

# The Price of Justified Representation

Edith Elkind,<sup>1</sup> Piotr Faliszewski,<sup>2</sup> Ayumi Igarashi,<sup>3</sup> Pasin Manurangsi,<sup>4</sup>  
Ulrike Schmidt-Kraepelin,<sup>5</sup> Warut Suksompong<sup>6</sup>

<sup>1</sup> University of Oxford, <sup>2</sup> AGH University of Science and Technology, <sup>3</sup> National Institute of Informatics

<sup>4</sup> Google Research <sup>5</sup> TU Berlin <sup>6</sup> National University of Singapore

## Abstract

In multiwinner approval voting, the goal is to select  $k$ -member committees based on voters' approval ballots. A well-studied concept of proportionality in this context is the *justified representation (JR)* axiom, which demands that no large cohesive group of voters remains unrepresented. However, the JR axiom may conflict with other desiderata, such as *coverage* (maximizing the number of voters who approve at least one committee member) or *social welfare* (maximizing the number of approvals obtained by committee members). In this work, we investigate the impact of imposing the JR axiom (as well as the more demanding EJR axiom) on social welfare and coverage. Our approach is threefold: we derive worst-case bounds on the loss of welfare/coverage that is caused by imposing JR, study the computational complexity of finding 'good' committees that provide JR (obtaining a hardness result, an approximation algorithm, and an exact algorithm for one-dimensional preferences), and examine this setting empirically on several synthetic datasets.

## 1 Introduction

What do the tasks of electing members of a parliament, making a shortlist of job candidates to invite for an interview, and selecting dishes for a banquet have in common? They can all be seen as instances of multiwinner voting: there is a set of voters with preferences over the candidates, and the goal is to select a fixed-size subset of the candidates based on these preferences. Recently, axiomatic and algorithmic questions related to multiwinner voting have attracted significant attention from the AI community (Faliszewski et al. 2017).

A common form of multiwinner voting involves *approval ballots*, wherein each voter specifies the subset of candidates that she finds acceptable. Depending on the application, one may want the winning set (also called a *committee*) to maximize the *utilitarian social welfare* (i.e., to select the candidates with the largest number of approvals) or *coverage* (i.e., to maximize the number of voters who approve at least one of the selected candidates). Another objective, which is somewhat more difficult to capture, is to select a committee that represents the set of voters in a proportional manner. A fairly basic proportionality requirement is the *justified representation (JR)* axiom, formulated by Aziz et al. (2017): it

says that in an  $n$ -voter election where the goal is to select a committee of size  $k$ , no group of  $n/k$  voters who jointly approve some candidate should remain unrepresented in the elected committee.

**Our Contribution** In this paper, we seek to understand the effect of imposing the JR axiom on the objectives of social welfare and coverage. We observe that requiring a committee to provide JR has no impact on coverage: there is a committee that offers optimal coverage as well as JR. In contrast, the impact on welfare is substantial: an example of Lackner and Skowron (2020b) shows that imposing the JR axiom may result in a committee whose social welfare differs from optimal by a factor of  $\sqrt{k}/2$ , and we prove that this bound is tight; in fact, we can obtain a nearly  $\sqrt{k}/2$ -approximation to optimal social welfare, while providing both JR and a  $4/3$ -approximation to optimal coverage. For coverage, we also show that the loss caused by imposing a more demanding axiom of *extended justified representation* is bounded by  $4/3$ .

We then investigate the complexity of computing a welfare-maximizing JR committee. While both the problem of computing a JR committee and the problem of maximizing the social welfare are in P, combining these two objectives results in a problem that is NP-hard, even to approximate up to a factor of  $k^{1/2-\epsilon}$  for  $\epsilon > 0$ . On the positive side, we show that it can be efficiently approximated up to a factor of nearly  $\sqrt{k}/2$ , which means that the inapproximability bound is asymptotically tight. Moreover, we present a polynomial-time exact algorithm for one-dimensional preferences (namely, for the fairly large 1D-VCR domain, recently proposed by Godziszewski et al. (2021)).

We complement our theoretical findings by an empirical analysis. We consider three standard models of preference distributions, and estimate the impact of imposing the JR axiom on welfare and coverage. In contrast to our theoretical results, this analysis paints a much more optimistic picture: one can achieve a good approximation to the Pareto frontier of social welfare and coverage even under the JR constraint.

**Related Work** There is a quickly growing body of work on approval-based multiwinner voting; we point the reader to the survey of Lackner and Skowron (2020a). Our work builds on several predecessor papers, which study trade-offs among different desiderata in the context of committee selection, but take a somewhat different perspective.

Lackner and Skowron (2020b) consider a number of approval-based multiwinner voting rules and, for each of them, establish theoretical worst-case guarantees on its performance with respect to social welfare and coverage. They also provide some experimental results, showing that in practice these rules often perform much better than their guarantees suggest. The main difference between their work and ours is that they consider specific voting rules (including some that provide JR), whereas we analyze the impact of the JR axiom itself. Furthermore, Lackner and Skowron (2020b) take the worst-case approach, i.e., they measure the performance of a rule by the *minimum* welfare/coverage provided by committees that it outputs, whereas we are interested in the *maximum* welfare/coverage that can be obtained by committees that provide JR. Nevertheless, some of their results and examples turn out to be relevant for our analysis.

Kocot et al. (2019) consider ordinal elections rather than the approval ones and analyze the complexity of finding a specific committee that achieves good scores according to several different rules. In particular, they consider the problem of maximizing the  $k$ -Borda score subject to achieving a given Chamberlin–Courant score (Chamberlin and Courant 1983). This is quite similar to our problem of finding committees that maximize social welfare subject to JR. Our plots illustrating the trade-off between the social welfare and the coverage that JR committees may provide are inspired by analogous plots of Kocot et al. (2019).

Bredereck et al. (2019) initiate the study of committees that provide JR. In particular, they show that maximizing welfare or coverage subject to JR is computationally hard; we strengthen their NP-hardness result for welfare to a  $k^{1/2-\varepsilon}$ -inapproximability result. They also conduct experiments in which they study trade-offs between coverage and social welfare, with and without imposing the JR axiom. However, for the purposes of their analysis, they only consider three specific elections, whereas we sample many elections from three different distributions; thus, our empirical results are much more robust.

## 2 Preliminaries

We consider elections with a finite set of candidates  $C$  of size  $m$  and a finite set of voters  $N = [n]$ , where we write  $[t] := \{1, \dots, t\}$  for any positive integer  $t$ . Each voter  $i \in N$  submits a non-empty ballot  $A_i \subseteq C$ , and the goal is to select a subset of  $C$  of size  $k$ , which we will refer to as the *winning committee*. Thus, an instance  $I$  of our problem can be described by a set of candidates  $C$ , a list of ballots  $\mathcal{A} = (A_1, \dots, A_n)$ , and a positive integer  $k$ ; we write  $I = (C, \mathcal{A}, k)$ . If  $c \in A_i$  for some  $c \in C$  and  $i \in N$ , we say that  $c$  *covers*  $i$ , or, equivalently, that  $i$  *approves*  $c$ .

We consider the following two measures of committee quality:

1. (Utilitarian) social welfare: For each committee  $W \subseteq C$ , we define the (*utilitarian*) *social welfare* of  $W$  as

$$\text{sw}(W) = \sum_{i \in N} |A_i \cap W|.$$

2. Coverage: For each committee  $W \subseteq C$ , we define the

*coverage* of  $W$  as

$$\text{cvr}(W) = |\{i \in N : A_i \cap W \neq \emptyset\}|.$$

Given an instance  $I = (C, \mathcal{A}, k)$  with  $\mathcal{A} = (A_1, \dots, A_n)$ , we say that a group of voters  $N' \subseteq N$  is *cohesive* if  $\bigcap_{i \in N'} A_i \neq \emptyset$ . Further, we say that a committee  $W$  *represents* a group of voters  $N' \subseteq N$  if  $W \cap A_i \neq \emptyset$  for some  $i \in N'$ . We are now ready to state the justified representation axiom of Aziz et al. (2017).

**Definition 2.1** (JR). Given an instance  $I = (C, \mathcal{A}, k)$  with  $\mathcal{A} = (A_1, \dots, A_n)$ , we say that a committee  $W \subseteq C$ ,  $|W| = k$ , provides *justified representation* (JR) for  $I$  if it represents every cohesive group of voters  $N' \subseteq N$  such that  $|N'| \geq n/k$ . Let  $\text{JR}(I)$  be the set of all committees that provide justified representation for  $I$ .

Fix an instance  $I = (C, \mathcal{A}, k)$ . We write

$$\begin{aligned} \text{sw}(I) &= \max_{\substack{W \subseteq C \\ |W|=k}} \text{sw}(W), & \text{sw}_{\text{JR}}(I) &= \max_{W \in \text{JR}(I)} \text{sw}(W), \\ \text{cvr}(I) &= \max_{\substack{W \subseteq C \\ |W|=k}} \text{cvr}(W), & \text{cvr}_{\text{JR}}(I) &= \max_{W \in \text{JR}(I)} \text{cvr}(W), \\ P_{\text{sw}}(I) &= \frac{\text{sw}(I)}{\text{sw}_{\text{JR}}(I)}, & P_{\text{cvr}}(I) &= \frac{\text{cvr}(I)}{\text{cvr}_{\text{JR}}(I)}. \end{aligned}$$

We refer to  $P_{\text{sw}}(I)$  and  $P_{\text{cvr}}(I)$  as the *social welfare price of JR on  $I$*  and the *coverage price of JR on  $I$* , respectively. Given a committee size  $k$ , the *social welfare price of JR* and the *coverage price of JR* are defined as the largest values  $P_{\text{sw}}$  and  $P_{\text{cvr}}$  can take on instances with committee size  $k$ :

$$P_{\text{sw}}(k) = \sup_{I=(C, \mathcal{A}, k)} P_{\text{sw}}(I), \quad P_{\text{cvr}}(k) = \sup_{I=(C, \mathcal{A}, k)} P_{\text{cvr}}(I).$$

We note that the sheer fact that a committee provides JR does not imply anything about its utilitarian social welfare or coverage. Indeed, there exist instances where the JR axiom is non-binding, in the sense that there are no cohesive groups of size  $n/k$ . For such instances even a committee that consists of candidates not approved by *any* voter provides JR. This is why we defined the social welfare price (the coverage price) of JR in terms of the maximum welfare (coverage) achievable by a JR committee and not the minimum one.

In our proofs, we will often refer to a family of algorithms known as GreedyCC. Each algorithm in this family takes an instance  $I = (C, \mathcal{A}, k)$  as input, outputs a committee  $W \subseteq C$  of size  $k$ , and proceeds in two stages. In the first stage, it adds candidates to  $W$  one by one, starting with  $W = \emptyset$ . At each iteration, for each candidate  $c \notin W$  it computes  $\text{cvr}(W \cup \{c\}) - \text{cvr}(W)$ ; let  $c$  be a candidate that maximizes this quantity. If  $\text{cvr}(W \cup \{c\}) - \text{cvr}(W) \geq n/k$  then the algorithm adds  $c$  to the committee and proceeds to the next iteration. The first stage ends when  $\text{cvr}(W \cup \{c\}) - \text{cvr}(W) < n/k$  for all  $c \notin W$ ; let  $W'$  be the committee obtained at this point, and let  $k' = |W'|$ . As each candidate added in the first stage covers  $n/k$  ‘fresh’ voters, we have  $k' \leq k$ , and we have  $W \in \text{JR}(I)$  for any size- $k$  committee  $W$  with  $W' \subseteq W$ . In the second stage,  $k - k'$  further candidates are added to the committee. Specifically, we denote by GreedyCC<sup>sw</sup> the variant of GreedyCC that adds  $k - k'$  candidates with the

highest number of approvals among candidates in  $C \setminus W'$  (breaking ties in favor of lower-indexed candidates).

All omitted proofs can be found in the full version of our paper (Elkind et al. 2021b).

### 3 Price of Justified Representation

In this section we provide nearly tight bounds on  $\mathcal{P}_{\text{sw}}(k)$  and  $\mathcal{P}_{\text{cvr}}(k)$ . Lackner and Skowron (2020b) give an example showing that  $\mathcal{P}_{\text{sw}}(k) \geq \sqrt{k}/2$ ; we provide an upper bound on  $\mathcal{P}_{\text{sw}}(k)$  that (nearly) matches this lower bound. On the other hand, we observe that  $\mathcal{P}_{\text{cvr}}(k) = 1$ . Then we show that social welfare and coverage are surprisingly compatible when considering JR committees. We also discuss how to extend our analysis to extended justified representation (EJR).

#### 3.1 Social Welfare

We first consider  $\mathcal{P}_{\text{sw}}(k)$ . In addition to the lower bound of  $\sqrt{k}/2$ , Lackner and Skowron (2020b) also show that the social welfare of the committees output by the *Proportional Approval Voting* rule (PAV), which is well-known to provide JR (Aziz et al. 2017), is within a factor of  $\sqrt{k} + 2$  from the optimal social welfare; this implies that  $\mathcal{P}_{\text{sw}}(k) \leq \sqrt{k} + 2$ . We improve this upper bound to one that, in essence, matches the bound from Lackner and Skowron's example.

**Theorem 3.1.** *For each positive number  $\beta < 2$ , there is a committee size  $k'$  such that for all  $k \geq k'$ , we have  $\mathcal{P}_{\text{sw}}(k) \leq \frac{1}{\beta}(1 + \sqrt{k})$ .*

*Proof.* Consider an  $n$ -voter instance  $I = (C, \mathcal{A}, k)$ . Let  $W$  be the committee computed by GreedyCC<sup>sw</sup> (so that  $W \in \text{JR}(I)$ ), and let  $S = \{s_1, \dots, s_k\}$  be a committee such that  $\text{sw}(S) = \text{sw}(I)$ . For each  $i \in [k]$ , let  $a_i$  be the number of approvals received by candidate  $s_i$ . We can assume without loss of generality that  $a_1 \geq a_2 \geq \dots \geq a_k$ . For each  $i \in [k]$ , let  $S^i = \{s_1, \dots, s_i\}$  be the set of  $i$  candidates with the highest approval scores; note that  $\text{sw}(S^i) \geq \frac{i}{k} \cdot \text{sw}(S)$ . By definition, some candidate with  $a_1$  approvals is selected in the first iteration of GreedyCC<sup>sw</sup>; we can assume without loss of generality that this candidate is  $s_1$ .

Suppose first that GreedyCC<sup>sw</sup> selects at most  $k - \lfloor 2\sqrt{k} \rfloor$  candidates during the first stage. Consider a candidate  $s_i$ ,  $i \leq \lfloor 2\sqrt{k} \rfloor$ . If  $s_i$  is not selected during the first stage, then  $s_i$  will necessarily be selected during the second stage.<sup>1</sup> Hence, we have:

$$\text{sw}(W) \geq (\lceil 2\sqrt{k} \rceil / k) \cdot \text{sw}(S) \geq (2/\sqrt{k}) \cdot \text{sw}(S).$$

So, in this case  $\mathcal{P}_{\text{sw}}(I) \leq \sqrt{k}/2 \leq \frac{1+\sqrt{k}}{\beta}$  for each  $\beta < 2$ . Thus let us assume that GreedyCC<sup>sw</sup> selects  $k - x\sqrt{k}$  first-stage candidates for some  $0 \leq x < 2$ . Let  $r = x\sqrt{k}$ ; note that  $r$  is an integer. We will assume that  $k \geq 2r + 1$ ; this is true for large enough values of  $k$ .

Recall that in the first iteration, GreedyCC<sup>sw</sup> chooses candidate  $s_1$ , who is approved by  $a_1$  voters. Afterwards, it selects  $k - r - 1$  further first-stage candidates, with each

<sup>1</sup>We bound  $i$  by  $\lfloor 2\sqrt{k} \rfloor$  and not  $\lfloor 2\sqrt{k} \rfloor$  because we know that  $s_1$  was selected in the first stage.

of these candidates covering at least  $n/k$  additional voters. Hence, it must be the case that:

$$n - a_1 \geq \frac{n}{k} (k - r - 1).$$

Recalling that  $r = x\sqrt{k}$ , we obtain  $a_1 \leq \frac{n}{k} + \frac{xn}{\sqrt{k}}$  and hence  $a_i \leq \frac{n}{k} + \frac{xn}{\sqrt{k}}$  for  $i \in [r]$ . On the other hand, since each of the  $k - r - 1 \geq r$  candidates selected in the first stage besides  $s_1$  covers at least  $\frac{n}{k}$  additional voters, for each  $i \in [r]$  we have  $a_i \geq \frac{n}{k}$ . Hence, there exists a real value  $\alpha \in [0, 1]$  such that the average number of approvals obtained by candidates in  $S^r$  is:

$$\frac{1}{r} \sum_{i=1}^r a_i = \frac{n}{k} + \frac{\alpha xn}{\sqrt{k}}.$$

Thus the social welfare of  $S$  can be upper-bounded as:

$$\text{sw}(S) \leq k \cdot \left( \frac{n}{k} + \frac{\alpha xn}{\sqrt{k}} \right) = n(1 + \alpha x \sqrt{k}).$$

We will now establish a lower bound on the social welfare of  $W$ . Since we add  $r$  candidates in the second stage,  $W$  necessarily contains all candidates in  $S^r$ , who contribute a total of at least  $r(\frac{n}{k} + \frac{\alpha xn}{\sqrt{k}})$  to the social welfare. Moreover, in the first stage, we add at least  $k - r - |S^r| = k - 2r$  candidates who are not contained in  $S^r$ . Each of the first-stage candidates covers at least  $n/k$  additional voters and therefore contributes at least  $n/k$  to the social welfare. Hence, we can lower-bound  $\text{sw}(W)$  as:

$$\text{sw}(W) \geq \left( \frac{n}{k} + \frac{\alpha xn}{\sqrt{k}} \right) \cdot r + (k - 2r) \cdot \frac{n}{k},$$

or, rewriting,

$$\text{sw}(W) \geq n - \frac{xn}{\sqrt{k}} + \alpha x^2 n = n \left( 1 - \frac{x}{\sqrt{k}} + \alpha x^2 \right).$$

Now,  $\mathcal{P}_{\text{sw}}(I)$  can be upper-bounded as:

$$\frac{\text{sw}(S)}{\text{sw}(W)} \leq \frac{1 + \alpha x \sqrt{k}}{1 - \frac{x}{\sqrt{k}} + \alpha x^2} = \frac{\sqrt{k} + \alpha x k}{\sqrt{k} - x + \alpha x^2 \sqrt{k}}.$$

We claim that for each  $\beta \in (0, 2)$  and sufficiently large  $k$  it holds that:

$$\frac{\sqrt{k} + \alpha x k}{\sqrt{k} - x + \alpha x^2 \sqrt{k}} \leq \frac{1 + \sqrt{k}}{\beta}.$$

Note first that this inequality can be rewritten as:

$$\sqrt{k} - x + \alpha x^2 \sqrt{k} + k - x\sqrt{k} + \alpha x^2 k \geq \beta \sqrt{k} + \alpha \beta x k,$$

or, regrouping the terms and dividing by  $\sqrt{k}$ ,

$$-1 - \alpha x^2 + x + \beta + \frac{x}{\sqrt{k}} \leq \sqrt{k}(\alpha x^2 - \alpha \beta x + 1). \quad (1)$$

Now, observe that

$$\begin{aligned} \alpha x^2 - \alpha \beta x + 1 &= \left( \alpha x^2 - \alpha \beta x + \frac{\alpha \beta^2}{4} \right) + \left( 1 - \frac{\alpha \beta^2}{4} \right) \\ &= \alpha \left( x - \frac{\beta}{2} \right)^2 + \left( 1 - \frac{\alpha \beta^2}{4} \right) \\ &\geq 1 - \frac{\alpha \beta^2}{4} \geq 1 - \frac{\beta^2}{4} > 0. \end{aligned}$$

As a consequence, inequality (1) holds for sufficiently large  $k$  since the left-hand side does not increase as  $k$  grows, whereas the right-hand side grows proportionally to  $\sqrt{k}$ .  $\square$

For an explicit bound on  $k$  in the statement of Theorem 3.1, we refer to the full version of our paper (Elkind et al. 2021b). One may wonder if it is possible to strengthen the theorem so that it matches the  $\sqrt{k}/2$  lower bound exactly, for all values of  $k$ . While it might be possible, it would likely require an approach that is different from ours: the approximation of the social welfare provided by GreedyCC<sup>sw</sup> is quite good, but insufficient for small committees.

### 3.2 Coverage

Next we consider the coverage price of justified representation. Since the committees selected by the Chamberlin–Courant rule are known to both provide JR and maximize the coverage (Aziz et al. 2017), we have the following corollary.

**Corollary 3.2.** For every instance  $I = (C, \mathcal{A}, k)$ , there exists a committee  $W \in \text{JR}(I)$  such that  $\text{cvr}(W) \geq \text{cvr}(S)$  for all  $S \subseteq C$  with  $|S| = k$ . Thus,  $\mathcal{P}_{\text{cvr}}(k) = 1$  for all  $k \in \mathbb{N}$ .

### 3.3 Combining Social Welfare and Coverage

Social welfare and coverage are often viewed as two desiderata that are at the opposite ends of the spectrum: maximizing the social welfare may lead to low coverage and vice versa. We show that for JR committees these two properties are surprisingly compatible: if the committee size is sufficiently large, there always exists a JR committee that provides nearly optimal social welfare as well as coverage that is nearly  $3/4$  of the maximum possible coverage.

**Theorem 3.3.** For every  $\beta \in (0, 2)$  and  $\varepsilon > 0$ , there exists a  $k_0 \in \mathbb{N}$  such that for every instance  $I = (C, \mathcal{A}, k)$  with  $k \geq k_0$ , there exists a committee  $W \in \text{JR}(I)$  with  $\text{sw}(I)/\text{sw}(W) \leq \frac{1}{\beta}(1 + \sqrt{k})$  and  $\text{cvr}(I)/\text{cvr}(W) \leq 4/3 + \varepsilon$ .

The proof of Theorem 3.3 relies on a variant of GreedyCC that runs in exponential time. This is inevitable since, unless  $\text{P} = \text{NP}$ , the best approximation ratio for coverage that is guaranteed in polynomial time is  $1 - 1/e$  (see, e.g., Skowron and Faliszewski 2017), and  $1/(1 - 1/e) > 4/3$ . It is thus desirable to have a variant of Theorem 3.3 for committees computable in polynomial time. It turns out that in this case we can achieve a near-perfect result.

**Theorem 3.4.** Let  $\beta < 2$  be a positive number. For every instance  $I = (C, \mathcal{A}, k)$  with sufficiently large  $k$ , there is a committee  $W \in \text{JR}(I)$  with  $\text{sw}(I)/\text{sw}(W) \leq \frac{1}{\beta}(1 + \sqrt{k})$  and  $\text{cvr}(I)/\text{cvr}(W) \leq 1/(1 - (1/e)^{(1-2/\sqrt{k}-1/k)})$ .

### 3.4 Extended Justified Representation

We also study a strengthening of the JR axiom, known as *extended justified representation (EJR)* (Aziz et al. 2017), which is typically interpreted as a proportionality axiom. Let us consider an election with  $n$  voters, where we seek a committee of size  $k$ . We say that a group of voters is  $\ell$ -cohesive if there are at least  $\ell$  candidates that are approved by all the voters in this group, and we say that this group is  $\ell$ -large if it contains at least  $\ell n/k$  members.

**Definition 3.5 (EJR).** Given an instance  $I = (C, \mathcal{A}, k)$ , we say that a committee  $W \subseteq C$ ,  $|W| = k$ , provides *extended justified representation (EJR)* for  $I$  if for each  $\ell \in [k]$  and

each  $\ell$ -cohesive,  $\ell$ -large group of voters, there is at least one voter in this group who approves at least  $\ell$  members of  $W$ . Let  $\text{EJR}(I)$  denote the set of all committees that provide extended justified representation for  $I$ .

We define the social welfare price of EJR and the coverage price of EJR analogously to these notions for JR. The results of Lackner and Skowron (2020b) imply that the social welfare price of EJR is between  $\sqrt{k}/2$  and  $2 + \sqrt{k}$ , as well as that the coverage price of EJR is at least  $4/3$ . We show that, in fact, this latter bound is tight.

**Theorem 3.6.** The coverage price of EJR is  $4/3$ .

The proof is very similar to that of Theorem 3.3. The main difference is that we replace GreedyCC with an exponential-time greedy algorithm that is guaranteed to output a committee that provides EJR, and we do not have a stage where we select candidates so as to maximize the social welfare.

## 4 Optimizing Social Welfare under JR

In this section we focus on the complexity of maximizing the social welfare over the set of all committees that provide JR. We do not discuss maximizing coverage, as this topic is already covered in the rich literature on the approval-based Chamberlin–Courant rule; see, e.g., the works of Procaccia, Rosenschein, and Zohar (2008), Lu and Boutilier (2011), Betzler, Slinko, and Uhlmann (2013), Skowron et al. (2015), and Godziszewski et al. (2021).

### 4.1 Hardness of Approximation

Interestingly, while there are polynomial-time algorithms for finding committees that provide JR (Aziz et al. 2017), and we can find a social welfare-maximizing committee by simply picking  $k$  candidates with the highest number of approvals, finding a social welfare-maximizing JR committee turns out to be NP-hard. This was first observed by Bredereck et al. (2019); we will now strengthen their hardness result to an inapproximability result. We show that, for any constant  $\varepsilon > 0$ , it is NP-hard to approximate  $\text{sw}_{\text{JR}}(I)$  to within a factor of  $k^{1/2-\varepsilon}$ . This hardness result holds even when  $n = 2k$ .

**Theorem 4.1.** For any  $\varepsilon \in (0, 1/2)$ , the following problem is NP-hard:<sup>2</sup> Given an instance  $I = (C, \mathcal{A}, k)$ , find a committee  $W \in \text{JR}(I)$  such that  $\text{sw}(W) \geq \text{sw}_{\text{JR}}(I)/k^{1/2-\varepsilon}$ . This problem remains NP-hard even if  $k = n/2$ .

Since GreedyCC<sup>sw</sup> runs in polynomial time and the optimal social welfare under JR is at most the optimal social welfare overall, Theorem 3.1 implies that the inapproximability factor in Theorem 4.1 is asymptotically tight.

### 4.2 The case $n = k$

Our  $k^{1/2-\varepsilon}$ -inapproximability result holds for the case where  $n = 2k$ , and it is easy to extend it to some cases with smaller  $k$ . For example, for  $n = 4k$  it suffices to clone each voter. Yet, if  $k$  is extremely small, e.g., if it is bounded

<sup>2</sup>When saying that this search problem is NP-hard, we mean that a polynomial-time algorithm for the problem would enable us to solve all problems in NP in polynomial time.

by a constant, then, of course, we can find a social welfare-maximizing committee in  $\text{JR}(I)$  in polynomial time.

Larger values of  $k$  can also become easier. In what follows, we will focus on the case  $n = k$ . In this case, the JR axiom is equivalent to requiring perfect coverage: every voter forms a group of size  $n/k$  and thus must be represented in the committee (recall that we require that  $A_i \neq \emptyset$  for all  $i \in [n]$ ). It turns out that maximizing the social welfare under this condition remains hard.

**Theorem 4.2.** *For some constant  $\varepsilon > 0$ , the following problem is NP-hard: Given an instance  $I = (C, \mathcal{A}, k)$  with  $n = k$ , find  $W \in \text{JR}(I)$  with  $\text{sw}(W) \geq \text{sw}_{\text{JR}}(I)/(1 + \varepsilon)$ .*

In spite of the hardness, the case  $n = k$  admits a 2-approximation algorithm, which means that it is easier than the case  $n = 2k$  (Theorem 4.1).

**Theorem 4.3.** *There is a polynomial-time algorithm that, given an instance  $I = (C, \mathcal{A}, k)$  where  $n = k$ , finds a committee  $W \in \text{JR}(I)$  such that  $\text{sw}(W) \geq \text{sw}_{\text{JR}}(I)/2$ .*

### 4.3 Structured Preferences

Another approach to circumvent intractability is to focus on instances where voters have structured preferences. In this section, we consider one such domain, namely, the 1D-VCR domain, recently introduced by Godziszewski et al. (2021). The name of the domain, 1D-VCR, stands for 1-dimensional voter/candidate range model.

**Definition 4.4.** Given an instance  $I = (C, \mathcal{A}, k)$ , with voter set  $N$ , we say that the voters have *1D-VCR preferences* if there exists a collection  $(x_a, r_a)_{a \in C \cup N}$  of points  $x_a \in \mathbb{R}$  and nonnegative real values  $r_a \in \mathbb{R}$  such that for each voter  $i \in N$  and candidate  $c \in C$ , it holds that  $i$  approves  $c$  if and only if  $|x_c - x_i| \leq r_c + r_i$ .

Given an agent  $a \in C \cup N$ , we refer to  $x_a$  and  $r_a$  as the *position* and *radius* of  $a$ , respectively. We assume that the positions and radii of the agents are specified in the input instance. This assumption carries no computational cost, since we can decide in polynomial time whether a given instance  $I$  is a 1D-VCR instance and compute the respective positions and radii; this follows from the work of Müller (1997) and Rafey (2012).

The 1D-VCR domain is a generalization of both the candidate interval (CI) and voter interval (VI) domains of Elkind and Lackner (2015) (which translate the classic definitions of single-peaked (Black 1958) and single-crossing elections (Mirrlees 1971; Roberts 1977) to the approval setting): the CI domain (resp., the VI domain) corresponds to all candidates (resp., voters) having zero radii.

Given a 1D-VCR instance  $I = (C, \mathcal{A}, k)$  and an agent  $a \in C \cup N$ , we denote by  $s(a) = x_a - r_a$  and  $t(a) = x_a + r_a$  the leftmost and the rightmost point of  $a$ 's range of influence, respectively. We call  $[s(a), t(a)]$  the *interval* of  $a$ . Note that voter  $i$  approves candidate  $c$  if and only if  $i$ 's interval  $[s(i), t(i)]$  and  $c$ 's interval  $[s(c), t(c)]$  have a non-empty intersection, i.e.,

$$s(c) \leq t(i) \text{ and } s(i) \leq t(c). \quad (2)$$

The main result of this section is that the problem of maximizing the social welfare under the JR constraint is tractable for 1D-VCR instances.

**Theorem 4.5.** *There is a polynomial-time algorithm that, given a 1D-VCR instance  $I = (C, \mathcal{A}, k)$ , computes a committee that maximizes the social welfare under JR.*

We first prove Theorem 4.5 for the case where the intervals of the candidates have a non-nested structure (Lemma 4.6). Below, we denote by  $\text{JR}_{\leq k}(I)$  the set of committees of size at most  $k$  that satisfy JR (with parameter  $k$ ) for an instance  $I = (C, \mathcal{A}, k)$ . Note that for this lemma, we do not make the usual assumption that  $|C| \geq k$ .

**Lemma 4.6.** *Let  $I = (C, \mathcal{A}, k)$  be a 1D-VCR instance such that for every pair of candidates  $c, c' \in C$  with  $t(c) \leq t(c')$  it holds that  $s(c) \leq s(c')$ . Then, there is a polynomial-time algorithm that, given  $k^* \in [\min\{k, |C|\}]$ , computes a committee  $W \subseteq C$  that maximizes the social welfare under the constraints that  $W \in \text{JR}_{\leq k}(I)$  and  $|W| = k^*$  (or reports that no such committee exists).*

Our general algorithm consists of two phases: First, we compute a committee (possibly of size smaller than  $k$ ) that provides JR for the original instance and only contains candidates whose intervals are maximal. Second, we supplement this committee with highly-approved candidates.

To formalize this idea, we introduce the notion of *representatives*. Given a 1D-VCR instance  $I = (C, \mathcal{A}, k)$ , we say that candidate  $c^*$  is a *representative* of candidate  $c$  if the interval of  $c^*$  strictly includes that of  $c$  and is a maximal interval with this property. We denote the set of all representatives by  $C^*$ ; we have  $C^* = \{c \in C \mid \nexists c' \in C : [s(c), t(c)] \subsetneq [s(c'), t(c')]\}$ . We denote the instance restricted to the representatives by  $I^* = (C^*, \mathcal{A}|_{C^*}, k)$ . By Lemma A.1 of Elkind et al. (2021a), if voter  $i$  approves  $c$ , then  $i$  also approves its representative  $c^*$ . The following lemma states that a committee that consists of representatives satisfies JR if and only if it satisfies JR for the restricted instance.

**Lemma 4.7.** *Let  $I = (C, \mathcal{A}, k)$  be a 1D-VCR instance, and let  $I^* = (C^*, \mathcal{A}|_{C^*}, k)$  be the instance restricted to the representatives. Then for each  $W \subseteq C^*$  it holds that  $W \in \text{JR}_{\leq k}(I^*)$  if and only if  $W \in \text{JR}_{\leq k}(I)$ .*

Lemma 4.7 allows us to decompose our instance into its restrictions to  $C^*$  and  $C \setminus C^*$ .

**Lemma 4.8.** *Let  $I = (C, \mathcal{A}, k)$  be a 1D-VCR instance, and let  $I^* = (C^*, \mathcal{A}|_{C^*}, k)$  be the instance restricted to the representatives. Then, we have  $\text{sw}_{\text{JR}}(I) = \alpha$ , where*

$$\alpha = \max\{\text{sw}(W^*) + \text{sw}(W') \mid W^* \in \text{JR}_{\leq k}(I^*), \\ |W^*| + |W'| = k, W' \subseteq C \setminus C^*\}.$$

We can now use Lemma 4.8 in order to prove Theorem 4.5 by iterating over the size of  $W^*$ .

*Proof of Theorem 4.5.* Let  $I = (C, \mathcal{A}, k)$  be a 1D-VCR instance and let  $I^* = (C^*, \mathcal{A}|_{C^*}, k)$  be the instance restricted to the representatives. By Lemma 4.8, we can proceed as follows. We try all values of  $k^*$  in  $\{0, 1, \dots, \min\{|C^*|, k\}\}$ .

For each value of  $k^*$ , since the intervals of the candidates in  $C^*$  are not properly contained in each other, we can use Lemma 4.6 to compute a committee  $W^*$  that maximizes the social welfare subject to the requirements that  $W^* \in \text{JR}_{\leq k}(I^*)$  and  $|W^*| = k^*$  (or determine that no such committee exists). We can then find a committee  $W' \subseteq C \setminus C^*$  of size  $k - k^*$  maximizing the social welfare among such committees, by choosing  $k - k^*$  candidates with the highest number of approvals. For each value of  $k^*$  for which  $W^*$  exists, we evaluate  $\text{sw}(W^* \cup W')$ , and return the best of these committees.  $\square$

## 5 Experimental Evaluation

We analyze the trade-off between maximizing social welfare and maximizing coverage while requiring JR in randomly generated instances. We illustrate this trade-off in the form of Pareto curves, as depicted in Figure 1. In addition, we explore how well a variant of GreedyCC performs in selecting committees that are close to these Pareto curves.

**Setup** We consider three different models for generating random elections. All take as input the number of voters  $n$ , the number of candidates  $m$ , and one additional parameter.

- The *impartial culture (IC)* model is parameterized by  $p \in [0, 1]$ . Each candidate is approved by every voter independently with probability  $p$ .
- The  *$n$ -dimensional Euclidean* model takes as input a parameter  $r \in [0, \sqrt{n}]$ , also referred to as the *radius*. Every voter  $i$  and every candidate  $c$  is associated with points  $x_i, x_c \in [0, 1]^n$ , respectively, which are drawn uniformly at random. Candidate  $c$  is approved by voter  $i$  iff  $d(x_i, x_c) \leq r$ , where  $d(\cdot, \cdot)$  is the Euclidean distance. We utilize the 1-dimensional (1D) and 2-dimensional (2D) Euclidean model in our experiments.

All three models are frequently used in the literature (see, e.g., Bredereck et al. 2019; Godziszewski et al. 2021). We focus on elections with parameters  $n = m = 100$  and  $k = 10$ . To make the results for the three models comparable, we chose the parameters  $p$  and  $r$  so that on average each voter approves  $n/k = 10$  candidates. Bredereck et al. (2019) observe that, for the above models, JR is especially demanding for this ratio. This leads to parameters  $p = 0.1$  (for IC),  $r = 0.054$  (for 1D) and  $r = 0.195$  (for 2D).

Part of our implementation builds upon code contributed by Andrzej Kaczmarczyk (Bredereck et al. 2019). Our code can be found at <https://github.com/Project-PRAGMA/PriceOfJR-AAAI-2022>.

### Trade-off between social welfare and coverage under JR

The Pareto curves depicted in the top row of Figure 1 are created as follows. For each of the three models we sample 100 instances. For a given instance  $I$ , we iterate over  $\alpha \in [0, 1]$  (in 0.01 increments) and compute a size- $k$  committee  $W_{I,\alpha}$  so as to maximize the social welfare subject to the constraint that at least  $\alpha \cdot \text{cvr}(I)$  coverage is achieved and JR is satisfied. To compute  $W_{I,\alpha}$ , we use an IP formulation. Then, for each  $\alpha$  we define  $f(\alpha)$  as the average of  $\text{sw}(W_{I,\alpha})/\text{sw}(I)$  over all generated instances of the model and illustrate  $f$  by a blue

line in Figures 1a to 1c. For comparison, we repeat this process without requiring committees to provide JR, and depict the resulting curve by a red dashed line.

The bottom row of Figure 1 is created by exchanging the roles of the two objectives. That is, for  $\alpha \in [0, 1]$  (with 0.005 increments) the committee  $W_{I,\alpha}$  is chosen so as to maximize coverage subject to the constraints that the social welfare is at least  $\alpha \cdot \text{sw}(I)$  and JR is satisfied. There is one important difference to the previous experiments, namely, that for large values of  $\alpha$ , such a committee need not exist since high social welfare and JR can be incompatible. In these cases, we set  $W_{I,\alpha} = \emptyset$ . Then, we compute  $g(\alpha)$  as the average of  $\text{cvr}(W_{I,\alpha})/\text{cvr}(I)$  over 100 instances, and illustrate it by a blue line in the bottom row of Figure 1. Again, for comparison we drop the JR constraint and repeat the computation; the result is depicted by a red dashed line. Observe that the two figures in each column are not symmetric.

**Greedy Heuristic** In order to find committees that are close to the Pareto curves, is it necessary to solve an IP? To answer this question, we implement a greedy heuristic and analyze the distance from its selected committees to the Pareto curves. This heuristic proceeds as follows.

First, the algorithm computes  $\text{sw}(I)$  and a greedy estimate of  $\text{cvr}(I)$ , which is done by computing a committee  $\bar{W}$  returned by GreedyCC. Then, the algorithm starts with an empty committee  $W = \emptyset$ , and in each iteration decides between one of two steps: In a *coverage step*, the algorithm selects the next candidate  $c$  so as to maximize  $\text{cvr}(W \cup \{c\})$ ; among all such candidates, it chooses one with maximum approval score. In a *social welfare step*, the algorithm selects a candidate  $c$  with maximum approval score; among all such candidates  $c$  it chooses one that maximizes  $\text{cvr}(W \cup \{c\})$ . As long as JR is not satisfied, the greedy algorithm performs coverage steps. After that, it compares  $\text{cvr}(W)/\text{cvr}(\bar{W})$  with  $\text{sw}(W)/\text{sw}(I)$ . If the former is smaller, it performs a coverage step; otherwise it performs a social welfare step.

After termination, we compute the exact approximation ratio of  $W$ , i.e., the point  $(\text{cvr}(W)/\text{cvr}(I), \text{sw}(W)/\text{sw}(I))$  and indicate it by a blue triangle or a red dot in Figures 1a to 1c. More precisely, we depict the point by a red dot if it is not on the Pareto curve of the instance  $I$ . That is, we check whether there exists a committee  $W'$  achieving coverage at least  $\text{cvr}(W)$  and social welfare strictly greater than  $\text{sw}(W)$ . In this case we also note the distance  $(\text{sw}(W') - \text{sw}(W))/\text{sw}(I)$  for the optimal  $W'$ . Otherwise, the point lies on the Pareto curve and is indicated by a blue triangle. For the bottom row of Figure 1 we perform the analogous procedure.

### Maximizing social welfare with respect to $\alpha$ -coverage

Consider the top row of Figure 1. For the impartial culture model with approval probability  $p = 0.1$  (Figure 1a), the restriction to JR committees has no effect as the two Pareto curves coincide. Also, the trade-off between social welfare and coverage does not appear to be very strong. Even when demanding optimal coverage, we can obtain (on average) 88% of optimal social welfare. Regarding the performance of the greedy algorithm, 52% of the committees are positioned on the Pareto curve of their instance. Among the re-

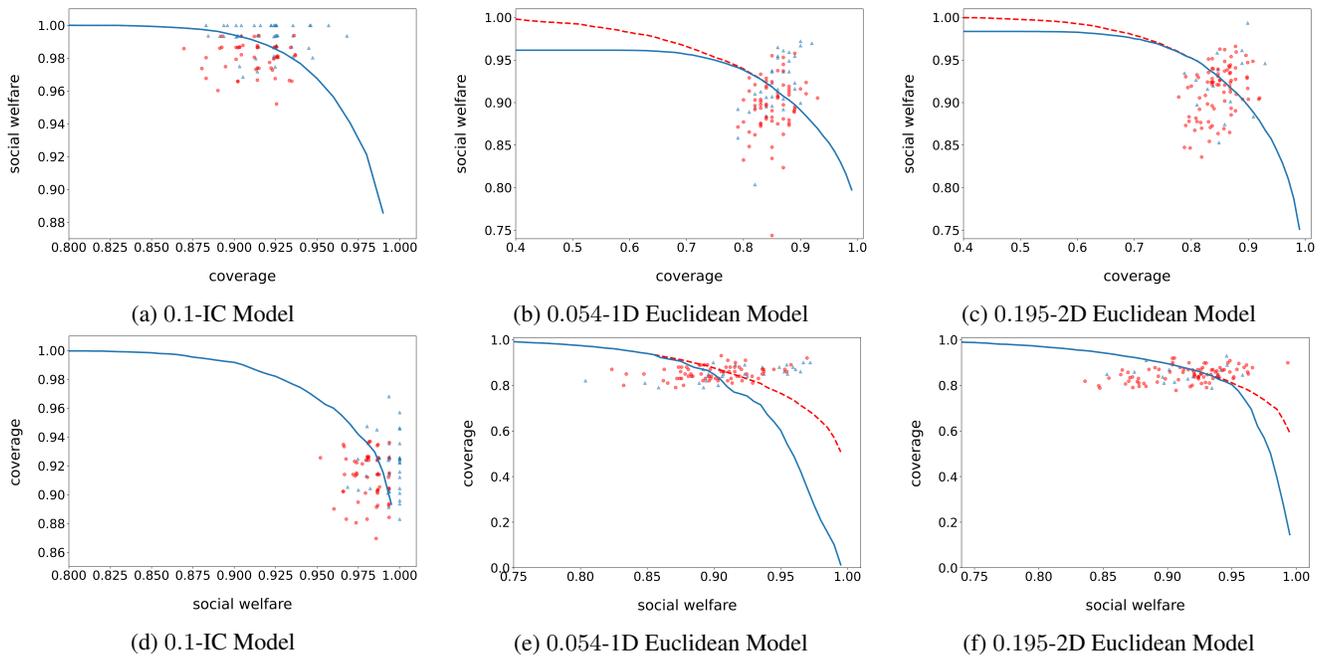


Figure 1: In the top row (resp., bottom row) we indicate the function  $f$  (resp.,  $g$ ) by a blue solid line and the benchmark line (where we drop the JR requirement) by a red dashed line. Dots and triangles indicate the committees computed by the greedy algorithm: a committee is indicated by a blue triangle if it lies on the Pareto curve, and by a red dot otherwise.

maining points, the average distance towards their respective Pareto curve (along the  $sw$ -axis) is 0.012.<sup>3</sup>

The trade-off between the two objectives becomes more apparent for the Euclidean models (Figures 1b and 1c). Even without any constraints regarding the coverage of the committee, the constraint that the committee has to provide JR induces a social welfare loss of roughly 4% (for the 1D model) and 1.5% (for the 2D model). Although the absolute loss, as compared to the best achievable social welfare, becomes larger for higher values of  $\alpha$ , namely 21.5% for the 1D model and 25% for the 2D model, the distance from the benchmark line diminishes. Regarding the performance of the greedy heuristic, 33% (for 1D) and 15% (for 2D) of the committees lie on the Pareto curve of their instance. The average distance of the remaining points from their respective Pareto curve is 0.022 (for 1D) and 0.024 (for 2D).

#### Maximizing coverage with respect to $\alpha$ -social welfare

Figure 1d shows that, again, for the IC model, the JR constraint has no impact on maximizing coverage under the constraint that an  $\alpha$ -fraction of social welfare is achieved. Regarding the performance of the greedy heuristic, 45% of the committees are positioned on the Pareto curve of their instance. For the remaining points, the distance towards the Pareto curve (along the  $cvr$ -axis) is 0.026 on average. For the 1D model (Figure 1e), requiring JR has no impact on the maximum coverage achievable up to a social welfare fraction of 85%. Starting from there, more and more instances do not admit a committee that provides JR while attaining

<sup>3</sup>We observe a grid-like pattern of the greedy points along both axes; see the full version for an explanation of this effect.

sufficient social welfare. For the 2D model (Figure 1f) the situation is similar, though JR becomes restrictive only for  $\alpha \geq 0.94$ . Considering the greedy heuristic, 29% (for 1D) and 16% (for 2D) of the committees are positioned on the Pareto curves of their instance. For the remaining points, the average distance from their Pareto curve (along the  $cvr$ -axis) is 0.036 and 0.038 for the 1D and 2D model, respectively.

**Conclusions of experiments** In contrast to our theoretical results, the experiments show that in practice, providing JR is often compatible with high social welfare. Moreover, for all three models, if we require at least 80% coverage, demanding JR on top does not lead to additional social welfare loss. Our greedy heuristic performs well in finding committees that are close to the Pareto curve: For all models, more than 20% of the committees are optimal and the remaining ones are close to the Pareto curve on average.

The full version of our paper (Elkind et al. 2021b) contains a comparison to experiments from the literature and details on the computational infrastructure.

## 6 Conclusions and Future Work

We have provided an extensive analysis of the impact of the justified representation (JR) axiom on both the (utilitarian) social welfare and the coverage, both from a worst-case perspective and from an algorithmic perspective, and complemented our theoretical results with an empirical analysis.

It would be interesting to perform a similar analysis for the EJR axiom, as well as other proportionality axioms such as PJR (Fernández et al. 2017) and the core (Aziz et al. 2017); Theorem 3.6 is the first step in this direction.

## Acknowledgments

This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No 101002854), from the Deutsche Forschungsgemeinschaft under grant BR 4744/2-1, from JST PRESTO under grant number JPMJPR20C1, and from an NUS Start-up Grant. We would like to thank the anonymous reviewers for their valuable comments.



## References

- 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.
- Betzler, N.; Slinko, A.; and Uhlmann, J. 2013. On the computation of fully proportional representation. *Journal of Artificial Intelligence Research*, 47: 475–519.
- Black, D. 1958. *The Theory of Committees and Elections*. Cambridge University Press.
- Bredereck, R.; Faliszewski, P.; Kaczmarczyk, A.; and Niedermeier, R. 2019. An experimental view on committees providing justified representation. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI)*, 109–115.
- Chamberlin, J. R.; and Courant, P. N. 1983. Representative deliberations and representative decisions: Proportional representation and the Borda rule. *American Political Science Review*, 77(3): 718–733.
- Elkind, E.; Faliszewski, P.; Igarashi, A.; Manurangsi, P.; Schmidt-Kraepelin, U.; and Suksompong, W. 2021a. Justifying groups in multiwinner approval voting. *arXiv preprint arXiv:2108.12949*.
- Elkind, E.; Faliszewski, P.; Igarashi, A.; Manurangsi, P.; Schmidt-Kraepelin, U.; and Suksompong, W. 2021b. The price of justified representation. *arXiv preprint arXiv:2112.05994*.
- Elkind, E.; and Lackner, M. 2015. Structure in dichotomous preferences. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI)*, 2019–2025.
- Faliszewski, P.; Skowron, P.; Slinko, A.; and Talmon, N. 2017. Multiwinner voting: a new challenge for social choice theory. In Endriss, U., ed., *Trends in Computational Social Choice*, chapter 2, 27–47. AI Access.
- Fernández, L. S.; Elkind, E.; Lackner, M.; García, N. F.; Arias-Fisteus, J.; Basanta-Val, P.; and Skowron, P. 2017. Proportional justified representation. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI)*, 670–676.
- Godziszewski, M.; Batko, P.; Skowron, P.; and Faliszewski, P. 2021. An analysis of approval-based committee rules for 2D-Euclidean elections. In *Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI)*, 5448–5455.
- Kocot, M.; Kolonko, A.; Elkind, E.; Faliszewski, P.; and Talmon, N. 2019. Multigoal committee selection. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI)*, 385–391.
- Lackner, M.; and Skowron, P. 2020a. Approval-based committee voting: Axioms, algorithms, and applications. *arXiv preprint arXiv:2007.01795*.
- Lackner, M.; and Skowron, P. 2020b. Utilitarian welfare and representation guarantees of approval-based multiwinner rules. *Artificial Intelligence*, 288: 103366.
- 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.
- Mirrlees, J. 1971. An exploration in the theory of optimal income taxation. *Review of Economic Studies*, 38(2): 175–208.
- Müller, H. 1997. Recognizing interval digraphs and interval bigraphs in polynomial time. *Discrete Applied Mathematics*, 78(1–3): 189–205.
- Procaccia, A.; Rosenschein, J.; and Zohar, A. 2008. On the complexity of achieving proportional representation. *Social Choice and Welfare*, 30(3): 353–362.
- Rafiey, A. 2012. Recognizing interval bigraphs by forbidden patterns. *arXiv preprint arXiv:1211.2662*.
- Roberts, K. W. S. 1977. Voting over income tax schedules. *Journal of Public Economics*, 8(3): 329–340.
- Skowron, P.; and Faliszewski, P. 2017. Chamberlin–Courant rule with approval ballots: Approximating the MaxCover problem with bounded frequencies in FPT time. *Journal of Artificial Intelligence Research*, 60: 687–716.
- Skowron, P.; Yu, L.; Faliszewski, P.; and Elkind, E. 2015. The complexity of fully proportional representation for single-crossing electorates. *Theoretical Computer Science*, 569: 43–57.