is computationally intractable (Sánchez-Fernández et al. 2017).

6.3 Results for max-Phragmén

In Example 4, max-Phragmén selects the committee providing perfect representation. We now show that max-Phragmén satisfies PR in general.

Theorem 8. *max-Phragmén satisfies PR.*

The proof consists of (1) observing that the existence of a committee providing PR implies the existence of a perfect load distribution, and (2) showing that *every* perfect load distribution corresponds to a committee providing PR. For the latter, we invoke the Birkhoff–von-Neumann theorem.

Since EJR is incompatible with PR (see Example 4), max-Phragmén fails EJR. However, it satisfies property PJR.

Theorem 9. *max-Phragmén satisfies PJR.*

The proof is by contradiction. Assuming that there is a cohesive group N^* such that not enough candidates approved by voters in N^* are in the committee, the corresponding load distribution can be improved upon (with respect to leximax comparisons) by shifting load from $N \setminus N^*$ to N^*.

Corollary 10. *max-Phragmén satisfies JR.*

We note that Example 1 shows that simply minimizing the maximal voter load (without leximax tie-breaking; see Definition 1) does not even yield committees satisfying JR.

6.4 Results for var-Phragmén

The proof of Theorem 8 directly applies to var-Phragmén.

Theorem 11. *var-Phragmén satisfies PR.*

Unlike max-Phragmén, var-Phragmén fails PJR.

Example 6. *Consider the following example with 100 voters, $C = \{a, b, c, d, e, f, g\}$, and $k = 6$. 67 voters approve $\{a, b, c, d\}$, 12 voters approve $\{e\}$, 11 voters approve $\{f\}$, and 10 voters approve $\{g\}$. Let N^* be the set of voters approving $\{a, b, c, d\}$. We have $|N^*| = 67 \geq 4\frac{n}{k}$ and $|\bigcap_{i \in N^*} A_i| = 4$. Thus, PJR requires that all four candidates in $\bigcap_{i \in N^*} A_i = \{a, b, c, d\}$ are selected. However, var-Phragmén selects $\{a, b, c, e, f, g\}$.*

Example 6 also shows that the sequential version of var-Phragmén violates PJR. Example 6 is an instance of the *party-list* setting (with four disjoint parties). An alternative proof that var-Phragmén violates PJR consists in noting that in the party-list setting, var-Phragmén reduces to Sainte-Laguë's apportionment method (Sainte-Laguë 1910), and PJR is equivalent to lower quota (Brill, Laslier, and Skowron 2017; Sánchez-Fernández, Fernández, and Fisteus 2016). It is well known that Sainte-Laguë's method violates lower quota (Balinski and Young 1982).

Finally, we prove that var-Phragmén satisfies JR.

Theorem 12. *var-Phragmén satisfies JR.*

7 Conclusion

We have shown that Phragmén's load-balancing methods satisfy interesting representation axioms. In particular, the polynomial-time computable variant seq-Phragmén satisfies PJR. Moreover, we have shown that both max-Phragmén and var-Phragmén satisfy PR and that max-Phragmén additionally satisfies PJR. Arguably, max-Phragmén is the first known example of a "natural" rule satisfying both PR and PJR—the only other rule known to satisfy these two properties is an artificial construct that returns a PR committee if one exists and otherwise runs PAV (Sánchez-Fernández et al. 2017). The Monroe rule (i.e., the optimization variant of *Greedy Monroe*) satisfies PR by definition, but fails PJR if the committee size does not divide the number of voters (Sánchez-Fernández et al. 2017).

Since seq-Phragmén violates EJR, it remains an open problem whether committees providing EJR can be computed efficiently. The intricate nature of Example 5 seems to suggest that instances on which seq-Phragmén violates EJR are rare. It would be interesting to see whether seq-Phragmén satisfies EJR for realistic distributions of preferences and/or for reasonable domain restrictions.

Acknowledgments

References

Aziz, H.; Brill, M.; Conitzer, V.; Elkind, E.; Freeman, R.; and Walsh, T. 2015a. Justified representation in approval-based committee voting. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence (AAAI)*, 784–790. AAAI Press.

Aziz, H.; Gaspers, S.; Gudmundsson, J.; Mackenzie, S.; Mattei, N.; and Walsh, T. 2015b. Computational aspects of multi-winner approval voting. In *Proceedings of the 14th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, 107–115. IFAAMAS.

Balinski, M., and Young, H. P. 1982. *Fair Representation: Meeting the Ideal of One Man, One Vote*. Yale University Press.

Betzler, N.; Slinko, A.; and Uhlmann, J. 2013. On the computation of fully proportional representation. *Journal of Artificial Intelligence Research* 47:475–519.

Brill, M.; Laslier, J.-F.; and Skowron, P. 2017. Multiwinner approval rules as apportionment methods. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI)*. AAAI Press. Forthcoming.

Caragiannis, I.; Nath, S.; Procaccia, A. D.; and Shah, N. 2016. Subset selection via implicit utilitarian voting. In *Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI)*, 151–157. AAAI Press.

Conitzer, V. 2010. Making decisions based on the preferences of multiple agents. *Communications of the ACM* 53(3):84–94.

Elkind, E.; Faliszewski, P.; Skowron, P.; and Slinko, A. 2014. Properties of multiwinner voting rules. In *Proceedings of the 13th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, 53–60. IFAAMAS.

Janson, S. 2012. Proportionella valmetoder. Available at http://www2.math.uu.se/~svante/papers/sjV6.pdf.

Janson, S. 2016. Phragmén's and Thiele's election methods. Technical Report arXiv:1611.08826 [math.HO], arXiv.org.

Kilgour, D. M. 2010. Approval balloting for multi-winner elections. In *Handbook on Approval Voting*. Springer. Chapter 6.

Lu, T., and Boutilier, C. 2011. Budgeted social choice: From consensus to personalized decision making. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI)*, 280–286. AAAI Press.

Monroe, B. L. 1995. Fully proportional representation. *The American Political Science Review* 89(4):925–940.

Mora, X., and Oliver, M. 2015. Eleccions mitjançant el vot d'aprovació. El mètode de Phragmén i algunes variants. *Butlletí de la Societat Catalana de Matemàtiques* 30(1):57–101.

Mora, X. 2016. Phragmén's sequential method with a variance criterion. Technical Report arXiv:1611.06833 [math.OC], arXiv.org.

Phragmén, E. 1894. Sur une méthode nouvelle pour réaliser, dans les élections, la représentation proportionnelle des partis. *Öfversigt af Kongliga Vetenskaps-Akademiens Förhandlingar* 51(3):133–137.

Phragmén, E. 1895. *Proportionella val. En valteknisk studie*. Svenska spörsmål 25. Lars Hökersbergs förlag, Stockholm.

Phragmén, E. 1896. Sur la théorie des élections multiples. *Öfversigt af Kongliga Vetenskaps-Akademiens Förhandlingar* 53:181–191.

Phragmén, E. 1899. Till frågan om en proportionell valmetod. *Statsvetenskaplig Tidskrift* 2(2):297–305.

Potthof, R. F., and Brams, S. J. 1998. Proportional representation: Broadening the options. *Journal of Theoretical Politics* 10(2):147–178.

Procaccia, A. D.; Rosenschein, J. S.; and Zohar, A. 2008. On the complexity of achieving proportional representation. *Social Choice and Welfare* 30:353–362.

Sainte-Laguë, A. 1910. La représentation proportionnelle et la méthode des moindres carrés. In *Annales scientifiques de l'École Normale Supérieure*, volume 27, 529–542.

Sánchez-Fernández, L.; Fernández, N.; Fisteus, J. A.; and Basanta Val, P. 2016. Some notes on justified representation. In *Proceedings of the 10th Multidisciplinary Workshop on Advances in Preference Handling (MPREF)*.

Sánchez-Fernández, L.; Elkind, E.; Lackner, M.; Fernández, N.; Fisteus, J. A.; Basanta Val, P.; and Skowron, P. 2017. Proportional justified representation. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI)*. AAAI Press. Forthcoming.

Sánchez-Fernández, L.; Fernández, N.; and Fisteus, J. A. 2016. Fully open extensions to the D'Hondt method. Technical Report arXiv:1609.05370 [cs.GT], arXiv.org.

Skowron, P. K.; Faliszewski, P.; and Lang, J. 2015. Finding a collective set of items: From proportional multirepresentation to group recommendation. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence (AAAI)*, 2131–2137. AAAI Press.

Thiele, T. N. 1895. Om flerfoldsvalg. *Oversigt over det Kongelige Danske Videnskabernes Selskabs Forhandlinger* 415–441.