The Complexity of Proportionality Degree in Committee Elections

Łukasz Janeczko, Piotr Faliszewski

AGH University, Kraków, Poland
{ljaneczk,faliszew}@agh.edu.pl

Abstract

Over the last few years, researchers have put significant effort into understanding of the notion of proportional representation in committee election. In particular, recently they have proposed the notion of proportionality degree. We study the complexity of computing committees with a given proportionality degree and of testing if a given committee provides a particular one. This way, we complement recent studies that mostly focused on the notion of (extended) justified representation. We also study the problems of testing if a cohesive group of a given size exists and of counting such groups.

1 Introduction

If we consider a parliamentary election where about 45% voters support party A, 30% support party B, and the remaining 25% support party C, then there are well-understood ways of assigning seats to the parties in a proportional manner. However, if instead of naming the supported parties the voters can approve each candidate individually, e.g., depending on non-partisan agendas, the situation becomes less clear. While such free-form elections are not popular in political settings, they do appear in the context of artificial intelligence. For example, they can be used to organize search results (Skowron et al. 2017), assure fairness in social media (Chakraborty et al. 2019), help design online Q&A systems (Israel and Brill 2021), or suggest movies to watch (Gawron and Faliszewski 2022). As a consequence, seeking formal understanding of proportionality in multiwinner elections is among the most active branches of computational social choice (Lackner and Skowron 2020). In this paper we extend this line of work by analyzing the computational complexity of one of the recent measures of proportionality, the proportionality degree (Aziz et al. 2018).

We consider the model of elections where, given a set of candidates, each of the n voters specifies which candidates he or she approves, and the goal is to choose a size-k subset of the candidates, called the winning committee. Since the committee is supposed to represent the voters proportionally, Aziz et al. (2017) proposed the following requirement: For each positive integer ℓ and each group of $\ell \cdot n/k$ voters who agree on at least ℓ common candidates, the committee should contain at least ℓ candidates approved by at least one of the group's members. In other words, such a group, known as an ℓ-cohesive group, deserves at least ℓ candidates, but it suffices that a single member of the group approves them. If this condition holds, then we say that the committee provides *extended justified representation* (provides EJR; if we restrict our attention to $\ell = 1$, then we speak of providing *justified representation*, JR). The key to the success of this notion is that committees providing EJR always exist and are selected by voting rules designed to provide proportional representation, such as the proportional approval voting rule (PAV) of Thiele (1895). Thus many researchers followed Aziz et al. (2017) either in defining new variants of the justified representation axioms (Sánchez-Fernández et al. 2017; Peters, Pierczynski, and Skowron 2020) or in analyzing and designing rules that would provide committees satsifying these properties (Aziz et al. 2018; Sánchez-Fernández et al. 2021; Brill et al. 2017). Others study these notions experimentally (Bredereck et al. 2019).

Yet, JR and EJR are somewhat unsatisfying. After all, to provide them it suffices that a single member of each cohesive group approves enough committee members, irrespective of all the other voters. To address this issue, Aziz et al. (2018) introduced the notion of *proportionality degree* (PD). They said that the satisfaction of a voter is equal to the number of committee members that he or she approves and they considered average satisfactions of the voters in cohesive groups. More precisely, they said that a committee has PD f, where f is a function from positive integers to nonnegative reals, if for each ℓ-cohesive group, the average satisfaction of the voters in this group is at least $f(\ell)$. So, establishing the PD of a rule gives quantitative understanding of its proportionality, whereas JR and EJR only give qualitative information. That said, Aziz et al. (2018) did show that if a committee provides EJR then it has PD at least $f(\ell) = {\ell-1}/2$, so the two notions are related. Further, they showed that PAV committees have PD $f(\ell) = \ell - 1$ and Skowron (2021) established good bounds on the PDs of numerous other proportionality-oriented rules. In particular, their results allow one to order these rules according to their theoretical guarantees, from those providing strongest guarantees to those providing the weakest.

We extend this line of work by studying the complexity of problems pertaining to the proportionality degree:

1. We show that, in general, deciding if a committee with

a given PD exists is both NP-hard and coNP-hard and we suspect it to be $\mathrm{NP}^{\mathrm{NP}}$-complete (but we only show membership). Nonetheless, we find the problem to be NP-complete for (certain) constant PD functions. These results contrast those for JR and EJR, for which analogous problems are in P.

2. We show that verifying if a given committee provides a given PD is coNP complete, which is analogous to the case of EJR. We also provide ILP formulations that may allow one to compute PDs in practice (thus, one could use them to establish an empirical hierarchy of proportionality among multiwinner voting rules).
3. We show that many of our problems are polynomial-time solvable for the candidate interval and voter interval domains of preferences (Elkind and Lackner 2015) and are fixed-parameter tractable for the parameterizations by the number of candidates or the number of voters.

We also study the complexity of finding and counting cohesive groups. We omit some proofs due to restricted space.

2 Preliminaries

For an integer t, we write $[t]$ to denote the set $\{1, \dots, t\}$. An election $E = (C, V)$ consists of a finite set C of candidates and a finite collection V of voters. Each voter $v \in V$ is endowed with a set $A(v) \subseteq C$ of candidates that he or she approves. Analogously, for each candidate c we write $A(c)$ to denote the set of voters that approve c; value $|A(c)|$ is known as the approval score of c. The election considered in the $A(\cdot)$ notation will always be clear from the context.

Multiwinner Voting Rules. A multiwinner voting rule is a function that given an election $E = (C, V)$ and a committee size $k \leq |C|$ outputs a family of winning committees, i.e., a family of size-k subsets of C. (While in practice some form of tie-breaking is necessary, theoretical studies usually disregard this issue.) Generally, we do not focus on specific rules, but the following three provide appropriate context for our discussions (we assume that $E = (C, V)$ is some election and we seek a committee of size k):

1. Multiwinner Approval Voting (AV) selects size-k committees whose members have highest total approval score. Intuitively, AV selects committees of individually excellent candidates.
2. The Approval-Based Chamberlin–Courant rule (CC) selects those size-k committees that maximize the number of voters who approve at least one member of the committee. Originally, the CC rule was introduced by Chamberlin and Courant (1983) and its approval variant was discussed, e.g., by Procaccia, Rosenschein, and Zohar (2008) and Betzler, Slinko, and Uhlmann (2013). CC selects committees of diverse candidates, that *cover* as many voters as possible.
3. Proportional Approval Voting (PAV) selects those size-k committees S that maximize the value $\sum_{i \in V} w(|S \cap A(i)|)$, where for each natural number t we have $w(t) = \sum_{j=1}^{t} 1/j$. PAV selects committees that, in a certain sense, represent the voters proportionally; see, e.g., the works of Brill, Laslier, and Skowron (2018) and Lackner and Skowron (2021). The rule is due to Thiele (1895).

All the above rules belong to the family of Thiele rules (Thiele 1895; Lackner and Skowron 2021), but there are also many other (families of) rules. For more details, we point to the survey of Lackner and Skowron (2020). Classifying multiwinner rules as focused on individual excellence, diversity, or proportionality is due to Faliszewski et al. (2017).

(Extended) Justified Representation. Let E be an election with n voters and let k be the committee size. For an integer $\ell \in [k]$, called the *cohesiveness level*, we say that a group of voters forms an ℓ-cohesive group if (a) the group consists of at least $\ell \cdot n/k$ voters, and (b) there are at least ℓ candidates approved by each member of the group. Intuitively, ℓ-cohesive groups are large enough to demand representation by at least ℓ candidates (as they include a large-enough proportion of the voters) and they can name these ℓ candidates (as there are at least ℓ common candidates that they approve). Thus many proportionality axioms focus on satisfying such demands. In particular, we are interested in the notions of (extended) justified representation, due to Aziz et al. (2017).

Definition 1. *Let $E = (C, V)$ be an election, let k be a committee size, and let S be some committee:*

1. *We say that S provides justified representation (JR) if each 1-cohesive group contains at least one voter who approves at least one member of S;*
2. *We say that S provides extended justified representation (EJR) if for each $\ell \in [k]$, each ℓ-coheseive group contains at least one voter that approves at least ℓ members of S.*

Researchers also consider the notion of *proportional justified repereserntation* (PJR) due to Sánchez-Fernández et al. (2017). Recently, Peters, Pierczynski, and Skowron (2020) also introduced the axiom of *fully justified representation* (FJR). JR is the weakest of these (in the sense that if a committee satisfies any of the other ones then it also provides JR), followed by PJR, EJR, and FJR. We focus on JR and EJR as they will suffice for our purposes. For every election and every committee size there always exists at least one committee providing EJR (thus, also JR). Indeed, all CC committees provide JR and all PAV committees also provide EJR (Aziz et al. 2017), but AV committees may fail to provide (E)JR.

Proportionality Degree. We mostly focus on the notion of a proportionality degree of a committee, introduced by Aziz et al. (2018). Let us consider some voter v and a committee S. We define v's satisfaction with S as $|A(v) \cap S|$, i.e., the number of committee members that v approves.

Definition 2. *Let E be an election, let S be a committee of size k, and let $f\colon [k] \to \mathbb{R}$ be a function. We say that S has proportionality degree f if for each ℓ-cohesive group of voters X (where $\ell \in [k]$) the average satisfaction of the voters in X is at least $f(\ell)$.*

In other words, if a committee has a certain proportionality degree f for a given election, then members of the cohesive groups in this election are guaranteed at least a certain average level of satisfaction. We are interested in several special types of proportionality degree (PD) functions:

1. We say that f is a *nonzero PD* if $f(\ell) > 0$ for all ℓ.
2. We say that f is a *unit PD* if $f(\ell) = 1$ for all ℓ.
3. We say that f is *(nearly) perfect* PD if $f(\ell) = \ell$ (if $f(\ell) = \ell - 1$) for all ℓ.

One can verify that every CC committee (or, in fact, every JR committee) has nonzero PD, and Aziz et al. (2018) have shown that every PAV committee has nearly perfect PD. It is also known that if a committee provides EJR then its proportionality degree f satisfies, at least, $f(\ell) = {}^{\ell-1}/_{2}$ (Sánchez-Fernández et al. 2017). Yet there exist elections for which no committee has unit PD or perfect PD (Aziz et al. 2018). For a detailed analysis of proportionality degrees of various multiwinner rules, we point to the work of Skowron (2021).

Computational Complexity. We assume knowledge of classic and parameterized computational complexity theory, including classes P and NP, the notions of hardness and completeness for a given complexity class, and FPT algorithms. Occasionally, we also refer to the coNP class and to higher levels of the Polynomial Hierarchy. Given a problem X from NP, where we ask if a certain mathematical object exists, we write $\#X$ to denote its variant where we ask for the number of such objects. Such problems belong to the class $\#\mathrm{P}$. It is commonly believed that if a counting problem is $\#\mathrm{P}$-complete then it cannot be solved in polynomial time. We mention that $\#\mathrm{P}$-completeness is defined via Turing reductions (Valiant 1979); this contrasts NP-completeness, defined via many-one reductions.

Computational Aspects of JR, EJR, and PD. There are polynomial-time algorithms that given an election and a committee size compute committees which provide JR (Aziz et al. 2017) or EJR (Aziz et al. 2018). On the other hand, given a committee it is easy to verify if it provides JR, but doing the same for EJR is coNP-complete (Aziz et al. 2017). In this paper we answer analogous questions for the case of the proportionality degree.

3 Finding and Counting Cohesive Groups

As cohesive groups lay at the heart of JR, EJR, and PD, we start our discussion by analyzing the hardness of finding them. More precisely, we consider the following problem.

Definition 3. *An instance of the* COHESIVE-GROUP *problem consists of an election E, a committee size k, and a positive integer ℓ. We ask if E contains an ℓ-cohesive group.*

Somewhat disappointingly, this problem is NP-complete. This follows via a reduction inspired by that provided by Aziz et al. (2017) to show that testing if a given committee provides EJR is coNP-complete (we include the proof for the sake of completeness, as some of our further hardness proofs follow by reduction from COHESIVE-GROUP).

Theorem 1. COHESIVE-GROUP *is* NP*-complete*

Proof. We observe that COHESIVE-GROUP is in NP: Given an election E with n voters, committee size k, and cohesiveness level ℓ, it suffices to nondeterministically guess a group of at least $\ell \cdot {}^{n}/_{k}$ voters and check that the intersection of their approval sets contains at least ℓ candidates.

To show that the problem is NP-hard, we give a reduction from the NP-complete problem BALANCED-BICLIQUE (Johnson 1987). The input for the latter consists of a bipartite graph G and a nonnegative integer k. The vertices of G are partitioned into two sets, $L(G)$ and $R(G)$. We write $E(G)$ to denote the set of G's edges (each edge connects a vertex from $L(G)$ with one from $R(G)$. We ask if there is a size-k subset of $L(G)$ and a size-k subset of $R(G)$ such that each vertex from the former is connected with each vertex from the latter (such two sets are jointly referred to as a k-biclique of G).

Given an instance of BALANCED-BICLIQUE, we form an instance of COHESIVE-GROUP as follows. We construct an election E', where $R(G)$ is the set of candidates and $L(G)$ is a collection of voters. A voter $\ell_i \in L(G)$ approves a candidate $r_j \in R(G)$ if ℓ_i and r_j are connected in G. We extend E' by adding $\max(||L(G)| - |R(G)|, 0)$ candidates not approved by any voter. We set the committee size to be $k' = |L|$ and the desired cohesiveness level to be $\ell' = k$. This completes the construction.

Note that each ℓ'-cohesive group in our election consists of at least $\ell' \frac{|L|}{k'} = k\frac{|L|}{|L|} = k$ voters who approve at least k common candidates. Focusing on exactly k voters and k candidates, we see that such a group exists if and only if G has a k-biclique. This completes the proof. $\square$

On the positive side, Aziz et al. (2017) gave a polynomial-time algorithm for deciding if an election contains a 1-cohesive group (we refer to this variant of the problem as ONE-COHESIVE-GROUP): It suffices to check if there is a candidate c for whom $|A(c)| \geq {}^{n}/_{k}$ (where n is the total number of voters and k is the committee size); if so, then the voters from $A(c)$ form a 1-cohesive group and if no such candidate exists then there are no 1-cohesive groups.

Corollary 1 (Aziz et al. (2017)). ONE-COHESIVE-GROUP *is in* P.

We complement the above results by considering the complexity of the #COHESIVE-GROUP problem, i.e., the problem of counting cohesive groups. An efficient algorithm for this problem would imply an efficient uniform sampling procedure (Jerrum, Valiant, and Vazirani 1986), which would be useful, e.g., to experimentally study the distribution of cohesive groups in elections.[1] Naturally, #COHESIVE-GROUP is intractable (namely, $\#\mathrm{P}$-complete), since even deciding if a single cohesive group exists is hard. More surprisingly, the same holds for 1-cohesive groups.

Theorem 2. #ONE-COHESIVE-GROUP *is* $\#\mathrm{P}$*-complete*

[1]Formally, an approximate counting algorithm would suffice to obtain a nearly uniform sampling procedure. Our results do not preclude existence of such an algorithm, but we leave studies in this direction for future work.

An intuition as to why finding a single 1-cohesive group is easy but counting them is hard is as follows. Using the argument from Corollary 1, for each candidate we can count (in polynomial time) the number of 1-cohesive groups whose members approve this candidate. Yet, if we simply added these values, then some groups could be counted multiple times. If we used the inclusion-exclusion principle, then we would get the correct result, but doing so would take exponentially many arithmetic operations.

4 Computing a Committee with a Given PD

In this section we focus on the complexity of deciding if a committee with a given proportionality degree exists. At first, this problem may seem trivial as for each election there is a committee with nearly perfect PD (Aziz et al. 2018). Yet, we find that the answer is quite nuanced. This stands in sharp contrast to analogous decision questions for JR and EJR, which *are* trivial (a committee with the desired property always exists so the algorithm always accepts). Formally, we consider the following problem.

Definition 4. *In the* PD-COMMITTEE *problem we are given an election E, a committee size k, and a function $f\colon [k] \to \mathbb{Q}$, specified by listing its values. We ask if E has a size-k committee with proportionality degree at least f.*

We find that PD-COMMITTEE is both coNP-hard and NP-hard. For the former result, we use the fact that $f(\ell)$ value of a proportionality degree function is binding only if the given election contains ℓ-cohesive groups.

Theorem 3. PD-COMMITTEE *is* NP*-hard and* coNP*-hard*

Proof. We will show NP-hardness in Theorem 6 and here we focus on coNP-hardness. Specifically, we reduce COHESIVE-GROUP to the complement of PD-COMMITTEE. Let (E, k, ℓ) be our input instance, where $E = (C, V)$ is an election, k is the committee size, and ℓ is the cohesiveness level. The question is if there exists an ℓ-cohesive group for election E with committee size k. For convenience, set $n = |V|$, and $m = |C|$.

We create an instance of the complement of PD-COMMITTEE as follows. Let s be the smallest integer such that $s \cdot k > m$. We form an election E' by copying E and (a) adding $s \cdot k$ new candidates who are not approved by any voter and (b) $(s-1) \cdot n$ new voters who do not approve any candidate. Thus, in E' we have $n' = s \cdot n$ voters, and $m' = m + s \cdot k$ candidates. Further, we set the committee size to be $k' = s \cdot k$ and we set the PD function f so that for $i < \ell$ we have $f(i) = 0$ and for $i \geq \ell$ we have $f(i) = k'$. This completes the construction.

Note that the minimum size of an ℓ-cohesive group in E' is equal to the minimum size of an ℓ-cohesive group in E, because $\frac{\ell \cdot n'}{k'} = \frac{\ell \cdot s \cdot n}{s \cdot k} = \frac{\ell \cdot n}{k}$. Thus every ℓ-cohesive group from E is also an ℓ-cohesive group for E' and vice versa.

Further, any size-k' committee must contain at least one new candidate, because $k' = s \cdot k > m$. Yet, new candidates are not approved by any voter and, so, if E' has some ℓ-cohesive group, then its average satisfaction must be strictly below $f(\ell) = k'$. This means that if E' has a committee with PD f then there are no ℓ-cohesive groups in E' (and, thus, there are no ℓ-cohesive groups in E). In other words, the answer for the PD-COMMITTEE instance is "yes" if and only if the answer for the COHESIVE-GROUP instance is "no." That is, we have reduced COHESIVE-GROUP to the complement of PD-COMMITTEE. Since, by Theorem 1, the former is NP-complete, the latter is coNP-hard. □

Since PD-COMMITTEE is both NP-hard and coNP-hard, it is unlikely that it is complete for either of these classes (we would have $\mathrm{NP} = \mathrm{coNP}$ if it were). Indeed, we suspect that it is complete for $\mathrm{NP}^{\mathrm{NP}}$ and we show that it belongs to this class. A hardness result remains elusive, unfortunately.

Theorem 4. PD-COMMITTEE *is in* $\mathrm{NP}^{\mathrm{NP}}$.

Proof sketch. Consider an instance (E, k, f) of PD-COMMITTEE. It is a "yes"-instance exactly if there exists a size-k committee such that for every $\ell \in [k]$, every ℓ-cohesive group has average satisfaction at least $f(\ell)$. Since computing an average satisfaction of a given cohesive group can be done in polynomial time, this implies that PD-COMMITTEE belongs to $\mathrm{NP}^{\mathrm{NP}}$. □

Yet, for some classes of PD functions our problem can be significantly easier. As an extreme example, for nearly perfect ones it is trivially in P. Thus we consider the following restricted variants of PD-COMMITTEE: In CONSTANT-PD-COMMITTEE we require the desired PD functions to be constant, in UNIT-PD-COMMITTEE we require them to take value 1 for each argument, and in PERFECT-PD-COMMITTEE we require them to be perfect.

We find that both CONSTANT-PD-COMMITTEE and UNIT-PD-COMMITTEE are NP-complete and, thus, likely much easier than the general variant. To establish these results, it suffices to show membership in NP for the former and NP-hardness for the latter.

Theorem 5. CONSTANT-PD-COMMITTEE *is in* NP.

Proof. Consider an instance (E, k, f) of CONSTANT-PD-COMMITTEE, where $E = (C, V)$ is an election, k is the committee size, and f is a PD function. Since f is a constant function, there is a value x such that for each $\ell \in [k]$ we have $f(\ell) = x$. To show that CONSTANT-PD-COMMITTEE is in NP, we give a polynomial-time algorithm that given such an instance and size-k committee W verifies if W has PD f.

Let $n = |V|$ be the number of voters. For each candidate $c \in C$, we define $\mathrm{sat}(c)$ to be the average satisfaction of $\lceil \frac{n}{k} \rceil$ members of $A(c)$ that are least satisfied with W; if $A(c)$ contains fewer than $\frac{n}{k}$ voters then we set $\mathrm{sat}(c) = +\infty$. We set $y = \min_{c \in C} \mathrm{sat}(c)$. If $y = +\infty$ then election E has no cohesive groups and W has PD f trivially. Otherwise, y is the smallest average satisfaction that a 1-cohesive group from E has for W (indeed, every 1-cohesive group must have at least $\lceil \frac{n}{k} \rceil$ members and for each $c \in C$, each 1-cohesive group whose members approve c has satisfaction at least $\mathrm{sat}(c)$). For each $\ell \in [k]$, each ℓ-cohesive group also has satisfaction at least y (briefly put, because it is also an 1-cohesive group with average satisfaction at least y). Thus, if $y \geq x$ then we accept and otherwise we reject. This algorithm runs in polynomial time. □

Theorem 6. UNIT-PD-COMMITTEE *is* NP*-hard*

Proof. Since UNIT-PD-COMMITTEE is a special case of CONSTANT-PD-COMMITTEE, by Theorem 5 we know that it is in NP. To show NP-hardness, we give a reduction from a variant of the classic X3C problem, which we call RX3C and which is well-known to be NP-complete (Gonzalez 1985). An instance of RX3C consists of a universe set $U = \{u_1, u_2, ..., u_{3k}\}$ and a family $\mathcal{S} = \{S_1, S_2, ..., S_{3k}\}$ size-3 subsets of U. Each element from U belongs to exactly three sets from $\mathcal{S}$. We ask if there exist k subsets from $\mathcal{S}$ which sum up to universe U.

We form an instance of UNIT-PD-COMMITTEE with an election E, committee size k, and unit PD function. We let the set from $\mathcal{S}$ be the candidates in E, and we let the universe elements be the voters. A voter u_i approves a candidate S_j if $u_i \in S_j$. This completes the construction.

We note that all cohesive groups in E contain exactly three voters and have cohesiveness level one. This holds because each candidate is approved by exactly three voters and this is also the lower bound on the size of 1-cohesive groups in E (indeed, ${}^{3k}/_{k} = 3$).

It is clear that if there exist k subsets from $\mathcal{S}$ which sum up to U, then the corresponding candidates form a committee which has average satisfaction at least 1. Indeed, for each voter there is at least one candidate in the committee that he or she approves (in fact, exactly one). Otherwise the selected sets would not sum up to U. As a consequence, the average satisfaction of each (1-)cohesive group is at least 1.

Next, let us show that if there exists a committee W of size k such that each cohesive group has average satisfaction at least 1, then there is a collection T of k sets from $\mathcal{S}$ that sum up to U (i.e., there is an exact cover of U).

Let B be the sum of the total satisfactions of all the $3k$ 1-cohesive groups in E. Since each 1-cohesive group has average satisfaction at least one, its total satisfaction is at least 3. Since there are $3k$ such groups, we have that B is at least $9k$. Moreover, B is equal to $9k$ exactly if each 1-cohesive group has average satisfaction equal to 1. However, each committee member is approved by exactly three voters, and each of these voters belongs to exactly three 1-cohesive groups. Hence $B = 9k$ and each 1-cohesive group has average satisfaction equal to 1.

Consider some set $S_j = \{u_{j_1}, u_{j_2}, u_{j_3}\}$ such that candidate S_j is a member of committee W. Naturally, $\{u_{j_1}, u_{j_2}, u_{j_3}\}$ is a 1-cohesive group, all its member approve S_j, and, so, its average satisfaction is at least 1. Indeed, by previous discussion we know that it is exactly 1. Hence, for each voter in $\{u_{j_1}, u_{j_2}, u_{j_3}\}$, candidate S_j is the only member of W that he or she approves. If we repeat this reasoning for every member of W, we find that each of them is approved by exactly three voters and no two of them are approved by the same voters. This means that W corresponds to an exact cover of U. The proof is complete. □

Corollary 2. *Both* CONSTANT-PD-COMMITTEE *and* UNIT-PD-COMMITTEE *are* NP*-complete.*

As all the cohesive groups in the election constructed in the proof of Theorem 6 have cohesiveness level 1, we have a stronger result: Given a PD function f such that $f(1) = 1$, it is NP-hard to decide if there is a committee with proportionality degree f. In particular, we have the next corollary.

Corollary 3. PERFECT-PD-COMMITTEE *is* NP*-hard.*

We can extend Theorem 6 to work for any positive integer constant x and functions f such that $f(1) = x$. For example, for $x = 2$ it suffices to extend the constructed election with three voters that do not approve anyone and with a single candidate who is approved by all the other voters. It would also be interesting to consider functions f such that $f(1)$ is a constant between 0 and 1, but we leave it for future work. The above results are nicely aligned with existing polynomial-time algorithms for computing committees with guarantees on their PD. For example, there are polynomial-time algorithms for computing EJR committees, and EJR committees are guaranteed to have PD f such that $f(\ell) = \frac{\ell-1}{2}$ (Sánchez-Fernández et al. 2017). As we see, $f(1) = 0$ (though this could be improved very slightly).

5 Computing the PD of a Given Committee

Sometimes, instead of computing a committee with a specified PD, we would like to establish the PD of an already existing one. For example, this would be the case if we wanted to experimentally compare how well the committees provided by various voting rules represent the voters.

One way to proceed would be as follows: For a given election E and committee W, consider each cohesiveness level ℓ and, using binary search (up to a given accuracy level $\varepsilon > 0$), find value $f(\ell)$, $0 \le f(\ell) \le |W|$, such that each ℓ-cohesive group has average satisfaction at least $f(\ell)$, but there exists an ℓ-cohesive group with average satisfaction below $f(\ell) + \varepsilon$ (or there are no ℓ-cohesive groups in this election). [2] Doing so requires the ability to solve the following problem.

Definition 5. *In the* PD-FAILURE *problem we are given an election E, a committee W, a cohesiveness level ℓ, and a non-negative rational threshold $y \le k$. We ask if E contains an ℓ-cohesive group whose average satisfaction for W is lower than y.*

As one may expect, this problem is NP-complete. Membership in NP follows by nondeterministically guessing an ℓ-cohesive group and checking if its average satisfaction is below y. To show NP-hardness, we note that setting the y value to an impossible-to-achieve value makes the problem equivalent to testing if an ℓ-cohesive group exists.

Theorem 7. PD-FAILURE *is* NP*-complete.*

5.1 ILP Formulation

Fortunately, in practice we may be able to solve instances of our problem by expressing them as integer linear programs (ILPs) and solving them using off-the-shelf software.

[2] In fact, we can calculate the exact value of $f(\ell)$, since it is either unbounded or a fraction with the denominator equal to $s = \lceil \ell \cdot {}^{n}/_{k} \rceil$ (see the argument in Section 5.1) and the numerator in range $[0,\ s \cdot k]$. Such an approach would require only $O(log(s \cdot k + 2)) = O(log(\ell \cdot n))$ queries to a PD-FAILURE oracle.

Specifically, let us consider an instance of PD-FAILURE with election $E = (C, V)$, committee W, cohesiveness level ℓ, and threshold y. We set $m = |C|$, $n = |V|$, and $k = |W|$. For convenience, let A be the binary matrix of approvals for E, that is, we have $a_{ij} = 1$ if the i-th voter approves the j-th candidate, and we have $a_{ij} = 0$ otherwise. We note that if there is an ℓ-cohesive group X whose satisfaction for W is below y, then there is such a group of size exactly $s = \lceil \ell \cdot {}^{n}/_{k} \rceil$ (e.g., consider X and remove sufficiently many voters who approve the most members of W).

To form our ILP instance, we first specify the variables:

1. For each $i \in [n]$, we have a binary variable x_i, with the intention that $x_i = 1$ if the i-th voter is included in the sought cohesive group, and $x_i = 0$ otherwise.
2. For each $j \in [m]$, we have a binary variable y_i, with the intention that $y_j = 1$ if all the voters in the group specified by variables $x_1, \ldots, x_n$ approve the j-th candidate, and $y_j = 0$ otherwise.

We refer to the voters (to the candidates) whose x_i (y_j) variables are set to 1 as *selected*. Next, we specify the constraints. Foremost, we ensure that we select exactly s voters and at least ℓ candidates:

$$\textstyle\sum_{i=1}^{n} x_i = s, \qquad \text{and} \qquad \sum_{j=1}^{m} y_j \geq \ell.$$

Then, we ensure that each selected voter approves all the selected candidates. For each $j \in [m]$, we form constraint:

$$\textstyle\sum_{i=1}^{n} a_{ij} \cdot x_i \geq s \cdot y_j.$$

If the j-th candidate is not selected, then this inequality is satisfied trivially. However, if the j-th candidate is selected, then the sum on the left-hand side must be at least s, i.e., there must be at least s selected voters who approve the j-th candidate. Since there are exactly s selected voters, all of them must approve the j-th candidate.

Finally, we ensure that the average satisfaction of the selected voters is below y, by adding constraint $\frac{1}{s}\sum_{i=1}^{n}\sum_{j \in W} a_{ij} \cdot x_i < y$. If there is an assignment that satisfies these constraints, then the selected voters form an ℓ-cohesive group with average satisfaction below y. Otherwise, no such group exists.

5.2 Verification

For a comparison with previous studies regarding JR and EJR, we also consider the following verification problem.

Definition 6. *In the* PD-VERIFICATION *problem we are given an election E, a committee W, a PD function f, and we ask if W has proportionality degree f.*

As PD-VERIFICATION is very closely related to the complement of PD-FAILURE, we find that it is coNP-complete.

Corollary 4. PD-VERIFICATION *is* coNP*-complete.*

Testing if a committee provides EJR is coNP-complete as well (Aziz et al. 2017), so in this respect PD and EJR are analogous. There is also a polynomial-time algorithm for testing if a committee provides JR, and in the PD world this corresponds to a polynomial-time algorithm for checking if a committee admits a given constant PD function. Such an algorithm was included as part of the proof of Theorem 5.

Corollary 5. PD-VERIFICATION *for a constant PD functions (provided as input) is in* P.

6 Dealing With Computational Hardness

In this section we consider circumventing the computational hardness of our problems by studying their parameterized complexity and by considering structured elections.

6.1 Fixed-Parameter Tractability

Our two main problems, PD-COMMITTEE and PD-FAILURE, are fixed-parameter tractable with respect to the number of candidates and the number of voters.

Theorem 8. *There are FPT algorithms for* PD-COMMITTEE *and* PD-FAILURE *both for the parameterization by the number of candidates and for the parameterization by the number of voters.*

For PD-FAILURE and the parameterization by the number of candidates, we proceed similarly as in the proof of Theorem 5. Namely, for each set R of candidates we consider the set $V(R)$ of all the voters that approve members of R, one-by-one, remove from this set the voters with the highest satisfation, and watch if at any point we obtain an ℓ-cohesive group with average satisfaction below the required value. Using a similar approach, and trying every possible committee, we also obtain an algorithm for PD-COMMITTEE.

For the parameterization by the number of voters, we solve our problems by forming ILP instances and solving them using the classic algorithm of Lenstra, Jr. (1983). This is possible because with n voters there are at most 2^n cohesive groups and each candidate has one of 2^n types (where the type of a candidate is the set of voters that approve him; candidates with the same type are interchangeable).

Testing if a committee provides EJR is also fixed-parameter tractable for the parameterizations considered in Theorem 8. So, from this point of view, dealing with PD is not harder than dealing with EJR.

Finally, the problem of counting cohesive groups (and, thus, also the problem of deciding if groups with particular cohesiveness level exist) also is fixed-parameter tractable for our parameters. For parameterization by the number of candidates, we can use the inclusion-exclusion principle, and for the parameterization by the number of voters we can explicitly look at each subset of voters.

Theorem 9. *There are FPT algorithms for* (#)COHESIVE-GROUP, *for the parameterizations by the number of candidates and by the number of voters.*

6.2 Structured Preferences

Next we consider two domains of structured preferences, introduced by Elkind and Lackner (2015).

Definition 7 (Elkind and Lackner (2015)). *An election $E = (C, V)$ has candidate interval (CI) preferences (voter interval preferences, VI) if it is possible to order the candidates (the voters) so that for each voter v (for each candidate c) the set $A(v)$ (the set $A(c)$) is an interval w.r.t. this order.*

For an example of CI preferences, consider a political election where candidates are ordered according to the left-to-right spectrum of opinions and the voters approve ranges of candidates whose opinions are close enough to their own. Elkind and Lackner (2015) gave algorithms for deciding if a given election has CI or VI preferences, and for computing appropriate orders of candidates or voters. Thus, for simplicity, we assume that these orders are provided together with our input elections. We mention that a number of other preference domains are considered in the literature—see, e.g., the works of Yang (2019) and Godziszewski et al. (2021)—but the CI and VI ones are by far the most popular.

Unfortunately, even for CI and VI elections we do not know how to solve the PD-COMMITTEE problem in polynomial-time.[3] Nonetheless, we do have polynomial-time algorithms for the PD-FAILURE problem.

Theorem 10. PD-FAILURE *restricted to either CI or VI elections is in* P.

Proof. First, we give an algorithm for the CI case. Our input consists of an election $E = (C, V)$, where $C = \{c_1, \dots, c_m\}$ and $V = (v_1, \dots, v_n)$, a size-k committee W, cohesiveness level ℓ, and threshold value y. Without loss of generality, we assume that E is CI with respect to the order $c_1 \lhd c_2 \lhd \cdots \lhd c_m$.

Since E is a CI election, we observe that if X is some cohesive group whose all members approve some two candidates c_i and c_j, $i \leq j$, then all members of X also approve candidates $c_{i+1}, \dots, c_{j-1}$. For each $i \leq m - \ell + 1$, let $X(i)$ be the set of all voters who approve each of the candidates $c_i, \dots, c_{i+\ell-1}$. By the preceding argument, we see that every ℓ-cohesive group can be obtained by taking some set $X(i)$ and (possibly) removing some of its members.

Our algorithm proceeds as follows. For each set $X(i)$, we form a set $Y(i)$ by taking $X(i)$ and removing all but $\lceil \ell \cdot n/k \rceil$ voters that are least satisfied with W. (If a given $X(i)$ contains fewer than $\lceil \ell \cdot n/k \rceil$ voters then we set $Y(i) = \emptyset$ and we assume that the average satisfaction of its voters is $+\infty$.) If there is some i such that the average satisfaction of the voters in $Y(i)$ is below y, then we accept (indeed, we have just found an ℓ-cohesive group with average satisfaction below y). If there is no such $Y(i)$, then we reject (we do so because each nonempty $Y(i)$ has the lowest average satisfaction among all the ℓ-cohesive groups that can be obtained by removing voters from $X(i)$). Correctness and polynomial running time follow immediately.

Now let us consider the VI case. We use the same notation as before, except that we assume that E is VI with respect to the voter order $v_1 \lhd v_2 \lhd \cdots \lhd v_n$. We use the same algorithm as in the CI case, but for the $X(i)$ sets defined as follows (let $s = \lceil \ell \cdot n/k \rceil$): For each $i \in [n - s + 1]$, we let $X(i) = \{v_i, v_{i+1}, \dots, v_j\}$, where j is the largest value such that $|A(v_i) \cap A(v_j)| \geq \ell$ (if v_i approves fewer than ℓ candidates then $X(i)$ is empty). The algorithm remains correct because, as in the CI case, every ℓ-cohesive group is a subset of some $X(i)$. □

[3]The approach of Peters and Lackner (2020) based on solving totally unimodular ILP instances does not seem to work here.

	c_1	c_2	c_3	c_4	c_5	c_6	c_7
v_1	1	-	-	-	-	-	-
v_2	1	-	-	-	-	-	-
v_3	1	1	-	-	-	-	-
v_4	-	1	-	-	-	-	-
v_5	-	1	1	-	-	-	-
v_6	-	-	1	-	-	-	-
v_7	-	-	1	1	-	-	-
v_8	-	-	-	1	-	-	-
v_9	-	-	-	1	1	-	-
v_{10}	-	-	-	-	1	-	-
v_{11}	-	-	-	-	1	1	-
v_{12}	-	-	-	-	-	1	-
v_{13}	-	-	-	-	-	1	1
v_{14}	-	-	-	-	-	-	1
v_{15}	-	-	-	-	-	-	1

Table 1: Approval sets used in Example 1.

Similar reasoning and observations as in the above proof also give the algorithms for counting cohesive groups (and, thus, for deciding their existence).

Theorem 11. (#)COHESIVE-GROUP *restricted to either CI or VI elections is in* P.

Similar approach shows that testing if a committee provides EJR can be done in polynomial time for CI or VI elections (to the best of our knowledge, this is a folk result).

Perfect PD? Aziz et al. (2018) have shown that for each election and each committee size there is a committee with a nearly perfect PD, but there are scenarios where committees with perfect PDs do not exist. Unfortunately, this remains true even if the elections are CI and VI at the same time.

Example 1. *Consider an election $E = (C, V)$, where $C = \{c_1, \dots, c_7\}$, and $V = (v_1, \dots, v_{15})$. The set the committee size to be $k = 5$, and the approval sets are as in Table 1 Clearly, the election is both CI and VI. We see that $n/k = 3$ and, thus, for each $i \in [7]$, voters $v_{2i-1}, v_{2i}, v_{2i+1}$ form a cohesive group (for candidate c_i).*

Now consider some size-k committee. If it does not contain some candidate c_i, then the 1-cohesive group $\{v_{2i-1}, v_{2i}, v_{2i+1}\}$ must have average satisfaction below 1. Indeed, altogether members of this group give at most five approvals, of which three go to c_i. Thus, without c_i, the average satisfaction is at most $\frac{2}{3} < 1$. However, since the committee size is five and there are seven candidates, for each committee there is some 1-cohesive group with satisfaction below 1. Thus there is no committee with a perfect PD for this election and committee size five.

7 Conclusions and Future Work

We have shown that computing committees with a given proportionality degree is, apparently, more difficult that computing EJR committees, but verification problems for these two notions have the same complexity. Two most natural directions of future work would be to establish the exact complexity of the PD-COMMITTEE problem and experimentally analyze PDs of committees provided by various voting rules.

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).

References

Aziz, H.; Brill, M.; Conitzer, V.; andR. Freeman, E. E.; and Walsh, T. 2017. Justified Representation in Approval-Based Committee Voting. *Social Choice and Welfare*, 48(2): 461–485.

Aziz, H.; Elkind, E.; Huang, S.; Lackner, M.; Sánchez-Fernández, L.; and Skowron, P. 2018. On the Complexity of Extended and Proportional Justified Representation. In *Proceedings of AAAI-18*, 902–909.

Betzler, N.; Slinko, A.; and Uhlmann, J. 2013. On the Computation of Fully Proportional Representation. *Journal of Artificial Intelligence Research*, 47: 475–519.

Bredereck, R.; Faliszewski, P.; Kaczmarczyk, A.; and Niedermeier, R. 2019. An experimental view on committees providing justified representation. In *Proceedings of IJCAI-19*, 109–115.

Brill, M.; Freeman, R.; Janson, S.; and Lackner, M. 2017. Phragmén's Voting Methods and Justified Representation. In *Proceedings of AAAI-17*, 406–413.

Brill, M.; Laslier, J.; and Skowron, P. 2018. Multiwinner Approval Rules as Apportionment Methods. *Journal of Theoretical Politics*, 30(3): 358–382.

Chakraborty, A.; Patro, G.; Ganguly, N.; Gummadi, K.; and Loiseau, P. 2019. Equality of Voice: Towards Fair Representation in Crowdsourced Top-K Recommendations. In *Proceedings of FAT-19*, 129–138.

Chamberlin, B.; and Courant, P. 1983. Representative Deliberations and Representative Decisions: Proportional Representation and the Borda Rule. *American Political Science Review*, 77(3): 718–733.

Elkind, E.; and Lackner, M. 2015. Structure in Dichotomous Preferences. In *Proceedings of IJCAI-15*, 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*. AI Access Foundation.

Gawron, G.; and Faliszewski, P. 2022. Using Multiwinner Voting to Search for Movies. Technical Report arXiv:2202.03385 [cs.GT], arXiv.org.

Godziszewski, M.; Batko, P.; Skowron, P.; and Faliszewski, P. 2021. An analysis of approval-based committee rules for 2D-Euclidean elections. In *Proceedings of AAAI-21*, 5448–5455.

Gonzalez, T. 1985. Clustering to Minimize the Maximum Intercluster Distance. *Theoretical Computer Science*, 38: 293–306.

Israel, J.; and Brill, M. 2021. Dynamic Proportional Rankings. In *Proceedings of IJCAI-21*, 261–267.

Jerrum, M.; Valiant, L.; and Vazirani, V. 1986. Random Generation of Combinatorial Structures from a Uniform Distribution. *Theoretical Computer Science*, 43: 169–188.

Johnson, D. S. 1987. The NP-completeness column: An ongoing guide. *Journal of Algorithms*, 8: 438–448.

Lackner, M.; and Skowron, P. 2020. Approval-Based Committee Voting: Axioms, Algorithms, and Applications. *CoRR*, abs/2007.01795.

Lackner, M.; and Skowron, P. 2021. Consistent approval-based multi-winner rules. *Journal of Economic Theory*, 192: 105173.

Lenstra, Jr., H. 1983. Integer Programming with a Fixed Number of Variables. *Mathematics of Operations Research*, 8(4): 538–548.

Peters, D.; and Lackner, M. 2020. Preferences Single-Peaked on a Circle. *Journal of Artificial Intelligence Research*, 68: 463–502.

Peters, D.; Pierczynski, G.; and Skowron, P. 2020. Proportional Participatory Budgeting with Cardinal Utilities. Technical Report arXiv:2008.13276 [cs.GT], arXiv.org.

Procaccia, A.; Rosenschein, J.; and Zohar, A. 2008. On the Complexity of Achieving Proportional Representation. *Social Choice and Welfare*, 30(3): 353–362.

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

Sánchez-Fernández, L.; García, N. F.; Fisteus, J.; and Brill, M. 2021. The Maximin Support Method: An Extension of the D'Hondt Method to Approval-Based Multiwinner Elections. In *Proceedings of AAAI-21*, 5690–5697.

Skowron, P. 2021. Proportionality Degree of Multiwinner Rules. In *Proceedings of EC-21*, 820–840.

Skowron, P.; Lackner, M.; Brill, M.; Peters, D.; and Elkind, E. 2017. Proportional Rankings. In *Proceedings of IJCAI-17*, 409–415.

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

Valiant, L. 1979. The Complexity of Computing the Permanent. *Theoretical Computer Science*, 8(2): 189–201.

Yang, Y. 2019. On the tree representations of dichotomous preferences. In *Proceedings of IJCAI-19*, 644–650.