The Thirty-Second AAAI Conference
on Artificial Intelligence (AAAI-18)

On the Complexity of Extended and Proportional Justified Representation

Haris Aziz
Data61, CSIRO and UNSW
Sydney, Australia
haris.aziz@data61.csiro.com.au

Edith Elkind
University of Oxford
Oxford, UK
elkind@cs.ox.ac.uk

Shenwei Huang
University of New South Wales
Sydney, Australia
dynamichuang@gmail.com

Martin Lackner
TU Wien
Vienna, Austria
lackner@dbai.tuwien.ac.at

Luis Sánchez-Fernández
Universidad Carlos III
de Madrid, Spain
luiss@it.uc3m.es

Piotr Skowron
TU Berlin
Berlin, Germany
p.k.skowron@gmail.com

Abstract

We consider the problem of selecting a fixed-size committee based on approval ballots. It is desirable to have a committee in which all voters are fairly represented. Aziz et al. (2015a; 2017) proposed an axiom called *extended justified representation (EJR)*, which aims to capture this intuition; subsequently, Sánchez-Fernández et al. (2017) proposed a weaker variant of this axiom called *proportional justified representation (PJR)*. It was shown that it is coNP-complete to check whether a given committee provides EJR, and it was conjectured that it is hard to find a committee that provides EJR. In contrast, there are polynomial-time computable voting rules that output committees providing PJR, but the complexity of checking whether a given committee provides PJR was an open problem. In this paper, we answer open questions from prior work by showing that EJR and PJR have the same worst-case complexity: we provide two polynomial-time algorithms that output committees providing EJR, yet we show that it is coNP-complete to decide whether a given committee provides PJR. We complement the latter result by fixed-parameter tractability results.

Introduction

Consider an election where voters have simple preferences: each voter approves some of the candidates and disapproves the remaining candidates; she is indifferent among the candidates in each group. Suppose that the goal is to select a fixed-size set of winners, or *committee*. This model captures a number of applications: the candidates could be potential members of a governing body, items to be shown on a seller's homepage, or tunes to be played at a wedding. Accordingly, there is a number of natural voting rules that take approval ballots as their input and output a set of committees that are tied for winning (Kilgour 2010; Brams and Fishburn 2007; LeGrand, Markakis, and Mehta 2007; Aziz et al. 2015; Skowron, Faliszewski, and Lang 2016; Sánchez-Fernández, Fernández, and Fisteus 2016; Brill et al. 2017). Many of these voting rules attempt to ensure that all groups of voters are fairly represented in the selected committee. However, it has been far from clear how to best capture the representation requirements.

Aziz et al. (2017) proposed a compelling representation axiom called *justified representation (JR)*, as well as a stronger version of this axiom called *extended justified representation (EJR)*. Intuitively, the JR axiom says that every sufficiently large group of voters who jointly approve at least one candidate should be represented in a committee; EJR additionally requires that very large groups whose preferences exhibit significant agreement should be allocated several representatives. While EJR is a rather demanding axiom, every election admits a committee that provides EJR.

Aziz et al. (2017) show that it is easy to check if a given committee provides JR, and to find a committee with this property; indeed, many common approval-based voting rules are guaranteed to output committees that provide JR. The EJR axiom is considerably more challenging from a computational perspective: among the voting rules considered in the literature, there is only one rule, namely, Proportional Approval Voting (PAV) that satisfies EJR (in the sense that every committee output by PAV provides EJR), and computing the winners under this rule is NP-hard (Aziz et al. 2015). Moreover, Aziz et al. (2017) show that it is coNP-complete to check if a given committee provides EJR.

Sánchez-Fernández et al. (2017) put forward an intermediate property called *proportional justified representation (PJR)*: every committee that provides EJR also provides PJR, and every committee that provides PJR also provides JR, but converse implications are not true. An attractive property of PJR is that it is compatible with another important axiom called *perfect representation*, while for EJR this is not the case. Sánchez-Fernández et al. (2017) argued that two well-studied approval-based committee selection rules satisfy PJR when the target committee size k divides the number of voters n; one of these rules is polynomial-time computable. Other authors identified two polynomial-time computable rules that satisfy PJR for all values of k and n (Brill et al. 2017; Sánchez-Fernández, Fernández, and Fisteus 2016). However, the complexity of checking whether a given committee provides PJR remained open. Thus, the existing results seemed to suggest that, from a computational perspective, PJR is more tractable than EJR.

Contributions In this paper, we resolve two main open problems regarding the computational complexity of EJR and PJR.

First, we present two polynomial-time algorithms that output committees satisfying EJR. Our first algorithm is a simple local search algorithm that looks for committees with approximately optimal PAV scores (recall that PAV is the only voting rule that was known to satisfy EJR prior to our work). It exploits an interesting connection between extended justified representation and another important concept called *average satisfaction* (Sánchez-Fernández et al. 2017). Our second algorithm can be viewed as a variant of Single Transferable Vote with fractional vote transfers, adapted to the setting of approval ballots. This algorithm proceeds iteratively, adding candidates to the committee one by one and adjusting the voters' weights according to how well they are represented by the already selected candidates.

Second, we settle the complexity of testing PJR: we prove that checking whether a given committee provides PJR is coNP-complete. We complement this complexity result by showing that PJR and EJR can be tested efficiently if any of the following parameters are small: (1) n (number of voters) (2) m (number of candidates) (3) a (maximum number of candidates approved by a voter) (4) d (maximum number of voters approving a given candidate). More specifically, we provide FPT results for n and m, and XP results for a and d.

Preliminaries

An *election* is a pair $E = (N, C)$, where $N = \{1, \dots, n\}$ is a set of *voters* and $C = \{c_1, \dots, c_m\}$ is a set of *candidates*. Each voter $i \in N$ is associated with an *approval ballot* $A_i \subseteq C$: this is the set of candidates approved by i. For each $c \in C$, we write $N_c = \{i \in N \mid c \in A_i\}$. We are interested in procedures that, given an election E and a positive integer k, $1 \le k \le |C|$, output a non-empty collection of size-k subsets of candidates; such procedures are called *committee selection rules*.

Given an election $E = (N, C)$, we define the PAV-score of a committee $W \subseteq C$ as

$$\text{pav-sc}(W) = \sum_{i=1}^{n} \sum_{j=1}^{|A_i \cap W|} \frac{1}{j}.$$

Proportional Approval Voting (PAV) is the committee selection rule that, given an election E and a committee size k, outputs all size-k committees with the highest PAV-score; finding a committee in the output of this rule is NP-hard (Aziz et al. 2015; Skowron, Faliszewski, and Lang 2016).

In what follows, we consider an election $E = (N, C)$ and a target committee size k, $1 \le k \le |C|$. We say that a group of voters $V \subseteq N$ is *ℓ-large* if $|V| \ge \ell \cdot \frac{n}{k}$; V is *ℓ-cohesive* if $|\bigcap_{i \in V} A_i| \ge \ell$. The following representation axioms have been considered in the literature (Aziz et al. 2017; Sánchez-Fernández et al. 2017):

Justified representation (JR). A committee W provides *justified representation (JR)* if for every 1-large, 1-cohesive group of voters V there exists a voter $i \in V$ who approves a member of W, i.e., $|A_i \cap W| \ge 1$.

The rationale behind this definition is that if k candidates are to be selected, then, intuitively, each group of $\frac{n}{k}$ voters "deserves" a representative. Therefore, a set of $\frac{n}{k}$ voters that have at least one candidate in common should not be completely unrepresented.

Extended justified representation (EJR). A committee W provides *extended justified representation (EJR)* if for every $\ell \in [k]$ and every ℓ-large, ℓ-cohesive group of voters V there exists a voter $i \in V$ who approves at least ℓ members of W, i.e., $|A_i \cap W| \ge \ell$.

Proportional Justified Representation (PJR). A committee W provides *proportional justified representation (PJR)* if for every $\ell \in [k]$ and every ℓ-large, ℓ-cohesive group of voters V there are at least ℓ members of W who are approved by a voter from V, i.e., $|W \cap (\cup_{i \in V} A_i)| \ge \ell$.

The difference between EJR and PJR is that the former requires that some voter in an ℓ-large, ℓ-cohesive group has ℓ representatives in the committee, whereas the latter requires that voters in that group collectively approve of ℓ committee members. It is immediate that EJR implies PJR, and PJR implies JR.

We have defined different variants of justified representation as properties of committees. We extend these definitions to committee selection rules by saying that a committee selection rule R satisfies a property $\mathcal{P}$ if for every election $E = (N, C)$ and every k with $1 \le k \le |C|$ all committees in the output of $R(E, k)$ provide $\mathcal{P}$.

Further, Sánchez-Fernández et al. (2017) defined the notion of average satisfaction, which relates to the concepts of justified representation.

Definition 1 (Average satisfaction.)**.** *Given a committee W, the* average satisfaction of a group of voters $V \subseteq N$ with respect to W *is defined as*

$$\text{avs}_W(V) = \frac{1}{|V|} \sum_{i \in V} |A_i \cap W|.$$

Sánchez-Fernández et al. (2017) showed that if a committee W provides EJR then for every $\ell > 0$ and every ℓ-large, ℓ-cohesive group of voters V it holds that the average satisfaction of V with respect to W is at least $\frac{\ell-1}{2}$. We observe that, conversely, if a committee offers very high average satisfaction to all large cohesive groups then it provides EJR.

Lemma 1. *Consider an election $E = (N, C)$ and a committee W. If for every ℓ-large, ℓ-cohesive group of voters it holds that its average satisfaction with respect to W is strictly greater than $\ell - 1$, then W provides EJR.*

Proof. Let V be an ℓ-large, ℓ-cohesive group. Since the average satisfaction of voters in V is greater than $\ell - 1$, there is at least one voter $i \in V$ with $|A_i \cap W| \ge \ell$. □

Optimal Average Satisfaction and a Local Search Algorithm

In this section, we show that a committee providing EJR can be computed in polynomial time. We define a new rule, LS-PAV, which is an approximate local search algorithm for PAV (see Algorithm 1).

Algorithm 1: LS-PAV: a local search algorithm for PAV

$W \leftarrow k$ arbitrary candidates from C
while *there exist* $c \in W$ *and* $c' \in C \setminus W$ *such that*

$\text{pav-sc}\big((W \setminus \{c\}) \cup \{c'\}\big) \geq \text{pav-sc}(W) + \frac{n}{k^2}$ **do**
$W \leftarrow (W \setminus \{c\}) \cup \{c'\}$
return W

Theorem 1. *Consider an election $E = (N, C)$ and a positive integer k, $1 \leq k \leq |C|$. Let W be a winning committee chosen by* LS-PAV *on (E, k). Then for every $\ell > 0$ and every ℓ-large, ℓ-cohesive group V it holds that $\text{avs}_W(V) > \ell - 1$.*

Proof. Assume for the sake of contradiction that for some (E, k) LS-PAV outputs a committee W such that there exists an ℓ-large, ℓ-cohesive group V with $\text{avs}_W(V) \leq \ell - 1$. Let $w_i = |W \cap A_i|$.

As V is ℓ-cohesive, there exist ℓ candidates approved by all voters in V. At least one such candidate does not appear in W, since otherwise we would have $\text{avs}_W(V) \geq \ell$. Let c be some such candidate. Now consider a candidate $c' \in W$. If we remove c' from the committee and add c instead, we increase the PAV-score of W by

$$\Delta(c, c') \geq \underbrace{\sum_{i \in V : c' \notin A_i} \frac{1}{w_i + 1}}_{\text{adding } c} - \underbrace{\sum_{i \in N \setminus V \,:\, c' \in A_i} \frac{1}{w_i}}_{\text{removing } c'}$$
$$= \sum_{i \in V} \frac{1}{w_i + 1} - \sum_{i \in N \,:\, c' \in A_i} \frac{1}{w_i} + \sum_{i \in V \,:\, c' \in A_i} \left(\frac{1}{w_i} - \frac{1}{w_i + 1} \right).$$

Note that $\Delta(c, c')$ may be negative for some $c' \in W$. By the inequality between arithmetic and harmonic means we obtain

$$\sum_{i \in V} \frac{1}{w_i + 1} \geq \frac{|V|^2}{\sum_{i \in V}(w_i + 1)} = \frac{|V|}{\frac{\sum_{i \in V} w_i}{|V|} + 1} = \frac{|V|}{\text{avs}_W(V) + 1} \geq \frac{|V|}{\ell}. \tag{1}$$

Now, observe that

$$\sum_{c' \in W} \Delta(c, c') \geq \sum_{c' \in W} \left(\sum_{i \in V} \frac{1}{w_i + 1} - \sum_{i \in N \,:\, c' \in A_i} \frac{1}{w_i} + \sum_{i \in V \,:\, c' \in A_i} \left(\frac{1}{w_i} - \frac{1}{w_i + 1} \right) \right)$$
$$= k \sum_{i \in V} \frac{1}{w_i + 1} - \sum_{c' \in W} \sum_{i \in N \,:\, c' \in A_i} \frac{1}{w_i} + \sum_{c' \in W} \sum_{i \in V \,:\, c' \in A_i} \left(\frac{1}{w_i} - \frac{1}{w_i + 1} \right)$$
$$= k \sum_{i \in V} \frac{1}{w_i + 1} - \sum_{i \in N} \sum_{c' \in W \cap A_i} \frac{1}{w_i} + \sum_{i \in V} \sum_{c' \in W \cap A_i} \left(\frac{1}{w_i} - \frac{1}{w_i + 1} \right)$$
$$\geq k \sum_{i \in V} \frac{1}{w_i + 1} - n + |V| - \sum_{i \in V} \frac{w_i}{w_i + 1}$$
$$= k \sum_{i \in V} \frac{1}{w_i + 1} - n + |V| - \sum_{i \in V} \left(1 - \frac{1}{w_i + 1} \right)$$
$$= (k + 1) \sum_{i \in V} \frac{1}{w_i + 1} - n.$$

Further, by Equation (1) we obtain

$$k \sum_{i \in V} \frac{1}{w_i + 1} - n \geq \frac{k|V|}{\ell} - n \geq \left\lceil \frac{\ell n}{k} \right\rceil \cdot \frac{k}{\ell} - n \geq 0,$$

and hence

$$\sum_{c' \in W} \Delta(c, c') \geq (k + 1) \sum_{i \in V} \frac{1}{w_i + 1} - n \geq \sum_{i \in V} \frac{1}{w_i + 1} \geq \frac{|V|}{\ell} \geq \frac{n}{k}.$$

From the pigeonhole principle it follows that there exists a candidate $c' \in W$ such that $\Delta(c, c') \geq \frac{n}{k^2}$, which means that W could not have been returned by our local search algorithm. This completes the proof. □

Note that Theorem 1 applies not only to LS-PAV but also to PAV, since PAV selects a committee with maximum PAV-score.

Corollary 1. *For every committee in the output of PAV the average satisfaction of every ℓ-large ℓ-cohesive group of voters is strictly greater than $\ell - 1$.*

Corollary 2. LS-PAV *satisfies extended justified representation.*

Proof. Let $E = (N, C)$ be an election, let k be a positive integer with $1 \leq k \leq |C|$ and let W be a winning committee chosen by LS-PAV. Further, let V be an ℓ-large, ℓ-cohesive group. Then, by Theorem 1 it holds that $\text{avs}_W(V) > \ell - 1$. Consequently, by Lemma 1, W provides EJR. □

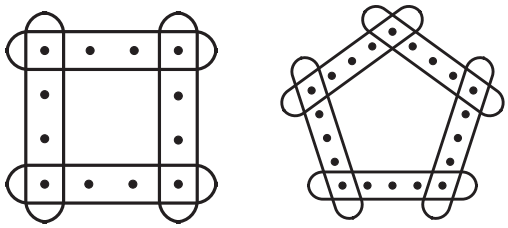
Proposition 1. LS-PAV *runs in polynomial time.*

Proof. A single improving swap can be found and executed in polynomial time. Now, let us assess how many improvements the local search algorithm may perform. Each improvement increases the total PAV-score of a committee by at least $\frac{n}{k^2}$. The maximum PAV-score of a size-k committee is $n \cdot (1 + 1/2 + \cdots + 1/k) = O(n \ln k)$. Thus, there can be at most $O(k^2 \ln k)$ improving swaps. □

Observe that Proposition 1 relies on having a threshold of $\frac{n}{k^2}$ in the definition of the local search algorithm. If we perform a swap each time when $\text{pav-sc}\big((W \setminus \{c\}) \cup \{c'\}\big) > \text{pav-sc}(W)$, this could potentially lead to a superpolynomial running time, as the improvement in score could be exponentially small in k.

Average Satisfaction Guarantees

In many scenarios it is desirable to have committees that provide high average satisfaction. One may then wonder if the guarantee provided by Theorem 1 can be further improved. We will now argue that this guarantee is, in fact, optimal.



(a) Example for $k = 3$ (b) Example for $k = 4$

Figure 1: A visualization of the profiles used in Example 1. Dots represent voters and boxes represent candidates; candidates are approved by voters contained in the respective boxes.

Example 1. Consider an election with the following approval ballots.

1 voter: $\{d, a\}$	2 voters: $\{a\}$
1 voter: $\{a, b\}$	2 voters: $\{b\}$
1 voter: $\{b, c\}$	2 voters: $\{c\}$
1 voter: $\{c, d\}$	2 voters: $\{d\}$

This profile is schematically shown in Figure 1a. For $k = 3$, we have $\frac{n}{k} = 4$ and consequently for each candidate c all voters in N_c form a 1-large, 1-cohesive group. The profile is symmetric with respect to candidates so without loss of generality assume that committee $\{a, b, c\}$ is chosen. There are four voters who approve d: one voter with ballot $\{d, a\}$, two voters with ballot $\{d\}$, and one voter with ballot $\{c, d\}$. The average satisfaction of this group is $1/2$. Thus, it is impossible to guarantee an average satisfaction to 1-large, 1-cohesive groups that is better than $\frac{1}{2}$.

To extend this example to $k = 4$, we move from a square shape to a pentagon (see Figure 1b).

1 voter: $\{e, a\}$	3 voters: $\{a\}$
1 voter: $\{a, b\}$	3 voters: $\{b\}$
1 voter: $\{b, c\}$	3 voters: $\{c\}$
1 voter: $\{c, d\}$	3 voters: $\{d\}$
1 voter: $\{d, e\}$	3 voters: $\{e\}$

By the same argument as before, we can assume without loss of generality that the winning committee does not contain e. The average satisfaction of the voters who approve e is $\frac{2}{5}$. In general, for each $k \geq 3$ we can construct an election E_k with $n = k(k+1)$ voters and $k+1$ candidates where for every committee the average satisfaction of some 1-large, 1-cohesive groups is $\frac{2}{k+1}$. Hence, there is no positive constant γ such that we can guarantee an average satisfaction of γ to 1-large, 1-cohesive groups for all values of k.

The construction in Example 1 shows that Theorem 1 is tight for $\ell = 1$. We can extend this result to all values of ℓ.

Proposition 2. *Let ℓ be a positive integer and let γ be a positive constant. There exists $k > 0$ and an election (N, C) such that no committee $W \subset C$ with $|W| = k$ provides average satisfaction of $\ell - 1 + \gamma$ to all ℓ-large, ℓ-cohesive groups of voters in V.*

Proof. By considering the sequence of elections $E_3, E_4, \ldots$ from Example 1 and choosing $k > \frac{2}{\gamma}$, we obtain our claim for $\ell = 1$. This argument extends to other values of ℓ as follows. Given an election E_k from our sequence, we set $k' = k \cdot \ell$, replace each candidate with ℓ copies, and modify the voters' preferences so that each voter who approved of c in $E_k = (N_k, C_k)$ now approves of all ℓ copies of c; denote the new election by E'_k. By construction, for each $c \in C_k$ we have $|N_c| = \frac{n}{k} = \frac{\ell n}{k'}$. As all voters in N_c approve all ℓ copies of c in E'_k, this group is ℓ-large and ℓ-cohesive in E'_k. Now, pick a committee W of size k' in E'_k. There is some candidate $c \in C_k$ such that at most $\ell - 1$ copies of c are included in W, so in E'_k there are $\frac{\ell n}{k'} - 2$ voters in N_c who approve at most $\ell - 1$ members of W, and the remaining two voters in N_c approve at most $2\ell - 1$ members of W. Thus, the average satisfaction of the voters in N_c in E'_k is at most

$$\frac{2(2\ell - 1) + (\frac{\ell n}{k'} - 2)(\ell - 1)}{\frac{\ell n}{k'}}$$
$$= \ell - 1 + \frac{k'(4\ell - 2 - 2\ell + 2)}{\ell n} = \ell - 1 + \frac{2k\ell}{n}.$$

As in E_k we have $n = k(k+1)$, this means that for $k > \frac{2\ell}{\gamma}$ the average satisfaction of this group in E'_k is less than $\ell - 1 + \gamma$, as claimed. □

An Iterative Algorithm

In this section, we describe a family of iterative voting rules, which we call **EJR-Exact**, and argue that all rules in this family provide EJR. Throughout this section, we fix an election (N, C) with $|N| = n$, $|C| = m$, and assume that the target committee size is k. For readability, we first present a proof outline, followed by technical proofs.

Briefly, each **EJR-Exact** rule starts with an empty committee $W = \emptyset$ and adds candidates to W one by one, until $|W| = k$ or none of the remaining candidates is 'safe', in a precise sense to be formalized below; if $|W| < k$ at that point, some $k - |W|$ candidates are added to W. Each voter is associated with a weight, which is initially set to 1; whenever a 'safe' candidate c is added to W, the total weight of all voters in N_c is reduced by $\frac{n}{k}$, following certain rules (this is why we call such rules **EJR-Exact**: at each step the removed weight is *exactly* the quota $\frac{n}{k}$). These rules ensure that voters who are not yet adequately represented in the committee get to keep some of their weight, to wield some power in future iterations. A necessary (but not sufficient!) condition for a candidate to be 'safe' is that the total weight of voters who approve him is at least $\frac{n}{k}$.

In more detail, for each $j \in [k] \cup \{0\}$ and each $i \in N$, let f_i^j denote the weight of voter i after the j-th iteration,

and let W_j denote the set of elected candidates after the j-th iteration; we set $f_i^0 = 1$ for all $i \in N$ and $W_0 = \emptyset$.

To explain what it means for a candidate to be safe at iteration j, we need some preliminary definitions.

Definition 2. *Given a set of candidates $W \subseteq C$ and $\ell \in [k]$, we say that a candidate $c \in C \setminus W$ is ℓ-*plausible with respect to W *if there is a set of voters $N' \subseteq N_c$ such that $|N'| \geq \ell \cdot \frac{n}{k}$ and $|A_i \cap W| < \ell$ for each $i \in N'$. The* plausibility level *of a candidate $c \in C \setminus W$ with respect to W is*

$$\mathrm{pl}(c, W) = \max\{\ell \mid c \text{ is } \ell\text{-plausible with respect to } W\},$$

with the convention that $\max \emptyset = 0$.

The entitlement of a voter $i \in N$ with respect to W *is*

$$\mathrm{en}(i, W) = \max_{c \in A_i \setminus W} \mathrm{pl}(c, W).$$

In words, if a candidate c is ℓ-plausible with respect to W, there exists an ℓ-large set of voters N' who all approve c, but have fewer than ℓ representatives in W; while N' may fail to be ℓ-cohesive, it is necessarily 1-cohesive, as all voters in it approve c. Thus, the plausibility level of a candidate provides some indication of how dangerous he is, in terms of causing violations of EJR. The concept of entitlement is motivated by similar reasoning: if $\mathrm{en}(i, W)$ is small, we do not have to worry about EJR violations involving i.

Specifically, it can be shown that if $\mathrm{en}(i, W_j) \leq |A_i \cap W_j|$ at some iteration j, voter i cannot be in a group witnessing a violation of EJR (we omit the proof). In contrast, if $\mathrm{en}(i, W_j) > |A_i \cap W_j|$, voter i should retain some weight after iteration j. To this end, for each $j \in [k-1] \cup \{0\}$, each $c \notin W_j$, and each $i \in N_c$ with $\mathrm{en}(i, W_j \cup \{c\}) > |A_i \cap (W_j \cup \{c\})|$, we set

$$g_i^{j+1}(c) = 1 - \frac{|A_i \cap (W_j \cup \{c\})|}{\mathrm{en}(i, W_j \cup \{c\})};$$

if $\mathrm{en}(i, W_j \cup \{c\}) \leq |A_i \cap (W_j \cup \{c\})|$, we set $g_i^{j+1}(c) = 0$. Intuitively, $g_i^{j+1}(c)$ is a lower bound on the weight that i should retain if c is added at iteration $j+1$.

We are now ready to classify candidates according to the amount of support they have after each iteration.

Definition 3. *Given a $j \in [k-1] \cup \{0\}$, we say that after iteration j a candidate $c \in C \setminus W_j$ is*

- weak *if $\sum_{i \in N_c} f_i^j < \frac{n}{k}$;*
- risky *if $\frac{n}{k} \leq \sum_{i \in N_c} f_i^j < \frac{n}{k} + \sum_{i \in N_c} g_i^{j+1}(c)$, and*
- safe *if $\frac{n}{k} + \sum_{i \in N_c} g_i^{j+1}(c) \leq \sum_{i \in N_c} f_i^j$.*

Note that we can compute the plausibility level of each candidate and hence the voters' entitlements at the end of each iteration in polynomial time; consequently, we can decide in polynomial time which category a given candidate belongs to at the end of a given iteration.

With these definitions in hand, we can describe **EJR-Exact** rules more precisely. We perform at most k iterations. After each iteration j, $j = 0, \ldots, k-1$, we identify all safe candidates. If this set is non-empty, then during iteration $j+1$ we add some safe candidate c to the committee (i.e., we set $W_{j+1} = W_j \cup \{c\}$), and update the voters' weights so that

(a) $\sum_{i \in N} f_i^{j+1} = \sum_{i \in N} f_i^j - \frac{n}{k}$,

(b) $f_i^{j+1} \geq g_i^{j+1}(c)$ for each $i \in N_c$, and

(c) $f_i^{j+1} = f_i^j$ for each $i \notin N_c$.

Note that conditions (a)–(c) can be satisfied simultaneously exactly because c is safe. If at the end of iteration j none of the candidates in $C \setminus W_j$ is safe, we stop and return some committee of size k that contains W_j.

This completes the formal description of **EJR-Exact** rules. The rules in this family may differ in (i) how they choose a safe candidate; (ii) how they update the voters' weights subject to conditions (a)–(c), and (iii) how they pick additional $k - |W_j|$ candidates when no safe candidates are available. Importantly, it is possible to make these choices so as to obtain a rule that is polynomial-time computable.

To establish that every **EJR-Exact** rule always returns a committee that satisfies EJR, we prove the following claims.

Proposition 3. *If after iteration j no candidate in $C \setminus W_j$ is safe, then all candidates in $C \setminus W_j$ are weak.*

Proposition 4. *Suppose that in each of the first j iterations, $j \in [k]$, we added a safe candidate to the committee, but after iteration j all remaining candidates are weak. Then every size-k committee that contains W_j provides EJR.*

Propositions 3 and 4 imply that all **EJR-Exact** rules satisfy EJR. Indeed, if we have performed k iterations, adding a candidate to the committee at each step, then we have removed $k \cdot \frac{n}{k} = n$ units of vote weight, and hence all remaining candidates are weak. On the other hand, if we stopped earlier because no candidate was safe, then by Proposition 3 all remaining candidates are weak. In either case Proposition 4 directly implies that the output committee provides EJR. Thus, we obtain the following theorem.

Theorem 2. *Every rule in the* **EJR-Exact** *family provides EJR. Moreover, this family contains rules that are polynomial-time computable.*

Proofs

In this section we provide proofs of Propositions 3 and 4. We first present three auxiliary lemmas.

Our first lemma provides lower bounds on voters' weights after iteration j.

Lemma 2. *Suppose that during each of the first j iterations, a safe candidate has been added to the committee. Then for each voter $i \in N$ with $\mathrm{en}(i, W_j) > |A_i \cap W_j|$ we have*

$$f_i^j \geq 1 - \frac{|A_i \cap W_j|}{\mathrm{en}(i, W_j)}.$$

Proof. Fix a voter $i \in N$ such that $\mathrm{en}(i, W_j) > |A_i \cap W_j|$. If $A_i \cap W_j = \emptyset$, then $f_i^j = 1$ and we are done. Otherwise, let r be the last iteration when a candidate $c \in A_i$ was added to the committee. Note that $W_r \subseteq W_j$ implies $\mathrm{en}(i, W_r) \geq \mathrm{en}(i, W_j)$ and hence $\mathrm{en}(i, W_r) > |A_i \cap W_j| = |A_i \cap W_r|$. As $W_r = W_{r-1} \cup \{c\}$, we have

$$f_i^j = f_i^r \geq g_i^r(c) = 1 - \frac{|A_i \cap W_r|}{\mathrm{en}(i, W_r)}.$$

As $|A_i \cap W_r| = |A_i \cap W_j|$ and $\mathrm{en}(i, W_r) \geq \mathrm{en}(i, W_j)$, the claim follows. □

Our next lemma shows that the current weight of each voter approving a candidate $c \in C \setminus W_j$ is at least as large as what it would have to be after adding c to the committee.

Lemma 3. *Suppose that during each of the first j iterations, a safe candidate has been added to the committee, and consider a candidate $c \in C \setminus W_j$ and a voter $i \in N_c$. Then*

$$f_i^j \geq g_i^{j+1}(c).$$

Proof. If $\mathrm{en}(i, W_j \cup \{c\}) \leq |A_i \cap (W_j \cup \{c\})|$, the claim is trivially true, because $g_i^{j+1}(c) = 0$ in this case, and our weight transfer procedure guarantees that all weights remain non-negative.

Thus, suppose that $\mathrm{en}(i, W_j \cup \{c\}) > |A_i \cap (W_j \cup \{c\})|$. Then we have $|A_i \cap W_j| < |A_i \cap (W_j \cup \{c\})|$ and $\mathrm{en}(i, W_j) \geq \mathrm{en}(i, W_j \cup \{c\})$, so

$$1 - \frac{|A_i \cap W_j|}{\mathrm{en}(i, W_j)} \geq 1 - \frac{|A_i \cap (W_j \cup \{c\})|}{\mathrm{en}(i, W_j \cup \{c\})} = g_i^{j+1}(c),$$

and our claim follows from Lemma 2. □

Our next lemma is used in the proof of Proposition 4.

Lemma 4. *Suppose that during each of the first j iterations, a safe candidate has been added to the committee. Consider a candidate $c \in C \setminus W_j$ with $\mathrm{pl}(c, W_j) \geq 1$ and define*

$$S_c = \{i \in N_c : |A_i \cap W_j| < \mathrm{pl}(c, W_j)\}.$$

Then it holds that

$$\sum_{i \in S_c} \left(f_i^j - \frac{\mathrm{pl}(c, W_j) - |A_i \cap W_j| - 1}{\mathrm{pl}(c, W_j)} \right) \geq \frac{n}{k}.$$

Proof. For each voter $i \in N_c$ we have $c \in A_i \setminus W_j$ and hence $\mathrm{en}(i, W_j) \geq \mathrm{pl}(c, W_j)$. As $S_c \subseteq N_c$, this means that for each voter $i \in S_c$ it holds that $|A_i \cap W_j| < \mathrm{en}(i, W_j)$. Further, for every voter $i \in N_c$ with $|A_i \cap W_j| < \mathrm{en}(i, W_j)$ we have

$$f_i^j \geq 1 - \frac{|A_i \cap W_j|}{\mathrm{en}(i, W_j)} \geq 1 - \frac{|A_i \cap W_j|}{\mathrm{pl}(c, W_j)}, \tag{2}$$

where the first inequality holds by Lemma 2 and the second inequality holds because $\mathrm{en}(i, W_j) \geq \mathrm{pl}(c, W_j)$. Using (2), we obtain

$$\begin{aligned} &\sum_{i \in S_c} \left(f_i^j - \frac{\mathrm{pl}(c, W_j) - |A_i \cap W_j| - 1}{\mathrm{pl}(c, W_j)} \right) \\ &\geq \sum_{i \in S_c} \left(1 - \frac{|A_i \cap W_j|}{\mathrm{pl}(c, W_j)} - \frac{\mathrm{pl}(c, W_j) - |A_i \cap W_j| - 1}{\mathrm{pl}(c, W_j)} \right) \\ &= |S_c| \cdot \frac{1}{\mathrm{pl}(c, W_j)}. \end{aligned} \tag{3}$$

By definition of $\mathrm{pl}(c, W_j)$, there are at least $\frac{n}{k} \cdot \mathrm{pl}(c, W_j)$ voters in N_c who have fewer than $\mathrm{pl}(c, W_j)$ approved candidates in W_j. Thus, $|S_c| \geq \frac{n}{k} \cdot \mathrm{pl}(c, W_j)$. Together with (3), this completes the proof. □

We are now ready to prove Proposition 4.

Proof of Proposition 4. Consider a committee W such that $|W| = k$ and $W_j \subseteq W$. Suppose for the sake of contradiction that W violates EJR. Then there exists an $\ell > 0$ and an ℓ-large ℓ-cohesive set of voters N' such that $|A_i \cap W| < \ell$ for each $i \in N'$. Consequently, there exists a candidate $c \in (\cap_{i \in N'} A_i) \setminus W$. This candidate is ℓ-plausible with respect to W and hence also with respect to W_j.

Let

$$S_c = \{i \in N_c : |A_i \cap W_j| < \mathrm{pl}(c, W_j)\}.$$

For each $i \in S_c$ we have

$$\frac{\mathrm{pl}(c, W_j) - |A_i \cap W_j| - 1}{\mathrm{pl}(c, W_j)} \geq 0.$$

As $\mathrm{pl}(c, W_j) > 0$, we can invoke Lemma 4 to obtain

$$\begin{aligned} \sum_{i \in N_c} f_i^j &\geq \sum_{i \in S_c} f_i^j \\ &\geq \sum_{i \in S_c} \left(f_i^j - \frac{\mathrm{pl}(c, W_j) - |A_i \cap W_j| - 1}{\mathrm{pl}(c, W_j)} \right) \geq \frac{n}{k}, \end{aligned}$$

so c would not be a weak candidate after iteration j, a contradiction. □

Finally, we prove Proposition 3.

Proof of Proposition 3. Suppose that we have run j iterations, adding a safe candidate to the committee at each iteration, and at the end of iteration j no candidate is safe. Suppose for the sake of contradiction that some candidate in $C \setminus W_j$ is risky at that point. Let c be a risky candidate that has the highest plausibility level among all risky candidates. To obtain a contradiction, we will identify another candidate c' so that either (a) c' is safe or (b) c' is risky, but $\mathrm{pl}(c', W_j) > \mathrm{pl}(c, W_j)$.

We will first argue that $\mathrm{pl}(c, W_j) > 0$. We have

$$\sum_{i \in N_c} f_i^j \geq \frac{n}{k}, \qquad \sum_{i \in N_c} \left(f_i^j - g_i^{j+1}(c) \right) < \frac{n}{k}.$$

Suppose for the sake of contradiction that $\mathrm{pl}(c, W_j) = 0$. For each voter $i \in N_c$ with $\mathrm{en}(i, W_j \cup \{c\}) = 0$ we have $g_i^{j+1}(c) = 0$. Therefore, there has to be a voter $i' \in N_c$ with $\mathrm{en}(i', W_j \cup \{c\}) > 0$, as otherwise we would have

$$\sum_{i \in N_c} \left(f_i^j - g_i^{j+1}(c) \right) = \sum_{i \in N_c} f_i^j \geq \frac{n}{k}.$$

Thus, there is a candidate $c' \in A_{i'} \setminus (W_j \cup \{c\})$ such that $\mathrm{pl}(c', W_j) > 0$. Let

$$S_{c'} = \{i \in N_{c'} : |A_i \cap W_j| < \mathrm{pl}(c', W_j)\}.$$

As $S_{c'} \subseteq N_{c'}$, from Lemma 4 we conclude

$$\sum_{i \in N_{c'}} f_i^j \geq \sum_{i \in S_{c'}} \left(f_i^j - \frac{\mathrm{pl}(c', W_j) - |A_i \cap W_j| - 1}{\mathrm{pl}(c', W_j)} \right) \geq \frac{n}{k},$$

so c' cannot be weak. However, if c' is safe, we obtain a contradiction with our assumption that there are no safe candidates at the end of iteration j, and if c' is risky, we have a contradiction with our choice of c, as $\mathrm{pl}(c', W_j) > 0 = \mathrm{pl}(c, W_j)$.

Thus, assume that $\mathrm{pl}(c, W_j) > 0$. Set

$$S_c = \{i \in N_c : |A_i \cap W_j| < \mathrm{pl}(c, W_j)\}.$$

Combining Lemma 4 with the fact that c is risky, we obtain

$$\sum_{i \in N_c} \left(f_i^j - g_i^{j+1}(c)\right) - \sum_{i \in S_c} \left(f_i^j - \frac{\mathrm{pl}(c, W_j) - |A_i \cap W_j| - 1}{\mathrm{pl}(c, W_j)}\right) < 0.$$

By Lemma 3, we can write

$$\sum_{i \in N_c} \left(f_i^j - g_i^{j+1}(c)\right) - \sum_{i \in S_c} \left(f_i^j - \frac{\mathrm{pl}(c, W_j) - |A_i \cap W_j| - 1}{\mathrm{pl}(c, W_j)}\right) \geq \sum_{i \in S_c} \left(\frac{\mathrm{pl}(c, W_j) - |A_i \cap W_j| - 1}{\mathrm{pl}(c, W_j)} - g_i^{j+1}(c)\right).$$

This implies that for some voter $i \in S_c$ we have

$$1 - \frac{|A_i \cap W_j| + 1}{\mathrm{pl}(c, W_j)} < g_i^{j+1}(c). \tag{4}$$

Since $|A_i \cap W_j| < \mathrm{pl}(c, W_j)$, we have

$$1 - \frac{|A_i \cap W_j| + 1}{\mathrm{pl}(c, W_j)} \geq 0.$$

Thus, $g_i^{j+1}(c) > 0$, which means that $\mathrm{en}(i, W_j \cup \{c\}) > |A_i \cap (W_j \cup \{c\})|$ and

$$g_i^{j+1}(c) = 1 - \frac{|A_i \cap (W_j \cup \{c\})|}{\mathrm{en}(i, W_j \cup \{c\})}.$$

Substituting this expression into (4), we obtain

$$1 - \frac{|A_i \cap (W_j \cup \{c\})|}{\mathrm{pl}(c, W_j)} < 1 - \frac{|A_i \cap (W_j \cup \{c\})|}{\mathrm{en}(i, W_j \cup \{c\})},$$

which implies that $\mathrm{en}(i, W_j \cup \{c\}) > \mathrm{pl}(c, W_j)$. Thus, there exists a candidate $c' \in A_i \setminus W_j$ such that $\mathrm{pl}(c', W_j) > \mathrm{pl}(c, W_j) > 0$. Again, using Lemma 4, we can argue that c' is not weak. Hence, we obtain a contradiction with our choice of c in this case as well. □

Complexity of Testing PJR and EJR

In this section, we settle the complexity of testing PJR by proving that this problem is coNP-complete. The proof is inspired by the proof of a similar statement for EJR (Aziz et al. 2017).

Theorem 3. *Given an election (N, C), a target committee size k, and a committee W, $|W| = k$, it is* coNP*-complete to check whether W provides PJR for (N, C) and k.*

Proof. It is easy to see that this problem is in coNP: A set of voters $X \subset N$ such that $|X| \geq \ell \frac{n}{k}$, $|\cap_{i \in X} A_i| \geq \ell$ and $|W \cap (\cup_{i \in X} A_i)| < \ell$ is a certificate that W violates PJR.

For the hardness proof, we reduce the BALANCED BICLIQUE problem ([GT24] in Garey and Johnson, 1979) to the complement of our problem. An instance of BALANCED BICLIQUE consists of a bipartite graph (L, R, E) with parts L and R and edge set E, as well as an integer ℓ; it is a "yes"-instance if there exist vertex subsets $L' \subseteq L$ and $R' \subseteq R$ such that $|L'| = |R'| = \ell$ and $(u, v) \in E$ for each $u \in L', v \in R'$, and a "no"-instance otherwise.

For each instance $\langle (L, R, E), \ell \rangle$ of BALANCED BICLIQUE with $R = \{v_1, \ldots, v_s\}$, we design an instance of our problem as follows. Assume without loss of generality that $s \geq 3$, $\ell \geq 3$. We construct three pairwise disjoint sets of candidates C_0, C_1 and C_2, so that $C_0 = L$, $|C_1| = \ell - 1$, $|C_2| = s\ell + \ell - 3s + (\ell - 2)$, and set $C = C_0 \cup C_1 \cup C_2$. We then construct three sets of voters N_0, N_1, N_2, so that $N_0 = \{1, \ldots, s\}$, $|N_1| = s\ell$, $|N_2| = s\ell + \ell - 3s + (\ell - 2)$; observe that $|N_2| \geq \ell - 1$ since we assume that $\ell \geq 3$. For every $i \in N_0$ we set $A_i = \{u_j \mid (u_j, v_i) \in E\}$, and for every $i \in N_1$ we set $A_i = C_0 \cup C_1$. The candidates in C_2 are matched to voters in N_2: each voter in N_2 approves exactly one candidate in C_2, and each candidate in C_2 is approved by exactly one voter in N_2. Denote the resulting election by F. Finally, we set $k = 2\ell - 2$, and let $W = C_1 \cup X$, where X is a subset of C_2 with $|X| = \ell - 1$. Note that the number of voters n is given by $s + s\ell + s\ell + \ell - 3s + (\ell - 2) = 2(s + 1)(\ell - 1)$, so $\frac{n}{k} = s + 1$.

Suppose first that we started with a "yes"-instance of BALANCED BICLIQUE, and let (L', R') be the respective ℓ-by-ℓ biclique. Let $C^* = L'$ and $N^* = R' \cup N_1$. Then $|N^*| = \ell(s + 1) = \ell \frac{n}{k}$, all voters in N^* approve all candidates in C^*, $|C^*| = \ell$, but all voters in N^* together are only represented by $\ell - 1$ candidates in W. Hence, for this value of ℓ the set W fails to provide PJR for F and k.

Conversely, suppose that W fails to provide PJR for F and k. Then there exists a value $j > 0$, a set N^* of $j(s+1)$ voters and a set C^* of j candidates so that all voters in N^* approve of all candidates in C^*, but all voters in N^* together are only represented by fewer than j candidates in W. Note that, since $s > 1$ and $j \geq 1$, we have $N^* \cap N_2 = \emptyset$. Further, since $|N^*| = j(s + 1) \geq s + 1$ and $|N_0| = s$, it follows that N^* contains one voter from N_1. So, all voters in N^* together are represented by exactly $\ell - 1$ candidates in W. This implies that $j \geq \ell$. As $N^* = j(s + 1) \geq \ell(s + 1)$, it follows that $|N^* \cap N_0| \geq \ell$. Since N^* contains voters from both N_0 and N_1, it follows that $C^* \subseteq C_0$. Thus, there are at least ℓ voters in $N^* \cap N_0$ who approve the same $j \geq \ell$ candidates in C_0; any set of ℓ such voters and ℓ such candidates corresponds to an ℓ-by-ℓ biclique in the input graph. □

We complement our hardness result by showing that testing PJR and EJR is computationally tractable if one of the following parameters is bounded: (1) $n = |N|$; (2) $m = |C|$; (3) $a = \max_{i \in N} |A_i|$ (maximum size of approval sets); (4) $d = \max_{c \in C} |N_c|$ (maximum number of approvals of a candidate). For the first two parameters, we show that

our problem belongs to the class *FPT (fixed-parameter tractable)*; for the remaining two parameters we place it in the class XP.

The proof of the following theorem is omitted due to space constraints.

Theorem 4. *PJR and EJR can be tested*

(1) in time $O(2^n mnk)$*;*

(2) in time $O(2^m m^2 n)$*;*

(3) in time $O(m^{a+2} n)$*;*

(4) in time $O(n^{d+1} mk)$*.*

Discussion

Our results show that, surprisingly, EJR and PJR have the same worst-case complexity: while a committee providing one of these properties can be computed in polynomial time, testing whether a given committee provides PJR or EJR is coNP-complete. One can still argue that PJR is somewhat more tractable because it is satisfied by a well-established polynomial-time computable voting rule, namely, the sequential version of Phragmén's rule (Phragmén 1894; Janson 2016; Brill et al. 2017). In contrast, the two polynomial-time procedures for computing committees that provide EJR that have been proposed in this paper have been specifically engineered with the goal of satisfying EJR in mind. Indeed, it seems that the local swap-based procedure is unlikely to be popular with voters: it aims to optimize a certain quantity, yet stops before reaching the optimum. In contrast, our sequential procedure is similar in spirit to existing voting rules; however, the weight update mechanism is too complicated for an average voter to comprehend.

We note that **EJR-Exact** is not just a single rule, but a family of rules. In particular, we can fine-tune the weight update mechanism to derive rules with further desirable properties; understanding the full power of this family of rules is a topic for future work. Similarly, we can tweak LS-PAV by selecting the initial committee in a particular way: for instance, we can use the output of another polynomial-time computable voting rule as a starting point. This can both improve the running time and bias the algorithm towards certain outputs; again, this is a research direction that deserves further attention.

Acknowledgments

This paper is based on a merge of four technical reports (Aziz and Huang 2016; 2017; Skowron et al. 2017; Sánchez-Fernández, Elkind, and Lackner 2017). The authors are grateful to the anonymous AAAI reviewers for their helpful comments. Aziz was supported by a Julius Career Award. Elkind and Lackner were supported by the European Research Council (ERC) under grant number 639945 (ACCORD); Lackner was also supported by the Austrian Science Foundation FWF, grants P25518 and Y698. Sánchez-Fernández was supported by the Spanish Ministerio de Economía y Competitividad (project AUDACity TIN2016-77158-C4-1-R) and by the Autonomous Community of Madrid (project e-Madrid S2013/ICE-2715). Skowron was supported by a Humboldt Research Fellowship for Postdoctoral Researchers.

References

Aziz, H., and Huang, S. 2016. Computational complexity of testing proportional justified representation. Technical Report arXiv:1612.06476, arXiv.org.

Aziz, H., and Huang, S. 2017. A polynomial-time algorithm to achieve extended justified representation. Technical Report 1703.10415, arXiv.org.

Aziz, H.; Gaspers, S.; Gudmundsson, J.; Mackenzie, S.; Mattei, N.; and Walsh, T. 2015. Computational aspects of multi-winner approval voting. In *AAMAS'15*, 107–115.

Aziz, H.; Brill, M.; Conitzer, V.; Elkind, E.; Freeman, R.; and Walsh, T. 2017. Justified representation in approval-based committee voting. *Social Choice and Welfare* 48(2):461–485.

Brams, S. J., and Fishburn, P. C. 2007. *Approval Voting*. Springer-Verlag, 2nd edition.

Brill, M.; Freeman, R.; Janson, S.; and Lackner, M. 2017. Phragmén's voting methods and justified representation. In *AAAI'17*, 406–413.

Garey, M. R., and Johnson, D. S. 1979. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman.

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

Kilgour, D. 2010. Approval balloting for multi-winner elections. In Laslier, J., and Sanver, R., eds., *Handbook on Approval Voting*. Springer. 105–124.

LeGrand, R.; Markakis, E.; and Mehta, A. 2007. Some results on approximating the minimax solution in approval voting. In *AAMAS'07*, 1193–1195.

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.

Sánchez-Fernández, L.; Elkind, E.; Lackner, M.; Fernández, N.; Fisteus, J. A.; Basanta Val, P.; and Skowron, P. 2017. Proportional justified representation. In *AAAI'17*, 670–676.

Sánchez-Fernández, L.; Elkind, E.; and Lackner, M. 2017. Committees providing EJR can be computed efficiently. Technical Report arXiv:1704.00356, arXiv.org.

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.; Lackner, M.; Elkind, E.; and Sánchez-Fernández, L. 2017. Optimal average satisfaction and extended justified representation in polynomial time. Technical Report arXiv:1704.00293, arXiv.org.

Skowron, P.; Faliszewski, P.; and Lang, J. 2016. Finding a collective set of items: From proportional multirepresentation to group recommendation. *Artificial Intelligence* 241:191–216.