Proportionality in Approval-Based Elections With a Variable Number of Winners

Rupert Freeman
Microsoft Research New York City

Anson Kahng
Carnegie Mellon University

David M. Pennock
Rutgers University

ABSTRACT

We study proportionality in approval-based multiwinner elections with a variable number of winners, where both the size and identity of the winning committee are informed by voters' opinions. While proportionality has been studied in multiwinner elections with a fixed number of winners, it has not been considered in the variable number of winners setting. The measure of proportionality we consider is *average satisfaction (AS)*, which intuitively measures the number of agreements on average between sufficiently large and cohesive groups of voters and the output of the voting rule. First, we show an upper bound on the AS that any deterministic rule can provide, and that straightforward adaptations of deterministic rules from the fixed number of winners setting do not achieve better than a 1/2 approximation to AS even for large numbers of candidates. We then prove that a natural randomized rule achieves a 29/32 approximation to AS.

1 INTRODUCTION

We study *multiwinner approval-based elections*, where a group of agents, or voters, selects a committee from a set of candidates based on the agents' preferences. Each agent expresses her preferences through an approval vote, where she designates a subset of candidates she approves for the committee, and all votes are then aggregated to select a winning committee from the pool of candidates.

Some multiwinner elections include a fixed committee size: voters must fill exactly k seats on a committee. This is known as the fixed number of winners (FNW) setting, and there is a large body of work on the complexity and proportionality of various voting rules in the FNW setting [1–3, 5, 11, 13, 15]. In contrast, we are interested in the setting in which there is no a priori fixed committee size, also known as the variable number of winners (VNW) setting. In this case, both the size of the committee and the candidates chosen to sit on the committee are informed by agents' votes.

We present two examples of settings where VNW elections are a natural fit; Faliszewski *et al.* [2017] discuss others.

The first example is a hiring scenario where the number of interviews responds to the quality of candidates. Each member of a hiring task force rates job candidates and, when many strong candidates have applied, more interviews will be granted, even if the number of jobs is limited. The second example is Hall of Fame balloting, where applicants are selected if they are deemed to be of a certain quality. For instance, in Major League Baseball, players are only inducted into the Hall of Fame if at least 75% of the voters approve them, resulting in class sizes that fluctuate based on the perceived qualities of eligible candidates each year.

In both of these examples (and, indeed, in any multiwinner election), it is important to ensure that the selected alternatives are chosen in a proportional manner. For instance, when screening candidates for a job interview, it may be important that all divisions in the company are satisfied with at least some of the candidates that made it past the initial screening process. In other words, if a large division of the company pushes for one subset of candidates A and a small division of the company pushes for another (disjoint) subset of candidates B, the set of candidates that progress past the screening phase should contain representatives from both A and B.

In order to study proportionality in FNW elections, researchers have proposed the axioms of justified representation (JR), proportional justified representation (PJR), extended justified representation (EJR), and average satisfaction (AS) [2, 13], which capture the intuition that all sufficiently large groups that agree on sufficiently many candidates should achieve some measure of satisfaction. However, to our knowledge, we are the first to study representation in VNW elections.

Our Contributions: Our main research goal is to study proportionality in multiwinner elections with a variable number of winners. In particular, we study the proportionality measure of average satisfaction (AS) and show that there is a separation between the performance of deterministic and randomized voting rules.

As our first contribution, we develop a framework for thinking about proportionality in VNW elections. Previous work on proportionality in FNW elections is largely based on the concept of justified representation (and extensions thereof). However, as we discuss in Section 3, JR-based notions of proportionality are less compelling in VNW elections than in FNW elections. Therefore, we instead base our approach on the concept of average satisfaction, which is arguably a more robust version of justified representation.

Second, in Section 4, we consider the proportionality guarantees of deterministic rules in the VNW setting. We extend three existing deterministic rules for the FNW setting to the VNW setting, and show that these rules do not guarantee good approximations to average satisfaction. We also prove upper bounds on the level of average satisfaction that any deterministic rule can provide.

Finally, in Section 5, motivated by the shortcomings of deterministic rules, we turn our attention to randomized rules and show that a natural randomized rule provides a good approximation to average satisfaction.

Related Work

There is a significant body of work studying proportionality in FNW elections. As mentioned above, Aziz *et al.* [2017] put forward the compelling axiom of justified representation (JR), as well as a stronger version of this axiom, extended justified representation

Proc. of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020), B. An, N. Yorke-Smith, A. El Fallah Seghrouchni, G. Sukthankar (eds.), May 2020, Auckland, New Zealand

https://doi.org/doi

(EJR) to capture the notion that any sufficiently large and cohesive group of voters deserves some measure of representation in the elected committee. This idea was built upon by Sánchez-Fernández *et al.* [2017], who introduced the intermediate axiom of proportional justified representation (PJR), a relaxation of EJR that is more stringent than JR.

Average satisfaction (AS) was first defined by Sánchez-Fernández *et al.* [2017]. In this paper, they studied the average satisfaction guaranteed by extended justified representation (EJR). Further work by Aziz *et al.* [2018] showed that Proportional Approval Voting (PAV) guarantees a level of average satisfaction that implies EJR. Additionally, Skowron *et al.* [2017a] extend the notion of average satisfaction to the context of complete rankings as opposed to committee selection. Further work by Skowron [2018] studies the proportionality degree of various multiwinner rules by considering the average satisfaction of all groups of a certain size.

There is also a significant body of work studying VNW elections; however, to the best of our knowledge, none of these works consider proportionality. Kilgour [2016] proposes a multitude of rules for VNW elections; however, none of the proposed VNW rules are proportional. Fishburn and Pekeč [2004] study threshold approaches to committee selection, which are VNW rules in the sense that the size of the selected committee depends on the approval votes. However, threshold approaches are also not proportional. Additionally, the Borda Mean Rule, which was characterized by Peters and Brandl [2019] and also studied by Duddy *et al.* [2016], can be seen as a VNW rule with approval votes, but is also not proportional. Finally, Faliszewski *et al.* [2017] study the computational complexity of various VNW rules, but do not consider proportionality in their analysis.

2 PRELIMINARIES

Let $N = \{v_1, \dots, v_n\}$ be a set of n voters and $C = \{c_1, \dots, c_m\}$ be a set of m candidates. For every voter v_i, denote by $A_i \subseteq C$ the set of candidates that are *approved* by v_i. A preference profile $\mathbf{A} = \{A_1, \dots, A_n\}$ is the set of all voter preferences A_i.

A variable number of winners (VNW) voting rule f takes as input a preference profile $\mathbf{A}$ and outputs some set of candidates $f(\mathbf{A}) \subseteq C$. Note that we allow $f(A) = \emptyset$ or $f(A) = C$. We will also consider randomized VNW voting rules that output a distribution over sets of candidates.

Throughout this paper, we will denote by W the set of candidates included on the committee, and we will denote by $C \setminus W$ the set of candidates who were excluded from the committee.

We say that a group of voters $V \subseteq N$ is ℓ-large if $|V| \geq \ell \cdot \frac{n}{m}$, and ℓ-cohesive if $|\cap_{i \in V} A_i| + |\cap_{i \in V} C \setminus A_i| \geq \ell$. We will also say that a group of voters V *agrees* on a candidate c_j if $c_j \in A_i$ for all $i \in V$ or $c_j \notin A_i$ for all $i \in V$. Otherwise, we say that V *disagrees* on c_j.

In our work, we consider a different measure of representation than in the FNW setting. In the FNW setting, voters derive utility from the number of their approved candidates who were elected to the committee. However, this definition does not make sense in the VNW setting because then a rule could maximally satisfy all voters by including all candidates on the committee. Therefore, we assume that voters derive utility from agreeing with the placement of candidates either on the committee or not on the committee. For instance, in an election with two candidates, c_1 and c_2, if a voter i has approval set $A_i = \{c_1\}$ (i.e., she approves c_1 and disapproves c_2), then she receives one unit of utility for the output committee $\{c_1, c_2\}$ because she agrees with the inclusion of c_1 but disagrees with the inclusion of c_2.

With this in mind, the following definition of average satisfaction is adapted from the definition of Sánchez-Fernández *et al.* [2017] in the FNW setting.

Definition 2.1. Given a set of candidates $W \subseteq C$, the *average satisfaction* of a group of voters $V \subseteq N$ is

$$avs_W(V) = \frac{1}{|V|} \sum_{i \in V} (|A_i \cap W| + |(C \setminus A_i) \cap (C \setminus W)|).$$

We can now define AS in the VNW setting. The intuition behind the following definition is that any sufficiently large and cohesive group of voters deserves to be adequately represented *on average*, which is a departure from justified representation-based axioms that have been studied in the FNW setting. Intuitively, JR-like notions of proportionality only require that some member of each cohesive group is represented to some extent, whereas average satisfaction requires all members of each cohesive group to be represented simultaneously (at least on average).

Definition 2.2. A set of candidates $W \subseteq C$ satisfies *α-AS* if, for all ℓ-large and ℓ-cohesive groups of voters $V \subseteq N$, $avs_W(V) \geq \alpha \cdot \ell$ for all $\ell \in [m]$. For brevity, we refer to the special case of 1-AS as AS.

The following example demonstrates cohesiveness and average satisfaction.

Example 2.3. Consider the following profile with $n = 8$ voters, $v_1, \dots, v_8$, and $m = 4$ candidates, $c_1, \dots, c_4$, with preferences

$$A_1 = A_2 = \{c_1, c_2, c_3, c_4\} \qquad A_6 = \{c_2, c_3\}$$
$$A_3 = A_4 = \{c_1, c_2\} \qquad A_7 = \{c_3\}$$
$$A_5 = \{c_1, c_3\} \qquad A_8 = \{c_4\}.$$

Now, consider the output $W = \{c_4\}$. Note that each voter agrees with the output on the placement of at least one candidate, so for any 1-large and 1-cohesive group $V_{(1)}$ (i.e., a group of $1 \cdot \frac{n}{m} = 2$ voters who agrees on the placement of 1 candidate), $avs_W(V_{(1)}) = 1$. Furthermore, note that there is only one 2-large and 2-cohesive group of voters: v_1, v_2, v_3, and v_4 agree on the placement of c_1 and c_2, but disagree on the placement of c_3 and c_4, so they constitute a 2-large group of voters who agree on 2 candidates. Let $V_{(2)} = \{v_1, v_2, v_3, v_4\}$. Note that $avs_W(V_{(2)}) = 1$ because each $v \in V$ agrees with W on exactly one placement, but because this group of voters is 2-large and 2-cohesive, we see that W only satisfies 1/2-AS in this scenario.

Additionally, we straightforwardly extend the following deterministic multiwinner rules from the FNW setting to the VNW setting.

Proportional Approval Voting (PAV):. Under the PAV rule [17], voter i derives utility $H_k = 1 + 1/2 + \dots + 1/k$ from a committee W and unchosen alternatives $C \setminus W$ if the total number of agreements between A_i and W plus the total number of agreements between

$C \setminus A_i$ and $C \setminus W$ is k; i.e., if voter i agrees with a total of k placements of candidates in the output. The goal of PAV is to maximize the sum of all voters' utilities, and thus PAV outputs the subset $W \subseteq C$ with highest PAV-score.

Sequential Phragmén (seq-Phragmén): The seq-Phragmén rule [5] is a sequential variant of a class of rules called Phragmén's methods, which select committees such that the "load" incurred by all members of the committee is distributed as evenly among voters as possible.[1] Each candidate carries a load of one unit, and this load is distributed among voters who agree with the placement of this candidate in either the included set or excluded set. The seq-Phragmén rule proceeds iteratively by, in each round, placing the candidate that results in the smallest increase in the maximal load of any voter.

Let $x_i^{(t)}$ denote the load of voter i at time t, and let $s^{(t)}$ denote the maximal voter load at time t. All voters start out with no load. Furthermore, let $N_j = \{i \in N : c_j \in A_i\}$ represent the set of voters that approve of candidate c_j. The maximal voter load if, on the $t + 1^{st}$ placement, candidate c_j is included in the committee is

$$s^{(t+1)}(c_j) = \frac{1 + \sum_{i \in N_j} x_i^{(t)}}{|N_j|},$$

and the maximal voter load if candidate c_j is excluded from the committee is

$$s^{(t+1)}(\bar{c_j}) = \frac{1 + \sum_{i \in C \setminus N_j} x_i^{(t)}}{|C \setminus N_j|}$$

because the load is distributed so as to equalize the loads of all voters who agree with the placement of c_j. At each step t, seq-Phragmén places the candidate c_j that minimizes $\min(s^{(t)}(c_j), s^{(t)}(\bar{c_j}))$ and updates voter loads accordingly: in the case that c_j is included in the committee,

$$x_i^{(t+1)} = \begin{cases} s^{(t+1)}(c_j) & \text{if } i \in N_j \\ x_i^{(t)} & \text{otherwise,} \end{cases}$$

and in the case that c_j is excluded from the committee,

$$x_i^{(t+1)} = \begin{cases} s^{(t+1)}(\bar{c_j}) & \text{if } i \in C \setminus N_j \\ x_i^{(t)} & \text{otherwise.} \end{cases}$$

This rule proceeds until all candidates have been placed, and then returns the included and excluded candidates.

Rule X:. Rule X [11] allocates each voter a budget of one dollar, which they then spend on placing candidates either in the included set or excluded set. Placing a candidate costs n/m dollars in total, and the set of voters who agree on the placement of this candidate must be able to collectively afford the placement price. The rule starts with an empty included set W and an empty excluded set $\bar{W}$, and it iteratively places candidates in either committee as follows.

Let $b_i(t)$ be the amount of money that voter i has remaining at the beginning of the t^{th} step; i.e., $b_i(1) = 1$ for all voters $v_i \in N$. At the t^{th} step, we say that a candidate $c \notin W \cup \bar{W}$ is q-affordable for some $q \geq 0$ if

$$\max\left(\sum_{i: c \in A_i} \min(q, b_i(t)), \sum_{i: c \in C \setminus A_i} \min(q, b_i(t))\right) \geq n/m.$$

In other words, candidate c is q-affordable if it can be placed in either the included or excluded set while voters who approve or disapprove of c each pay a maximum of q dollars. If no candidate is q-affordable for any $q \geq 0$, then the rule stops and returns the current set of included candidates W and places the rest of the candidates arbitrarily.[2] Else, the rule places the candidate which is q-affordable for the minimum value q in either the approved or disapproved committee. Each voter who agrees with this placement has $\min(q, b_i(t))$ deducted from their budget, and the process continues.

3 JUSTIFIED REPRESENTATION IN VNW ELECTIONS

In order to build intuition about why we focus on AS instead of (E/P)JR, we begin by defining JR, PJR, and EJR for VNW elections. In each case, the definition is a straightforward adaptation of the corresponding definition for the FNW setting, where we intuitively replace "agreement with members on the committee" with "agreement on the placement of each candidate." We slightly overload notation—namely, JR, EJR, and PJR—from the FNW setting in the following definitions.

Definition 3.1 (JR). Consider a ballot profile $\mathbf{A}$. A set of candidates $W \subseteq C$ satisfies *justified representation (JR)* with respect to $\mathbf{A}$ if, for all sets of 1-large and 1-cohesive voters N^*, there exists an $i \in N^*$ such that $|A_i \cap W| + |(C \setminus A_i) \cap (C \setminus W)| \geq 1$.

Definition 3.2 (PJR). Consider a ballot profile $\mathbf{A}$. A set of candidates $W \subseteq C$ satisfies *proportional justified representation (PJR)* if, for all sets of ℓ-large and ℓ-cohesive voters N^*, $|\bigcup_{i \in N^*} A_i \cap W| + |(\bigcup_{i \in N^*} (C \setminus A_i)) \cap (C \setminus W)| \geq \ell$ for all $\ell \in [m]$.

Definition 3.3 (EJR). Consider a ballot profile $\mathbf{A}$. A set of candidates $W \subseteq C$ satisfies *extended justified representation (EJR)* with respect to $\mathbf{A}$ if, for all sets of ℓ-large and ℓ-cohesive voters N^*, there exists an $i \in N^*$ such that $|A_i \cap W| + |(C \setminus A_i) \cap (C \setminus W)| \geq \ell$ for all $\ell \in [m]$.

The following example illustrates these definitions, and provides intuition about why PJR is less compelling in the VNW setting than in the FNW setting.

Example 3.4. Consider the same profile as in Example 2.3 with $n = 8$ voters, $v_1, \ldots, v_8$, and $m = 4$ candidates, $c_1, \ldots, c_4$.

Again, consider the output $W = \{c_4\}$. W satisfies JR because each voter agrees with the output on the placement of at least one candidate. Furthermore, W satisfies PJR because, on the only 2-large and 2-cohesive group of voters, $\{v_1, v_2, v_3, v_4\}$, two of them agree with the placement of c_3 and two of them agree with the placement of c_4. However, W does not satisfy EJR because no voter in the coalition agrees with two placements of W—they all agree with exactly one placement.

[1] For a survey, see [9, 12].

[2] Other natural options are to include or exclude all unplaced candidates; we note that our results about Rule X do not depend on the exact choice of what is done with unplaced candidates.

We also study the relationship between natural extensions of FNW rules (namely, PAV, seq-Phragmén, and Rule X) and different notions of justified representation in the VNW setting.

Lemma 3.5. *PAV satisfies PJR.*

Proof. Assume that there exists an ℓ-large, ℓ-cohesive coalition, G, that agrees on $\ell' \geq \ell$ candidates C_G, but only receives $k < \ell$ satisfaction from the output W. We will show that it is always possible to find a new output W' with a higher PAV score than W for which G is $(k+1)$-satisfied. In particular, we show that there exists a W' that differs from W only on the placement of one candidate (i.e., W' includes one additional candidate that was previously excluded from W or excludes one additional candidate that was previously included in W) that satisfies these properties.

Let W agree with the placement of $\beta < \ell$ of the candidates that G agrees upon. Note that $\beta + (m - \ell') = k$ because, according to the definition of PJR, G derives satisfaction from all candidates they disagree on, which means that $\beta + (m - \ell') < \ell$ and therefore $\ell + \ell' > \beta + m$.

In order to argue that there exists a W' that satisfies (a) and (b), we examine what happens if we change the placement of each of the $\ell' - \beta$ candidates that G agrees upon, but W does not agree with G on. Changing the placement of each of these candidates clearly results in $(k+1)$-representation for G, and we will show that, on average, changing the placement of one of these candidates increases the PAV score. We let $\Delta_{PAV}(G)$ denote the total change in PAV score for voters in G and $\Delta_{PAV}(N \setminus G)$ denote the total change in PAV score for voters in $N \setminus G$, and analyze these two quantities separately.

For each voter in G, her PAV score increases by at least $1/(k+1) \geq 1/\ell$ on each of the $\ell' - \beta$ possible changes because each member is at most k-satisfied. Furthermore, there are at least $\ell n/m$ voters in G. Therefore,

$$\Delta_{PAV}(G) \geq \frac{\ell n}{m} \cdot \frac{1}{\ell}.$$

For each voter v_i in $N \setminus G$, let v_i agree with the placement of x_i candidates in C_G and y_i candidates in $C \setminus C_G$ for a total of $x_i + y_i$ agreements with W. Initially, the PAV score of v_i is $H_{x_i+y_i}$. However, the average PAV score of each v_i in $N \setminus G$ over all $\ell' - \beta$ possible changes is

$$\frac{x_i H_{x_i+y_i-1} + ((\ell'-\beta) - x_i) H_{x_i+y_i+1}}{\ell' - \beta},$$

so the average change in PAV score for each v_i not in G is

$$\begin{aligned}
&\frac{x_i H_{x_i+y_i-1} + ((\ell'-\beta) - x_i) H_{x_i+y_i+1}}{\ell' - \beta} - H_{x_i+y_i} \\
&= \frac{1}{\ell'-\beta}\left((\ell'-\beta)H_{x_i+y_i-1} + (\ell'-\beta-x_i)\left(\frac{1}{x_i+y_i} + \frac{1}{x_i+y_i+1}\right)\right) \\
&\quad - \frac{\ell'-\beta}{\ell'-\beta} H_{x_i+y_i} \\
&= \frac{1}{\ell'-\beta}\left(\frac{-x_i}{x_i+y_i} + \frac{\ell'-\beta-x_i}{x_i+y_i+1}\right) \\
&> \frac{1}{\ell'-\beta} \cdot \frac{-x_i}{x_i+y_i} \\
&> \frac{-1}{\ell'-\beta}.
\end{aligned}$$

Because there are at most $n - \ell n/m$ voters in $N \setminus G$, we have

$$\Delta_{PAV}(N \setminus G) \geq \left(n - \frac{\ell n}{m}\right)\left(-\frac{1}{\ell'-\beta}\right).$$

Therefore, the total change in PAV score over all $\ell' - \beta$ potential changes is

$$\begin{aligned}
\Delta_{PAV}(G) + \Delta_{PAV}(N \setminus G) &\geq \frac{\ell n}{m} \cdot \frac{1}{\ell} + \left(n - \frac{\ell n}{m}\right)\left(-\frac{1}{\ell'-\beta}\right) \\
&= \frac{n}{m} - \frac{nm - \ell n}{m(\ell'-\beta)} \\
&= \frac{n\ell' - n\beta}{m(\ell'-\beta)} - \frac{nm - \ell n}{m(\ell'-\beta)} \\
&= \frac{n(\ell + \ell' - \beta - m)}{m(\ell'-\beta)} \\
&> 0,
\end{aligned}$$

where the last step follows because $\ell + \ell' > \beta + m$ and $\ell' > \beta$. Therefore, we have shown the existence of W' with higher PAV score, which is a contradiction.

□

Lemma 3.6. *Seq-Phrágmen satisfies PJR.*

Proof. We consider any arbitrary ℓ-cohesive group G and prove that, after placing $m - k$ candidates, G must be at least $(\ell - k)$-represented. Therefore, after placing all m candidates, G must be ℓ-represented.

We proceed by induction. As our base case, after $m - \ell$ placements, G must be at least $\ell - \ell = 0$ represented, which vacuously holds true.

For the induction step, assume that after $m - k - 1$ placements, G is $\ell - k - 1$ satisfied. Call the candidate that is placed on the $(m-k)^{th}$ step c. If G does not agree on c, then no matter how c is placed, G will receive $\ell - k$ representation after $m - k$ placements. We now consider the case where G agrees on c. WLOG, assume that every member in G approves c.

Because G is $\ell - k - 1$ satisfied, the total load on G is at most $\ell - k - 1$. This means that $N \setminus G$, which is of size at most $n - \ell n/m$ voters, has total load at least $(m-k-1)-(\ell-k-1) = m-\ell$. Therefore, there exists a voter in $N \setminus G$ with load at least $\frac{m-\ell}{n-\ell n/m} = m/n$.

Now, there are two subcases: (a) everyone in G has load at most n/m; and (b) there exists a voter in G with load strictly greater than n/m.

In subcase (a), it is possible to approve c without increasing the maximal voter load. This is because the total load on the ℓ-cohesive group will be at most $\ell - k$, and the average voter load among members of G is at most

$$\frac{\ell - k}{\ell n/m} \leq \frac{\ell}{\ell n/m} = \frac{m}{n}.$$

This means that it is possible to redistribute the load such that no one in G has weight more than m/n, and therefore, seq-Phragmén will approve c, leading to $(\ell - k)$-representation for G after $m - k$ placements.

In subcase (b), return to the first step at which any voter in G's load exceeded m/n. Call this voter v_i. Now, replace that allocation by approving c instead. After approving c, the total load on G is at most $\ell - k$, and therefore the average voter load among members of G at this point in time is at most

$$\frac{\ell - k}{\ell n/m} \leq \frac{\ell}{\ell n/m} = \frac{m}{n}.$$

This means that it is possible to redistribute the load such that no voter in G has weight more than m/n, a contradiction to the order in which seq-Phragmén decides how to place candidates. Therefore, this subcase cannot occur. □

Lemma 3.7. *Rule X satisfies PJR but not EJR.*

Proof. **Rule X satisfies PJR:** Consider any ℓ-large and ℓ-cohesive group G of size $\ell n/m$. Let G agree on $\ell' \geq \ell$ candidates C_G. WLOG, let G approve all candidates in C_G.

We claim that, by the end of the algorithm, G will have purchased $\ell - (m - \ell')$ candidates in C_G. Because G derives representation from any candidate in $C \setminus C_G$, this suffices to satisfy PJR.

We now case on the number of candidates in $C \setminus C_G$ that are placed by Rule X.

If no candidates in $C \setminus C_G$ are placed, then, at best, $N \setminus G$ can only afford to exclude $m - \ell$ candidates in C_G before they run out of money, leaving $\ell' - (m - \ell) = \ell - (m - \ell')$ candidates for G to include. Because nothing in $C \setminus C_G$ was purchased, and all elements in C_G that were purchased were excluded, each voter in G still has her entire budget left, and collectively G can afford to place $\ell > \ell - (m - \ell')$ candidates. Therefore, G can purchase the $\ell - (m - \ell')$ remaining candidates in C_G.

Now, consider the case where some candidates in $C \setminus C_G$ are placed. Note that when candidates in $C \setminus C_G$ are placed, they must cost at most $1/\ell$ per voter because otherwise Rule X will approve a candidate in C_G, which costs each member of G at most $(n/m)/(\ell n/m) = 1/\ell$. Therefore, for each of the $m - \ell'$ elements in $C \setminus C_G$, each voter v_i in G spends at most $1/\ell$, and therefore retains at least $1 - (1/\ell)(m - \ell') = 1 - m/\ell + \ell'/\ell$ budget. Furthermore, note that, again, $N \setminus G$ can only afford to exclude $m - \ell$ candidates in C_G, leaving $\ell - (m - \ell')$ candidates in C_G for G to include. However, note that it costs $\frac{\ell - (m - \ell') \cdot (n/m)}{\ell n/m} = (1/\ell)(\ell - (m - \ell')) = 1 - m/\ell + \ell'/\ell$ per voter in G to place $\ell - (m - \ell')$ candidates in C_G and that each voter $v \in G$ has at least this amount of money left. This ensures that they will be able to purchase $\ell - (m - \ell')$ candidates in C_G, and therefore Rule X satisfies PJR.

Rule X does not satisfy EJR: Consider the following profile with $n = 12$ and $m = 6$ with the following preferences.

$$A_1 = \ldots = A_4 = \{c_1, c_2, c_5\}$$
$$A_5 = \ldots = A_8 = \{c_1, c_2, c_6\}$$
$$A_9 = A_{10} = \{c_3, c_4, c_5\}$$
$$A_{11} = A_{12} = \{c_3, c_4, c_6\}.$$

The first four actions of Rule X will be to include c_1 and c_2 and exclude c_3 and c_4. After these placements, voters $v_1, \ldots, v_8$ will have no money left, and voters $v_9, \ldots, v_{12}$ will have their entire budgets left. Now, Rule X will perfectly spend the money of $v_9, \ldots, v_{12}$ over candidates c_5 and c_6 (either including both or neither) and the coalition of voters $v_9, \ldots, v_{12}$ will not satisfy EJR because each voter in this coalition is represented exactly once, even though someone in this coalition deserves two representatives. □

Notably, in the VNW setting, JR and PJR are less compelling notions of representation than in the FNW setting. In particular, whenever an ℓ-cohesive group of voters does not agree on the placement of a particular candidate, PJR automatically counts that candidate for free toward the group's representation quota, since at least one member of the group agrees with the candidate's placement. In other words, any disagreement within an ℓ-cohesive group results in partial representation, no matter the outcome of the election. This is particularly problematic for JR: any 1-large, 1-cohesive group of voters that disagrees on even a single candidate will never be witness to a violation of JR.

Lemma 3.7 is also notable because Rule X satisfies EJR for FNW elections, but the straightforward extension of this rule does not satisfy EJR for VNW elections, demonstrating a qualitative difference between proportionality properties in the FNW and VNW settings. It is still an open question whether or not PAV and seq-Phragmén satisfy EJR for VNW elections; we leave this for future work. Additionally, we note that proofs of (E/P)JR in the VNW setting are more complicated than in the FNW setting. In particular, the proofs from the FNW setting do not easily carry over to the VNW setting because of the different definitions of voter utility—it is more difficult to argue about agreements with the placement of candidates than to argue about the number of approved candidates on the winning committee.

4 DETERMINISTIC RULES

We begin by showing an upper bound on the level of average satisfaction that deterministic rules can provide.

Theorem 4.1. *No deterministic rule satisfies* $(\frac{m-1}{m} + \epsilon)$*-AS for any* m *and any* $\epsilon > 0$.

Proof. First, suppose that m is odd. Then set $n = 2$, with $A_1 = \{c_1, \ldots, c_m\}$ and $A_2 = \emptyset$. Without loss of generality, suppose that the output W is such that $|W| > \frac{m}{2}$. But then voter v_2 is an $\frac{m}{2}$-large, $\frac{m}{2}$-cohesive group with average satisfaction at most $\frac{m-1}{2}$, which yields an $(\frac{m-1}{m})$-AS approximation.

Next, suppose m is even, and set $n = 4m$. Consider the profile

$$A_1 = \ldots = A_m = \{c_1, \ldots, c_m\}$$
$$A_{m+1} = \ldots = A_{2m} = \{c_1, \ldots, c_{m-1}\}$$
$$A_{2m+1} = \ldots = A_{3m} = \{c_m\}$$
$$A_{3m+1} = \ldots = A_{4m} = \{\}.$$

We consider two cases. In the first case, suppose that the output W has $|W| \geq \frac{m}{2} + 1$. Consider the $\frac{m}{2}$-large, $\frac{m}{2}$-cohesive group of voters $V = \{v_{2m+1}, \ldots, v_{4m}\}$. We have

$$avs_W(V) \leq \frac{1}{2m}(m(m - |W|) + m(m - |W| + 1)) \leq \frac{m-1}{2}$$

which yields an $(\frac{m-1}{m})$-AS approximation.

In the second case, suppose that the output W has $|W| = \frac{m}{2}$. Suppose without loss of generality that $c_m \notin W$. Then again consider $V = \{v_{2m+1}, \ldots, v_{4m}\}$. We have

$$avs_W(V) \leq \frac{1}{2m}\left(m\left(\frac{m}{2} - 1\right) + m\left(\frac{m}{2}\right)\right) = \frac{m-1}{2}$$

again yielding an $(\frac{m-1}{m})$-AS approximation. This completes the proof. □

Theorem 4.1 leaves open the possibility that there exists a deterministic rule that provides quite good average satisfaction guarantees when the number of candidates is large. Finding such a rule or lowering the upper bound is an interesting open question. However, we show that none of the natural adaptations of FNW rules that we consider is able to guarantee better than a 0.5 approximation to AS even when m is large.

Theorem 4.2. *PAV does not satisfy a* $0.5 + \epsilon$ *approximation to AS, for any* $\epsilon > 0$ *for* $m \geq 2$.

Proof. Consider a profile with $n = 2m$ voters with preferences

$$A_1 = \ldots = A_{m-1} = \{c_1, \ldots, c_m\}$$
$$A_m = \ldots = A_{2m-2} = \{c_1, \ldots, c_{m-1}\}$$
$$A_{2m-1} = \{c_m\}$$
$$A_{2m} = \{\}.$$

This profile is symmetric in c_m, so without loss of generality suppose that c_m is included. When all candidates are included, the PAV score of the committee is

$$1 + (m-1)H_{m-1} + (m-1)H_m. \quad (1)$$

Suppose instead that some $k-1 < m-1$ of the candidates $c_1, \ldots, c_{m-1}$ are included. The PAV score of the committee is

$$H_{m-k} + H_{m-k+1} + (m-1)H_{k-1} + (m-1)H_k. \quad (2)$$

Subtracting Equation 2 from Equation 1 yields at least

$$\begin{aligned} &\geq \frac{(m-1)(m-k)}{m-1} - H_{m-k} + \frac{(m-1)(m-k)}{m} + 1 - H_{m-k+1} \\ &\geq m-k-(m-k) + \frac{m-k}{2} - \frac{1}{2} - \ldots - \frac{1}{m-k+1} \\ &\geq 0, \end{aligned}$$

where the first inequality holds because $H_{m-k} \leq m-k$ and $\frac{m-1}{m} \geq \frac{1}{2}$ for $m \geq 2$.

Therefore, the highest PAV score is achieved when all candidates $c_1, \ldots, c_{m-1}$ are included. But then the group $N^* = \{v_{2m-1}, v_{2m}\}$ is 1-large and 1-cohesive but is only satisfied 0.5 times on average. □

Theorem 4.3. *seq-Phragmén does not satisfy a* $0.5 + \epsilon$ *approximation to AS, for any* $\epsilon > 0$ *for* $m \geq 2$.

Proof. Consider the same profile as in the proof of Theorem 4.2. It is easy to check that seq-Phragmén begins by including candidates $c_1, \ldots, c_{m-2}$, after which each voter of the first and second type has load $\frac{m-2}{2(m-1)}$. In the $(m-1)$-th round, the algorithm has four choices: to include or exclude c_{m-1}, or to include or exclude c_m.

Including c_{m-1} results in a load of $\frac{m-1}{2(m-1)} = \frac{1}{2}$ on voters $v_1, \ldots, v_{2m-2}$. Excluding c_{m-1} results in a load of $\frac{1}{2}$ to voters v_{2m-1} and v_{2m}. Including c_m (which is symmetric to excluding c_m) results in a load x to voters $v_1, \ldots, v_{m-1}, v_{2m-1}$, where x is the solution to

$$mx - (m-1)\frac{m-2}{2(m-1)} = 1,$$

which yields a solution of $x = \frac{1}{2}$.

The algorithm is therefore indifferent between all possible actions; breaking ties adversarially yields the inclusion of c_{m-1}. Regardless of the inclusion or exclusion of candidate c_m, the group $N^* = \{v_{2m-1}, v_{2m}\}$ is 1-large and 1-cohesive but is only satisfied 0.5 times on average. □

We note that the dependence on tiebreaking in the proof of Theorem 4.3 can be removed by taking multiple copies of the profile used in the proof and changing the preference of a single voter.

Theorem 4.4. *Rule X does not satisfy a* $0.5 + \epsilon$ *approximation to AS, for any* $\epsilon > 0$ *for* $m \geq 3$.[3]

Proof. Consider the same profile used in the proof of Theorem 4.2. Rule X begins by including each of candidates $c_1, \ldots, c_{m-1}$. Each of these candidates costs $\frac{n}{m(2m-2)} = \frac{1}{m-1}$ for each voter $v_1, \ldots, v_{2m-2}$. In comparison, placing the last candidate at any point costs $n/2$ voters $\frac{n/m}{n/2} = \frac{2}{m}$, which is a greater cost than $\frac{1}{m-1}$ when $m \geq 3$. Including each of $c_1, \ldots, c_{m-1}$ therefore costs $v_1, \ldots, v_{2m-2}$ one dollar each. Regardless of the placement of c_m, the 1-large and 1-cohesive group of voters $N^* = \{v_{2m-1}, v_{2m}\}$ is satisfied only 0.5 times on average. □

5 RANDOMIZED RULES

We now turn our attention to randomized rules in order to achieve better average satisfaction guarantees. A randomized rule is one that outputs a distribution over committees rather than a single committee, and our approximation guarantee will hold in expectation over the possible committees. We consider a simple and natural randomized rule that, for each candidate c_j, includes c_j in the set of winners W with probability equal to the fraction of the voters who approve j.

Definition 5.1. Given a preference profile $\mathbf{A}$, the *Proportional Random Rule (PRR)* independently adds each $c_j \in C$ to the winning committee W with probability

$$p_j = \frac{\left|\{v_i \in N \text{ s.t. } c_j \in A_i\}\right|}{n}.$$

Theorem 5.2. *PRR satisfies* $29/32$*-AS in expectation for any* $m \geq 1$.

In the proof of Theorem 5.2, it will be helpful to think about the effect that an individual candidate has on the satisfaction of a group G. For an outcome W, a group of voters G, and a candidate c_j, we say that the *contribution from* c_j *to the average satisfaction of* G is $avs_{c_j}(G) = |\{i : c_j \in A_i\}|/|G|$ if $c_j \in W$ or $avs_{c_j}(G) = |\{i : c_j \notin A_i\}|/|G|$ if $c_j \notin W$. Note that $avs_W(G) = \sum_{j=1}^{m} avs_{c_j}(G)$.

Proof. We prove the result in two steps. First, we show that when $\ell \leq m/3$, PRR achieves an average satisfaction of ℓ; second, we show that when $\ell > m/3$, PRR achieves an average satisfaction of $(29/32)\ell$.

[3]When $m = 2$, we know from Theorem 4.1 that no deterministic rule, including Rule X, can achieve better than a 0.5 approximation.

Case 1: $\ell \leq m/3$. Consider an ℓ-cohesive group, G, of size $\ell n/m$,[4] and a candidate c_j. Let $k_A = |\{v_i \in G : c_j \in A_i\}|$ denote the number of voters in G who approve c_j, and $k_D = \ell n/m - k_A$ denote the number of voters in G who disapprove c_j. Without loss of generality, let $k_A \leq k_D$. Further, suppose that x of the voters in $N \setminus G$ approve c_j and $y = n - \ell n/m - x$ voters in $N \setminus G$ disapprove c_j.

The expected contribution from c_j to the average satisfaction of G is

$$\mathbb{E}[avs_{c_j}(G)] = \frac{k_A}{|G|}\left(\frac{k_A + x}{n}\right) + \frac{k_D}{|G|}\left(\frac{k_D + y}{n}\right).$$

Because $k_A \leq k_D$ and $x + y$ is fixed, this expression is minimized when $y = 0$. We therefore have

$$\begin{aligned}\mathbb{E}[avs_{c_j}(G)] &= \frac{k_A}{|G|}\left(\frac{k_A + n - \ell n/m}{n}\right) + \frac{k_D}{|G|}\left(\frac{k_D}{n}\right) \\ &= \frac{1}{n|G|}\left(|G|^2 + k_A(n - \ell n/m - 2k_D)\right) \\ &\geq \frac{|G|}{n} = \frac{\ell}{m},\end{aligned}$$

where the inequality holds because $k_D \leq \ell n/m$ by definition, and we can assume $m \geq 3$ because ℓ must be at least 1. So, in both the case where G agrees on c_j and the case where G disagrees on c_j, the expected contribution of c_j to the average satisfaction of G is at least $\frac{\ell}{m}$. Summed over all candidates, the average satisfaction of G is at least ℓ, as required.

Case 2: $\ell > m/3$. Consider an ℓ-cohesive group, G, of size $\ell n/m$, and a candidate c_j. Let $k_A = |\{v_i \in G : c_j \in A_i\}|$ denote the number of voters in G who approve c_j, and $k_D = \ell n/m - k_A$ denote the number of voters in G who disapprove c_j. Without loss of generality, let $k_A \leq k_D$. As in the previous case, it is easy to show that the expected contribution from c_j to G's average satisfaction is minimized when all voters in $N \setminus G$ approve c_j.

We therefore have that

$$[\mathbb{E}avs_{c_j}(G)] = \frac{k_A}{|G|}\left(\frac{k_A + n - \ell n/m}{n}\right) + \frac{k_D}{|G|}\left(\frac{k_D}{n}\right).$$

Substituting $k_D = \ell n/m - k_A$, taking the derivative with respect to k_A, and setting to 0 yields

$$\frac{1}{n}\left(4k_A - 3(\ell n/m) + n\right) = 0 \implies k_A = \frac{3\ell n/m - n}{4} > 0,$$

where the inequality follows from the assumption that $\ell > m/3$. Furthermore, the second derivative with respect to k_A is $4/n > 0$, and therefore $k_A = (3\ell n/m - n)/4$ is a local minimum.

The expected contribution from c_j to G's average satisfaction can therefore be as low as

$$\begin{aligned}[\mathbb{E}avs_{c_j}(G)] &= \frac{k_A}{|G|}\left(\frac{k_A + n - \ell n/m}{n}\right) + \frac{k_D}{|G|}\left(\frac{k_D}{n}\right) \\ &= \frac{-\ell}{8m} + \frac{3}{4} - \frac{m}{8\ell}.\end{aligned}$$

We also note that, because G is ℓ-cohesive, there exist at least ℓ candidates that G agrees on. Each of these candidates has

$$avs_{c_j}(G) \geq |G|/n \geq \ell/m,$$

where the first inequality follows from G being ℓ-cohesive and the second from G being ℓ-large.

Summing over the contributions of all candidates, the average satisfaction of G is at least

$$\begin{aligned}&\ell\frac{\ell}{m} + (m - \ell)\left(\frac{3}{4} - \frac{\ell}{8m} - \frac{m}{8\ell}\right) \\ &= \left(\frac{9\ell}{8m} - \frac{m^2}{8\ell^2} - \frac{7}{8} + \frac{7m}{8\ell}\right)\ell. \qquad (3)\end{aligned}$$

Our goal is to lower bound the term in parentheses by $\frac{29}{32}$, thus providing the desired approximation guarantee. Setting $\ell = \alpha m$, where $\alpha \in (\frac{1}{3}, 1)$, and differentiating with respect to α yields

$$\frac{d}{d\alpha}\left(\frac{9\alpha}{8} - \frac{1}{8\alpha^2} - \frac{7}{8} + \frac{7}{8\alpha}\right) = \frac{9}{8} + \frac{2}{8\alpha^3} - \frac{7}{8\alpha^2}.$$

Setting equal to 0 yields

$$9\alpha^3 - 7\alpha + 2 = (1 + \alpha)(3\alpha - 2)(3\alpha - 1) = 0,$$

so the only critical point in the interval $\alpha \in (1/3, 1]$ is $\alpha = 2/3$. It is easy to check that the second derivative is positive at $\alpha = 2/3$, so average satisfaction is minimized at this point. Plugging $\ell = 2m/3$ into Equation 3 yields a 29/32 approximation to AS, as desired. □

Guided by the proof of Theorem 5.2, we show that the bound is tight.

Theorem 5.3. *PRR does not satisfy* $(29/32 + \epsilon) - AS$ *for any* $\epsilon > 0$.

Proof. Let $m = 3$ and $n = 12$. Consider the profile

$$\begin{aligned}A_1 = A_2 = A_3 = A_4 = A_5 &= \{c_1, c_2, c_3\} \\ A_6 = A_7 = A_8 &= \{c_1, c_2\} \\ A_9 = A_{10} = A_{11} = A_{12} &= \{\}.\end{aligned}$$

In particular, note that the first 8 voters form a 2-large and 2-cohesive group. Then the expected satisfaction of the first five voters is $\frac{2}{3} + \frac{2}{3} + \frac{5}{12}$ and the expected satisfaction of the next three voters is $\frac{2}{3} + \frac{2}{3} + \frac{7}{12}$. Taking the average yields $\frac{29}{16} = \frac{29}{32}\ell$ for $\ell = 2$. □

Whether there exists a randomized rule that achieves better than a 29/32-AS approximation remains an open problem.

6 DISCUSSION

We have initiated the study of representation in approval elections with a variable number of winners. We believe that this topic, and the study of VNW elections more generally, deserves further research.

Many open problems remain. In particular, we do not have matching upper and lower bounds for the average satisfaction guarantees that can be provided by deterministic and randomized rules. Determining the existence of rules that satisfy EJR is also an interesting question; while we have argued that natural extensions of JR and PJR make less sense for VNW elections than for FNW, EJR remains a compelling property.

More broadly, we have assumed voters gain utility whenever they agree with the placement of a candidate. A natural extension of this model would be one in which voters derive different levels of utility for an approved candidate being selected and a disapproved candidate being excluded. Extending our results to this setting appears nontrivial.

[4]It is sufficient to consider groups of size exactly $\ell n/m$ because if there exists an ℓ-cohesive larger group that violates the desired guarantee, there must exist a subset of size $\ell n/m$ that also violates the guarantee.

REFERENCES

[1] Haris Aziz and Shenwei Huang. A polynomial-time algorithm to achieve extended justified representation. *arXiv preprint arXiv:1703.10415*, 2017.

[2] Haris Aziz, Markus Brill, Vincent Conitzer, Edith Elkind, Rupert Freeman, and Toby Walsh. Justified representation in approval-based committee voting. *Social Choice and Welfare*, 48(2):461–485, 2017.

[3] Haris Aziz, Edith Elkind, Shenwei Huang, Martin Lackner, Luis Sánchez-Fernández, and Piotr Skowron. On the complexity of extended and proportional justified representation. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*, 2018.

[4] Florian Brandl and Dominik Peters. An axiomatic characterization of the Borda mean rule. *Social Choice and Welfare*, 52(4):685–707, 2019.

[5] Markus Brill, Rupert Freeman, Svante Janson, and Martin Lackner. Phragmén's voting methods and justified representation. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence*, 2017.

[6] Conal Duddy, Ashley Piggins, and William S Zwicker. Aggregation of binary evaluations: a Borda-like approach. *Social Choice and Welfare*, 46(2):301–333, 2016.

[7] Piotr Faliszewski, Arkadii Slinko, and Nimrod Talmon. The complexity of multiwinner voting rules with variable number of winners. *arXiv preprint arXiv:1711.06641*, 2017.

[8] Peter C Fishburn and Aleksandar Pekeč. Approval voting for committees: Threshold approaches. *Technical Report*, 2004.

[9] Svante Janson. Phragmén's and Thiele's election methods. *arXiv preprint arXiv:1611.08826*, 2016.

[10] D Marc Kilgour. Approval elections with a variable number of winners. *Theory and Decision*, 81(2):199–211, 2016.

[11] Dominik Peters and Piotr Skowron. Proportionality and the limits of welfarism. *arXiv preprint arXiv:1911.11747*, 2019.

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

[13] Luis Sánchez-Fernández, Edith Elkind, Martin Lackner, Norberto Fernández, Jesús A Fisteus, Pablo Basanta Val, and Piotr Skowron. Proportional justified representation. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence*, 2017.

[14] P Skowron, M Lackner, M Brill, D Peters, and E Elkind. Proportional rankings. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI)*, 2017.

[15] Piotr Skowron, Martin Lackner, Edith Elkind, and Luis Sánchez-Fernández. Optimal average satisfaction and extended justified representation in polynomial time. *arXiv preprint arXiv:1704.00293*, 2017.

[16] Piotr Skowron. Proportionality degree of multiwinner rules. *arXiv preprint arXiv:1810.08799*, 2018.

[17] Thorvald N Thiele. Om flerfoldsvalg. *Oversigt over det Kongelige Danske Videnskabernes Selskabs Forhandlinger*, 1895:415–441, 1895.