A Quantitative Analysis of Multi-Winner Rules

Martin Lackner, Piotr Skowron

Abstract

To choose a suitable multi-winner rule, i.e., a voting rule for selecting a subset of k alternatives based on a collection of preferences, is a hard and ambiguous task. Depending on the context, it varies widely what constitutes the choice of an "optimal" subset. In this paper, we offer a new perspective to measure the quality of such subsets and—consequently—multi-winner rules. We provide a quantitative analysis using methods from the theory of approximation algorithms and estimate how well multi-winner rules approximate two extreme objectives: diversity as captured by the (Approval) Chamberlin–Courant rule and individual excellence as captured by Multi-winner Approval Voting. With both theoretical and experimental methods we classify multi-winner rules in terms of their quantitative alignment with these two opposing objectives.

1 Introduction

A multi-winner rule is a voting method for selecting a fixed-size subset of alternatives, a so-called *committee*. More formally, it is a function that given a set of objects, preferences of a population of voters over these objects, and an integer k, returns a subset of exactly k objects. Multi-winner rules are applicable to problems from and beyond the political domain, for instance to selecting a representative body such as a parliament or university senate [13, 18], to shortlisting candidates (e.g., in a competition) [6], designing search engines [15, 37] and other recommendation systems [35], and as mechanisms for locating facilities [19].

Ideally, a multi-winner rule should select the "best" committee, but the suitability of a chosen committee strongly depends on the specific context. For instance, if voters are experts (e.g., judges in a sport competition) whose preferences reflect their estimates of the objective qualities of candidates, then the goal is typically to pick k individually best candidates, e.g., those candidates who receive the highest scores from judges. Intuitively and somehow simplified, in this and similar scenarios the quality of candidates can be assessed separately, and a suitable multi-winner rule should pick the k best-rated ones. On the contrary, if the voters are citizens and the goal is to choose locations for k public facilities (say, hospitals), then our goal is very different: assessing the candidates separately can result in building all the facilities in one densely populated area; yet, it is preferable to spread them in order to ensure that as many citizens as possible have access to *some* facility in their vicinity.

These two examples illustrate two very different goals of multi-winner rules, which can be informally described as follows [18]: *Diversity* requires that a rule should select a committee which represents as many voters as possible; this translates to choosing a hospital distribution that covers as many citizens as possible. *Individual excellence* suggests picking those candidates that individually receive the highest total support from the voters; this translates to selecting a group of best contestants in the previous example. However, many real-life scenarios do not fall clearly into one of the two categories. For example, rankings provided by a search engine should list the most relevant websites but also provide every user at least one helpful link. In such cases, a mechanism designer would be interested in choosing a rule that guarantees some degree of diversity and individual excellence at the same time, putting more emphasis on either of them depending on the particular context. Consequently, to properly match rules with specific applications, it is essential to understand to which degree committees chosen by established multi-winner rules are diverse or individually excellent. In this paper we (1) develop a set of tools that allow one to better understand the nature of multi-winner rules and to assess the tradeoffs between their diversity and individual excellence, and (2) provide a classification that clarifies the behavior of these rules with respect to the two criteria. We focus on the case where voters express their preferences by providing subsets of approved candidates (the approval-based model), yet our approach is applicable to other preference models as well.

1.1 Methodology and Contribution

In our approach we identify two multi-winner rules, the Chamberlin–Courant rule (CC) and Multi-winner Approval Voting (AV), as distinctive representatives of the principles of diversity and individual excellence, respectively. Next, we measure how close certain rules are to AV and CC—we measure this distance by using the concept of the worst-case approximation. Thus, by investigating how well certain rules approximate AV (resp. CC), we provide guarantees of how individually excellent (resp. diverse) these rules are. Such guarantees could be viewed as *quantitative properties* that measure the level of diversity and individual excellence of the studied rules. This is quite different from the traditional axiomatic approach to investigating properties of voting rules, which is qualitative: a rule can either satisfy a property or not. Our approach provides much more fine-grained information and allows us to estimate the degree to which a certain property is satisfied. With these methods, we understand voting rules as a compromise between different (often contradictory) goals.

Our main contribution lies in developing a new method for evaluating multi-winner rules. Specifically, we provide two types of analyses for a number of multi-winner rules:

(1) In Section 3, we derive theoretical upper bounds on how much an outcome of the considered multi-winner rules can differ from the outcomes of CC and AV. We call these bounds CC-guarantee and AV-guarantee. These can be interpreted as worst-case (over all possible preference profiles) guarantees for diversity and individual excellence. Our guarantees are given as functions of the committee size k and return values between 0 and 1. Intuitively, a higher CC-guarantee (resp. AV-guarantee) indicates a better performance in terms of diversity (resp. individual excellence), where 1 denotes that the rule performs as good as CC (resp., AV). Table 1 summarizes our results. We also prove bounds on how well *proportional* rules can approximate AV and CC.

(2) In Section 4, we complement the worst-case analysis from Section 3 with an experimental study yielding approximation ratios for actual data sets. In extensive experiments we estimate how on average the outcomes of the considered rules differ from the outcomes of CC and AV.

In Section 5, we complement our results with an analysis of the axiom of efficiency, which can be viewed as an incarnation of Pareto efficiency, in the context of multi-winner elections. We say that a committee W_1 dominates a committee W_2 if each voter approves as many members of W_1 as of W_2 and some voter approves strictly more members of W_1 than of W_2. Efficiency says that a rule should never select a dominated committee; thus efficiency could be viewed as a basic axiom for individual excellence. Since efficiency appears to be very fundamental, it may come as a surprise that many known rules (in particular, the Monroe rule and all sequential rules) do not satisfy this property. The result of this analysis is also summarized in Table 1.

Our most important findings can be summarized as follows. Proportional Approval Voting (PAV) achieves the best compromise between AV and CC; this can be observed both from theoretical and experimental results. The sequential rules seq-PAV and Phragmén's rule, however, achieve almost the same quality while being polynomial-time computable (in contrast to PAV, which is computationally intractable [4, 35]). Also the 2-Geometric rule achieves a very good compromise, but is slightly leaning towards diversity. More generally, we show that the p-Geometric rule spans the whole spectrum from AV to CC, controlled through the parameter p. Hence, by adjusting the parameter p, one can obtain any desired compromise between AV and CC.

1.2 Related Work

The normative study of multi-winner election rules typically focuses on axiomatic analysis. For approval-based rules a number of axioms describing proportionality have been recently identified and

	AV-guarantee		CC-guarantee		effic.
	lower	upper	lower	upper	
AV	1	1	$\frac{1}{k}$	$\frac{1}{k}$	✓
CC	$\frac{1}{k}$	$\frac{1}{k}$	1	1	✓
seq-CC	$\frac{1}{k}$	$\frac{1}{k}$	$1 - 1/e$	$1 - (1 - 1/k)^k$	✗
PAV	$\frac{1}{2+\sqrt{k}}$	$\frac{2}{\lfloor\sqrt{k}\rfloor} - \frac{1}{k}$	$\frac{1}{2}$	$\frac{1}{2} + \frac{1}{4k-2}$	✓
p-Geometric	$\frac{W(k\log(p))}{k\log(p)+W(k\log(p))}$	$\frac{1}{k} + \frac{2W(k\log(p))}{k\log(p)}$	$\frac{p-1}{p}$	$\frac{p}{p+\frac{k}{k+2}}$	✓
seq-PAV	$\frac{1}{2\sqrt{k}}$	$\frac{2}{\lfloor\sqrt{k}\rfloor} - \frac{1}{k}$	$\frac{1}{\log(k)+2}$	$\frac{1}{2} + \frac{1}{4k-2}$	✗
α-Monroe	$\frac{1}{k}$	$\frac{1}{k}$	$\frac{1}{2}$	$\frac{1}{2} + \frac{1}{k-1}$	✗
Greedy α-Monr.	$\frac{1}{k}$	$\frac{1}{k}$	$\frac{1}{2}$	$\frac{1}{2} + \frac{1}{k-1}$	✗
seq-Phragmén	$\frac{1}{5\sqrt{k}+1}$	$\frac{2}{\lfloor\sqrt{k}\rfloor} - \frac{1}{k}$	$\frac{1}{2}$	$\frac{1}{2} + \frac{1}{4k-2}$	✗

Table 1: Summary of worst-case guarantees for the considered multi-winner rules. The guarantees are functions of the committee size k. A higher value means a better guarantee, with 1 denoting the optimal performance. In most cases we could only find (accurate) estimates instead of the exact values of the guarantees: the "lower" and "upper" values in the table denote that the respective guarantee is between these two values. The formulas for the guarantees of the p-Geometric rule are depicted in Figure 1 (page 9). The column "efficiency" indicates whether the rule satisfies the efficiency axiom as discussed in Section 5.

explored, in particular in the context of the rules that we study in this paper [1, 3, 10, 22, 32, 33, 37]. Similar properties for the ordinal model have been discussed by Dummett [14], Elkind et al. [17], Aziz et al. [2] and in the original works by Monroe [26] and Chamberlin and Courant [13]; and for the model with weak preferences by Baumeister et al. [7]. For a survey on properties of multi-winner rules, with the focus on the ideas of individual excellence, diversity, and proportionality, we refer the reader to the book chapter by Faliszewski et al. [18].

Another approach to understanding the nature of different multi-winner rules is to analyze how these rules behave on certain subdomains of preferences, where their behavior is much easier to interpret, e.g., on two-dimensional geometric preferences [16], on party-list profiles [11], or on single-peaked and single-crossing domains [2]. Other approaches include analyzing certain aspects of multi-winner rules in specifically-designed probabilistic models [21, 23, 30, 34], quantifying regret and distortion in utilitarian models [12], assessing their robustness [9], and evaluating them based on data collected from surveys [31, 39].

2 Preliminaries

For each $t \in \mathbb{N}$, we let $[t] = \{1, \dots, t\}$. For a set X, we write $\mathcal{S}(X)$ to denote the powerset of X, i.e., the set of all subsets of X. By $\mathcal{S}_k(X)$ we denote the set of all k-element subsets of X.

Let $C = \{c_1, \dots, c_m\}$ and $N = \{1, \dots, n\}$ be sets of m *candidates* and n *voters*, respectively. Voters reveal their preferences by indicating which candidates they like: by $A(i) \subseteq C$ we denote the *approval set* of voter i (that is, the set of candidates that i approves of). For a candidate $c \in C$, by $N(c) \subseteq N$ we denote the set of voters who approve c. Given a set of candidates $X \subseteq C$, we write $N(X)$ to denote the set of voters who approve at least one candidate in X, that is $N(X) = \{i \in N :$

$X \cap A(i) \neq \emptyset\}$. We call the collection of approval sets $A = (A(1), A(2), \dots, A(n))$, one per each voter, an *approval profile*. We use the symbol $\mathcal{A}$ to represent the set of all possible approval profiles.

We call the elements of $\mathcal{S}_k(C)$ size-k *committees*. Hereinafter, we will always use the symbol k to represent the desired size of the committee to be elected. An *approval-based committee rule* (in short, an ABC rule) is a function $\mathcal{R}\colon \mathcal{A} \times \mathbb{N} \to \mathcal{S}(\mathcal{S}_k(C))$ that takes as an input an approval profile and an integer $k \in \mathbb{N}$ (the required committee size), and returns a set of size-k committees.[1] Below, we recall the definitions of ABC rules which are the objects of our study.

Multi-winner Approval Voting (AV). This rule selects k candidates which are approved by most voters. More formally, for a profile A the AV-score of committee W is defined as $\mathrm{sc}_{\mathrm{av}}(A, W) = \sum_{c \in W} |N(c)|$, and AV selects committees W that maximize $\mathrm{sc}_{\mathrm{av}}(A, W)$.

Approval Chamberlin–Courant (CC). For a profile A we define the CC-score of a committee W as $\mathrm{sc}_{\mathrm{cc}}(A, W) = \sum_{i \in N} \min\big(1, |A(i) \cap W|\big) = |N(W)|$; CC outputs $\mathrm{argmax}_W \mathrm{sc}_{\mathrm{cc}}(A, W)$. In words, CC aims at finding a committee W such that as many voters as possible have their representatives in W (a representative of a voter is a candidate she approves of). The CC rule was first mentioned by Thiele [38], and then introduced in a more general context by Chamberlin and Courant [13].

Proportional Approval Voting (PAV). This rule selects committees with the highest PAV-scores, defined as $\mathrm{sc}_{\mathrm{pav}}(A, W) = \sum_{i \in N} \mathrm{H}\left(|W \cap A(i)|\right)$, where $\mathrm{H}(t)$ is the t-th harmonic number, i.e., $\mathrm{H}(t) = \sum_{i=1}^{t} 1/i$. By using the harmonic function $\mathrm{H}(\cdot)$, voters who already have more representatives in the committee get less voting power than those with fewer representatives. While using other concave functions instead of $\mathrm{H}(\cdot)$ would give similar effects, the harmonic function is particularly well justified—it implies a number of appealing properties of the rule [1], and it allows one to view PAV as an extension of the famous d'Hondt method [11, 22].

p-Geometric. This rule, introduced by Skowron et al. [35], can be described similarly to PAV. The difference is that it uses an exponentially decreasing function instead of the harmonic function to describe the relation between the voting power of individual voters and the number of their approved representatives in the committee. Formally, for a given parameter $p \geq 1$ the p-geometric rule assigns to each committee W the score $\mathrm{sc}_{p\text{-geom}}(A, W) = \sum_{i \in N} \sum_{j=1}^{|A(i) \cap W|} \frac{1}{p^j}$, and picks the committees with the highest scores. It is easy to see that the 1-geometric rule is simply AV.

Sequential CC/AV/PAV/p-Geometric. For each rule $\mathcal{R} \in \{\mathrm{CC}, \mathrm{AV}, \mathrm{PAV}, p\text{-geometric}\}$, we define its sequential variant, denoted as seq-$\mathcal{R}$, as follows. We start with an empty solution $W = \emptyset$ and in each of the k consecutive steps we add to W a candidate c that maximizes $\mathrm{sc}_{\mathcal{R}}(A, W \cup \{c\})$, i.e., the candidate that improves the committee's score most. We break ties lexicographicly.

Monroe. Monroe's rule [26], similarly to CC, aims at maximizing the number of voters who are represented in the elected committee. The difference is that for calculating the score of a committee, Monroe additionally imposes that each candidate should be responsible for representing roughly the same number of voters. Formally, a Monroe assignment of the voters to a committee W is a function $\phi\colon N \to W$ such that each candidate $c \in W$ is assigned roughly the same number of voters, i.e., that $\lfloor n/k \rfloor \leq |\phi^{-1}(c)| \leq \lceil n/k \rceil$. Let $\Phi(W)$ be the set of all possible Monroe assignments to W. The Monroe-score of W is defined as $\mathrm{sc}_{\mathrm{Monroe}}(A, W) = \max_{\phi \in \Phi(W)} \sum_{i \in N} |A(i) \cap \{\phi(i)\}|$; the rule returns $\mathrm{argmax}_W \mathrm{sc}_{\mathrm{Monroe}}(A, W)$.

Greedy Monroe [36]. This is a sequential variant of the Monroe's rule. It proceeds in k steps: In each step it selects a candidate c and a group G of $\lfloor n/k \rfloor$ or $\lceil n/k \rceil$ not-yet removed voters[2] so that

[1] Rules which for some profiles return multiple committees as tied winners are often called irresolute. In practice, one usually uses some tie-breaking mechanism to single out a winning committee.

[2] To be precise, for $n = k \cdot \lfloor n/k \rfloor + c$, the first c groups of voters to be removed have size $\lceil n/k \rceil$ and the remaining $k - c$ have size $\lfloor n/k \rfloor$.

$|N(c) \cap G|$ is maximal; next candidate c is added to the winning committee and the voters from G are removed from the further consideration.

Phragmén's Sequential Rule (seq-Phragmén). Perhaps the easiest way to define the family of Phragmén's rules [10, 20, 27–29] is by describing them as load distribution procedures. We assume that each selected committee member c is associated with one unit of load that needs to be distributed among those voters who approve c (though it does not have to be distributed equally). Seq-Phragmén proceeds in k steps. In each step it selects one candidate and distributes its load as follows: let $\ell_j(i-1)$ denote the total load assigned to voter j just before the i-th step. In the i-th step the rule selects a candidate c and finds a load distribution $\{x_j : j \in N\}$ that satisfies the following three conditions: (1) $x_j > 0$ implies that $c \in A(j)$, (2) $\sum_{j \in N} x_j = 1$ (3) the maximum load assigned to a voter, $\max_{j \in N}(\ell_j(i-1) + x_j)$, is minimized. The new total load assigned to a voter $j \in N$ after the i-th step is $\ell_j(i) = \ell_j(i-1) + x_j$.

3 Worst-Case Guarantees of Multi-winner Rules

The Chamberlin–Courant Rule and Approval Voting represent two extreme points in the spectrum of multi-winner rules [11, 16, 18, 22]. Specifically, CC and AV are prime examples of rules aiming at *diversity*, and at *individual excellence*, respectively. For a detailed discussion on these two principles we refer the reader to the book chapter of Faliszewski et al. [18], but below we also include a simple example which illustrates the difference between AV and CC. In short, AV cares about selecting candidates who receive the highest total support from the population of voters, and CC cares mostly about representing the minorities in the elected committee.

Example 1. *Consider a profile where 30 voters approve candidates $\{c_1, c_2, c_3\}$, 20 voters approve $\{c_4, c_5, c_6\}$, and 5 voters approve $\{c_7, c_8, c_9\}$. Let $k = 3$. For this profile AV selects candidates $\{c_1, c_2, c_3\}$, while CC selects the committee $\{c_1, c_4, c_7\}$ (among others).*

In this section we analyze the multi-winner rules from Section 2 with respect to how well they perform in terms of *diversity*, and *individual excellence*. In our study we use the established idea of approximation from computer science, but in a novel way: by estimating how well a given rule $\mathcal{R}$ approximates CC (resp., AV), we quantify how $\mathcal{R}$ performs with respect to diversity (resp., individual excellence). This differs from the typical use of the idea of approximation in the following aspects: (1) We do not seek new algorithms approximating a given objective function as well as possible, but rather analyze how well the existing known rules approximate given functions (even if it is apparent that better and simpler approximation algorithms exist, these algorithms might not share other important properties of the considered rules). (2) We are not approximating computationally hard rules with rules easier to compute. On contrary, we will be investigating how computationally hard rules (such as PAV, Monroe, etc.) approximate AV, which is easy to compute.

Definition 1. *Recall that for a profile A, $\mathrm{sc}_{\mathrm{av}}(A, W)$ and $\mathrm{sc}_{\mathrm{cc}}(A, W)$ denote the AV-score and CC-score of committee W, respectively. The* AV-guarantee *of an ABC rule $\mathcal{R}$ is a function $\kappa_{\mathrm{av}} : \mathbb{N} \to [0, 1]$ that takes as input an integer k, representing the size of the committee, and is defined as:*

$$\kappa_{\mathrm{av}}(k) = \inf_{A \in \mathcal{A}} \frac{\min_{W \in \mathcal{R}(A,k)} \mathrm{sc}_{\mathrm{av}}(A, W)}{\max_{W \in \mathcal{S}_k(C)} \mathrm{sc}_{\mathrm{av}}(A, W)}.$$

Analogously, the CC-guarantee *of $\mathcal{R}$ is defined by*

$$\kappa_{\mathrm{cc}}(k) = \inf_{A \in \mathcal{A}} \frac{\min_{W \in \mathcal{R}(A,k)} \mathrm{sc}_{\mathrm{cc}}(A, W)}{\max_{W \in \mathcal{S}_k(C)} \mathrm{sc}_{\mathrm{cc}}(A, W)}.$$

The AV and CC-guarantees can be viewed as quantitative properties of multi-winner rules. In comparison with the traditional qualitative approach (analyzing properties which can be either satisfied or not), a quantitative analysis provides much more fine-grained information regarding the behavior of a rule with respect to some normative criterion. In the remaining part of this section we evaluate the previously defined rules against their AV- and CC-guarantees.

3.1 Guarantees for CC and AV

Clearly, the AV-guarantee of Approval Voting and the CC-guarantee of the Chamberlin–Courant rule are the constant-one function. Below we establish the AV-guarantee of CC and vice versa.

Proposition 1. *The CC-guarantee of AV is $1/k$.*

Proof. Consider an approval profile A, and let W_{av} be an AV-winning committee for A. We know that W_{av} contains a candidate who is approved by most voters—let us call such a candidate $c_{\max}$. Clearly, it holds that $\mathrm{sc}_{\mathrm{cc}}(A, W_{\mathrm{av}}) \geq |N(c_{\max})|$. Further, for any size-$k$ committee $W \subseteq C$ we have that $\mathrm{sc}_{\mathrm{cc}}(A, W) \leq k|N(c_{\max})|$, which proves that the AV-guarantee of CC is at least $1/k$. To see that the guarantee cannot be higher than $1/k$ consider a family of profiles where the set of voters can be divided into k disjoint groups: $N_1, N_2, \ldots, N_k$, with $|N_1| = x + 1$ and $|N_i| = x$ for $i \geq 2$, for some large value x. Assume that $m = k^2$ and that all voters from N_i approve candidates $c_{(i-1)k+1}, c_{(i-1)k+2}, \ldots c_{ik}$. For this profile AV selects committee $\{c_1, \ldots c_k\}$ with the CC-score equal to $x + 1$. The optimal CC committee is e.g., $\{c_1, c_{k+1}, \ldots, c_{k(k-1)+1}\}$, with the CC-score equal to $kx + 1$. ☐

Proposition 2. *The AV-guarantee of CC and sequential CC is $1/k$.*

Proposition 1 and Proposition 2 give a baseline for our further analysis. In particular, we would expect that "good" rules implementing a tradeoff between diversity and individual excellence, should have AV and CC-guarantees better than $1/k$.

We conclude this section by noting that the CC-guarantee of the sequential Chamberlin–Courant rule is $1 - (1 - 1/k)^k$ (which approaches $1 - 1/e \approx 0.63$ for large k). This is the result of the fact that sequential CC is a $(1 - (1 - 1/k)^k)$-approximation algorithm for CC [24].

3.2 An Optimal Proportional Compromise

Next, we examine what are the possible AV- and CC-guarantees that a proportional rule could achieve. We consider a very weak definition of proportionality, called lower quota. This axiom is widely used [5] in the context of apportionment methods (which are special cases of approval-based multi-winner rules) and is strictly weaker than proportionality axioms typically used in the context of approval-based multiwinner rules (such as extended and proportional justified representation [1, 32]).

Definition 2. *We call a profile A a* party-list profile *if for each pair of voters $i, j \in N$ it holds that either $A(i) \cap A(j) = \emptyset$ or that $A(i) = A(j)$. For a given committee size k we say that a group of voters $V \subseteq N$ is ℓ-*cohesive*, if $|V| \geq \frac{n\ell}{k}$ and $\left|\bigcap_{i \in V} A(i)\right| \geq \ell$.*

An ABC rule $\mathcal{R}$ satisfies lower quota *if for each party-list profile A, each $k \in \mathbb{N}$ and each ℓ-cohesive group of voters $V \subseteq N$ it holds that at least ℓ members of each winning committee from $\mathcal{R}(A, k)$ are approved by the members of V.*

We obtain the following two upper bounds on the guarantees of proportional rules.

Proposition 3. *The AV-guarantee of a rule that satisfies lower quota is at most $\frac{2}{\lfloor\sqrt{k}\rfloor} - \frac{1}{k}$.*

Proof. Let us fix k, and consider the following approval-based profile A with $n = k \cdot x$ voters divided into k equal-size groups: $N = N_1 \cup \ldots \cup N_k$, with $|N_i| = x$ for each $i \in [k]$. All the voters from the first $\lfloor \sqrt{k} \rfloor$ groups approve k candidates denoted as $x_1, \ldots, x_k$. For each $i > \lfloor \sqrt{k} \rfloor$ all the voters from N_i approve a single candidate y_i.

Let $\mathcal{R}$ be a rule that satisfies lower quota. Let W and W_{av} denote the committees returned by $\mathcal{R}$ and by AV, respectively. Lower quota ensures that $y_i \in W$ for each $i > \lfloor \sqrt{k} \rfloor$. Thus,

$$\mathrm{sc}_{\mathrm{av}}(A, W) = \left(k - \lfloor \sqrt{k} \rfloor\right) x + \lfloor \sqrt{k} \rfloor \cdot x \cdot \lfloor \sqrt{k} \rfloor \leq 2kx - \lfloor \sqrt{k} \rfloor x.$$

On the other hand, one can observe that $W_{\mathrm{av}} = \{x_1, \ldots, x_k\}$, and so $\mathrm{sc}_{\mathrm{av}}(A, W_{\mathrm{av}}) = \lfloor \sqrt{k} \rfloor \cdot x \cdot k$. As a result we have:

$$\frac{\mathrm{sc}_{\mathrm{av}}(A, W)}{\mathrm{sc}_{\mathrm{av}}(A, W_{\mathrm{av}})} \leq \frac{2kx - \lfloor \sqrt{k} \rfloor x}{\lfloor \sqrt{k} \rfloor \cdot x \cdot k} = \frac{2}{\lfloor \sqrt{k} \rfloor} - \frac{1}{k}.$$

□

Proposition 4. *The CC-guarantee of a rule that satisfies lower quota is at most* $\frac{3}{4} + \frac{3}{8k-4}$.

3.3 Guarantees for Monroe and Greedy Monroe

Let us turn our attention to the Monroe rule and its greedy variant. Since Monroe is often considered a proportional rule, as it satisfies proportionality axioms such as proportional justified representation [32]. Hence, one could expect that in terms of AV and CC-guarantees this rule is between AV and CC. Surprisingly, this is not the case and in fact it does not offer a better AV-guarantee than CC.

Proposition 5. *The AV-guarantee of Greedy Monroe and Monroe is* $1/k$.

Proposition 6. *The CC-guarantee of Monroe and greedy Monroe is between* $\frac{1}{2}$ *and* $\frac{1}{2} + \frac{1}{k-1}$.

3.4 Guarantees for PAV

Let us now move to multi-winner voting systems offering asymptotically better guarantees than the (greedy) Monroe rule. As we will see, the examination of such rules requires a more complex combinatorial analysis. We start with Proportional Approval Voting.

Theorem 1. *The AV-guarantee of PAV is between* $\frac{1}{2+\sqrt{k}}$ *and* $\frac{2}{\sqrt{k}}$.

Proof. First, we show that the AV-guarantee of PAV is at least equal to $\frac{1}{2+\sqrt{k}}$. Consider an approval profile A and a PAV-winning committee W_{pav}; let $n_{\mathrm{pav}} = |N(W_{\mathrm{pav}})|$ denote the number of voters who approve some member of W_{pav}. For each $i \in N$ we set $w_i = |A(i) \cap W_{\mathrm{pav}}|$. Let W_{av} be a committee with the highest AV-score. W.l.o.g., we can assume that $W_{\mathrm{av}} \neq W_{\mathrm{pav}}$. Now, consider a candidate $c \in W_{\mathrm{av}} \setminus W_{\mathrm{pav}}$ with the highest AV-score, and let $n_c = |N(c)|$ denote the number of voters who approve c. If we replace a candidate $c' \in W_{\mathrm{pav}}$ with c, the PAV-score of W_{pav} will change by:

$$\begin{aligned}\Delta(c, c') &= \sum_{i\colon c \in A(i) \wedge c' \notin A(i)} \frac{1}{w_i + 1} - \sum_{i\colon c' \in A(i) \wedge c \notin A(i)} \frac{1}{w_i} \\ &= \sum_{i\colon c \in A(i)} \frac{1}{w_i + 1} - \sum_{i\colon c' \in A(i)} \frac{1}{w_i} + \sum_{i\colon \{c, c'\} \subseteq A(i)} \frac{1}{w_i} - \frac{1}{w_i + 1} \qquad (1) \\ &\geq \sum_{i \in N(c)} \frac{1}{w_i + 1} - \sum_{i \in N(c')} \frac{1}{w_i}.\end{aligned}$$

Let us now compute the sum:

$$\begin{aligned}\sum_{c' \in W_{\mathrm{pav}}} \Delta(c,c') &= \sum_{c' \in W_{\mathrm{pav}}} \sum_{i \in N(c)} \frac{1}{w_i+1} - \sum_{c' \in W_{\mathrm{pav}}} \sum_{i \in N(c')} \frac{1}{w_i} \\ &= k \sum_{i \in N(c)} \frac{1}{w_i+1} - \sum_{i \in N} \sum_{c' \in W_{\mathrm{pav}} \cap A(i)} \frac{1}{w_i} = k \sum_{i \in N(c)} \frac{1}{w_i+1} - n_{\mathrm{pav}}\end{aligned} \tag{2}$$

We know that for each $c' \in W$ we have $\Delta(c,c') \leq 0$, thus $k \sum_{i \in N(c)} \frac{1}{w_i+1} - n_{\mathrm{pav}} \leq 0$ and $\sum_{i \in N(c)} \frac{1}{w_i+1} \leq \frac{n_{\mathrm{pav}}}{k}$. We now use the inequality between harmonic and arithmetic mean to get:

$$\frac{n_{\mathrm{pav}}}{k} \geq \sum_{i \in N(c)} \frac{1}{w_i+1} \geq \frac{n_c^2}{\sum_{i \in N(c)}(w_i+1)} = \frac{n_c^2}{\sum_{i \in N(c)} w_i + n_c}.$$

This can be reformulated as:

$$kn_c \leq \frac{n_{\mathrm{pav}}\left(\sum_{i \in N(c)} w_i + n_c\right)}{n_c} = \frac{n_{\mathrm{pav}} \sum_{i \in N(c)} w_i}{n_c} + n_{\mathrm{pav}}$$

Now, let us consider two cases. If $n_{\mathrm{pav}} \leq n_c\sqrt{k}$, then we observe that:

$$\begin{aligned}\frac{\mathrm{sc}_{\mathrm{av}}(A, W_{\mathrm{av}})}{\mathrm{sc}_{\mathrm{av}}(A, W_{\mathrm{pav}})} &\leq \frac{\sum_{i \in N} w_i + kn_c}{\sum_{i \in N} w_i} = 1 + \frac{kn_c}{\sum_{i \in N} w_i} \leq 1 + \frac{\frac{n_{\mathrm{pav}} \sum_{i \in N(c)} w_i}{n_c} + n_{\mathrm{pav}}}{\sum_{i \in N} w_i} \\ &\leq 2 + \frac{\frac{n_{\mathrm{pav}} \sum_{i \in N(c)} w_i}{n_c}}{\sum_{i \in N} w_i} \leq 2 + \frac{n_{\mathrm{pav}}}{n_c} \leq \sqrt{k} + 2.\end{aligned}$$

On the other hand, if $n_{\mathrm{pav}} \geq n_c\sqrt{k}$, then:

$$\frac{\mathrm{sc}_{\mathrm{av}}(A, W_{\mathrm{av}})}{\mathrm{sc}_{\mathrm{av}}(A, W_{\mathrm{pav}})} \leq \frac{\sum_{i \in N} w_i + kn_c}{\sum_{i \in N} w_i} = 1 + \frac{kn_c}{\sum_{i \in N} w_i} \leq 1 + \frac{kn_c}{n_{\mathrm{pav}}} \leq 1 + \sqrt{k}.$$

In either case we have that $\frac{\mathrm{sc}_{\mathrm{av}}(A, W_{\mathrm{pav}})}{\mathrm{sc}_{\mathrm{av}}(A, W_{\mathrm{av}})} \geq \frac{1}{2+\sqrt{k}}$. This yields the required lower bound.

The fact that the AV-guarantee of PAV is at most equal to $\frac{2}{\lfloor\sqrt{k}\rfloor} - \frac{1}{k}$ follows from Proposition 3 and the fact that PAV satisfies lower-quota [11]. □

Theorem 2. *The CC-guarantee of PAV is between $\frac{1}{2}$ and $\frac{1}{2} + \frac{1}{4k-2}$.*

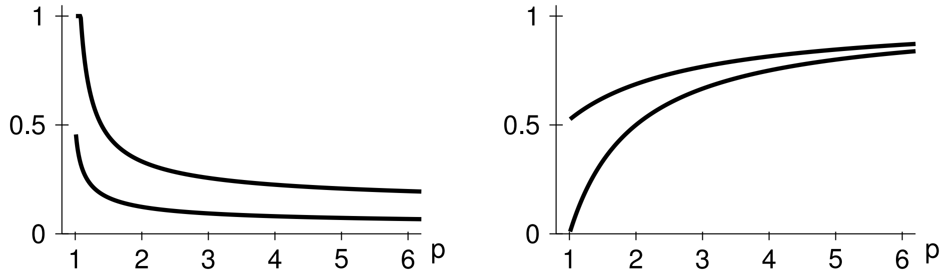
3.5 Guarantees for Sequential PAV

For sequential PAV we can prove qualitatively similar AV-guarantees to the ones for PAV.

Theorem 3. *The AV-guarantee of sequential PAV is between $\frac{1}{2\sqrt{k}}$ and $\frac{2}{\lfloor\sqrt{k}\rfloor} - \frac{1}{k}$.*

Let us now discuss the CC-guarantee of sequential PAV. One can observe that the construction for PAV from Theorem 2 also works for sequential PAV, which shows that the CC-guarantee of seq-PAV is at most equal to $\frac{1}{2} + \frac{1}{4k-2}$. Proposition 7 below establishes a lower bound. In this case however, the gap between the lower and upper bounds on the CC-guarantee of the rule is large. Finding a more accurate estimate remains an interesting open question.

Proposition 7. *The CC-guarantee of sequential PAV is at least equal to $\frac{1}{\log(k)+2}$.*



(a) AV-guarantee (b) CC-guarantee

Figure 1: Visualization of guarantees from Theorem 4 and Theorem 5: AV- and CC-guarantees for $k = 20$ and varying p. On each figure the upper and the lower line depict the upper and the lower bound, respectively, on the appropriate guarantee.

3.6 Guarantees for p-Geometric Rule

The following two theorems estimate the guarantees for the p-geometric rule. These guarantees are visualized in Figure 1. We can see that p-geometric rules, for $p \in [1, \infty)$, form a spectrum connecting AV and CC (with $p \to 1$ we approach AV and with $p \to \infty$ we approach CC): by adjusting the parameter p one can control the tradeoff between the diversity and individual excellence of the rule.

Let us recall that $\mathrm{W}(\cdot)$ denotes the Lambert W function. For each z it holds that $z = \mathrm{W}(z)e^{\mathrm{W}(z)}$. Intuitively, $\mathrm{W}(\cdot)$ is a function that asymptotically increases slower than the natural logarithm $\log$.

Theorem 4. *The AV-guarantee of the p-geometric rule is between:*

$$\frac{\mathrm{W}(k\log(p))}{k\log(p) + \mathrm{W}(k\log(p))} \qquad \textit{and} \qquad \frac{2\mathrm{W}(k\log(p))}{k\log(p)} + \frac{1}{k}.$$

Theorem 5. *The CC-guarantee of the p-geometric rule is between $\frac{p-1}{p}$ and $\frac{p}{p+\frac{k}{k+2}}$.*

3.7 Guarantees for the Sequential Phragmén's Rule

Finally we consider seq-Phragmén, another rule aimed at achieving proportionality of representation.

Theorem 6. *The AV-guarantee of seq-Phragmén is between $\frac{1}{5\sqrt{k}+1}$ and $\frac{2}{\lfloor\sqrt{k}\rfloor} - \frac{1}{k}$.*

The next theorem shows that the CC-guarantee of seq-Phragmén is asymptotically equal to $\frac{1}{2}$.

Theorem 7. *The CC-guarantee of seq-Phragmén is between $\frac{1}{2}$ and $\frac{1}{2} + \frac{1}{4k-2}$.*

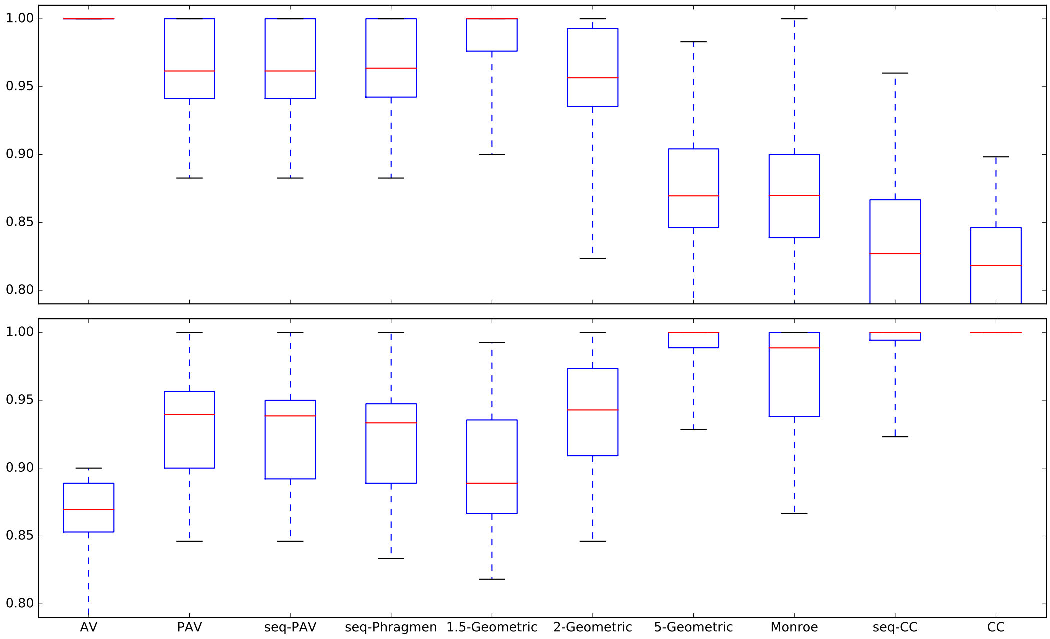
4 Average Guarantees: Experimental Analysis

To complement the theoretical analysis of Section 4, we have run experiments that aim at assessing *AV-ratios* and *CC-ratios* achieved by several voting rules. These two ratios are per-instance analogues of AV- and CC-guarantee and are defined as follows: Given a voting rule $\mathcal{R}$ and a profile A, the AV-ratio and the CC-ratio are defined as:

$$\frac{\min_{W \in \mathcal{R}(A,k)} \mathrm{sc}_{\mathrm{av}}(A, W)}{\max_{W \in \mathcal{S}_k(C)} \mathrm{sc}_{\mathrm{av}}(A, W)} \qquad \text{and} \qquad \frac{\min_{W \in \mathcal{R}(A,k)} \mathrm{sc}_{\mathrm{cc}}(A, W)}{\max_{W \in \mathcal{S}_k(C)} \mathrm{sc}_{\mathrm{cc}}(A, W)}.$$

In these experiments, we have calculated the AV- and CC-ratios for real-world and randomly generated profiles and compared them for different voting rules. We have used two data sets: profiles obtained from `preflib.org` [25] and profiles generated via an uniform distribution (see details below).

Figure 2: Results for the preflib dataset (upper boxplot shows AV-ratios, the lower CC-ratios).



Datasets. We restricted our attention to profiles where both the AV-ratio of CC and the CC-ratio of AV is at most 0.9. This excludes profiles where an (almost) perfect compromise between AV and CC exists. The uniform dataset consists of 500 profiles with 20 candidates and 50 voters, each. Voters' approval sets are of size 2–5 (chosen uniformly at random); the approval sets of a given size are also chosen uniformly at random. Experiments for the uniform dataset use a committee size of $k = 5$.

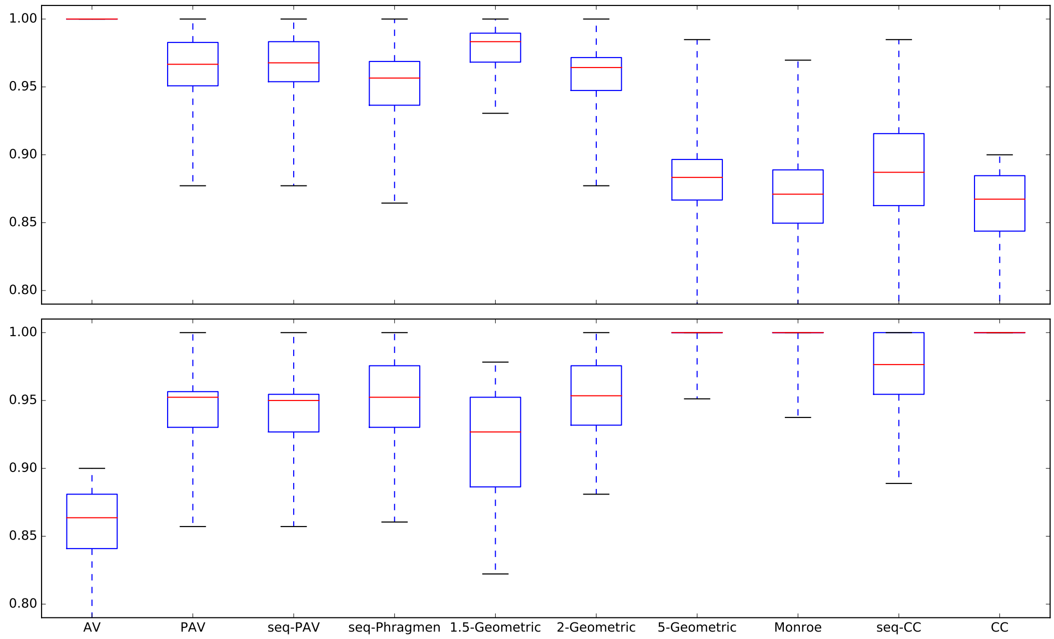
The preflib dataset is based on preferences obtained from `preflib.org`. Since their database does not contain approval-based datasets, we extracted approval profiles from ranked ballots as follows: for each ranked profile and $i \in \{1, \dots, k-1\}$, we generated an approval profile assuming that voters approve all candidates that are ranked in the top i positions. As before, we excluded profiles that allowed an almost perfect compromise between AV and CC. For the preflib dataset we considered $k \in \{3, \dots, 7\}$ and obtained a total number of 243 instances.

Results. We considered the following voting rules: AV, CC, seq-CC, PAV, seq-PAV, seq-Phragmén, Monroe's rule, as well as the 1.5-, 2-, and 5-Geometric rule. Our results are displayed as boxplots in Figure 2 for the preflib dataset and in Figure 3 for the uniform dataset. The top and bottom of boxes represent the first and third quantiles, the middle red bar shows the median. The dashed intervals (whiskers) show the range of all values, i.e., the minimum and maximum AV- or CC-ratio. The results for the preflib and random dataset are largely similar; we comment on the differences later on.

The main conclusion from the experiments is that the classification obtained from worst-case analytical bounds also holds in our (average-case) experiments. PAV, seq-PAV, and seq-Phragmén perform very well with respect to the AV-ratio, beaten only by 1.5-Geometric and AV itself. This is mirrored by our theoretical results as only PAV, seq-PAV, and seq-Phragmén achieve a $\Theta(1/\sqrt{k})$ AV-guarantee. For the uniform dataset, however, seq-Phragmén has slightly lower AV-ratios, but still comparable to PAV and seq-PAV. Also the 2-Geometric rule achieves comparable AV-ratios. Even better AV-ratios are achieved only by 1.5-Geometric and—by definition—by AV.

Considering the CC-ratio, we see almost optimal performance of seq-CC, Monroe, and 5-

Figure 3: Results for the uniform dataset (upper boxplot shows AV-ratios, the lower CC-ratios).



Geometric, and good performance of PAV, seq-PAV, seq-Phragmén, and 2-Geometric. Minor variations within these groups seem to depend on the chosen dataset. We also observe that 5-Geometric is better than Monroe's rule and seq-CC according to both criteria.

When looking at the three Geometric rules considered here, we see the transition from AV to CC as our theoretical findings predict (cf. Figure 1): 1.5-Geometric is close to AV, whereas 5-Geometric resembles CC; 2-Geometric performs very similarly to PAV, slightly favoring diversity over IE.

Our results indicate that PAV is the best compromise between AV and CC. Yet, seq-PAV, seq-Phragmén, and 2-Geometric achieve comparable ratios, and the former two are cheaper to compute.

5 A Pareto Efficiency Axiom

In this section, we provide a complementary axiomatic analysis concerning individual excellence. We formulate the axiom of efficiency, a form of Pareto efficiency with respect to the number of approved candidates in a committee. In other words, this axiom dictates that only committees can be chosen where a further improvement of the total AV-score implies that the AV-score of some individual voter is reduced. We analyze our rules with respect to this property, and, maybe surprisingly, show that many rules do not satisfy this basic axiom.

Definition 3. *Consider a committee size $k \in \mathbb{N}$, two committees $W_1, W_2 \in \mathcal{S}_k(C)$ and an approval profile $A \in \mathcal{A}$. We say that W_1 dominates W_2 in a A if for each voter $i \in N$ we have that $|W_1 \cap A(i)| \geq |W_2 \cap A(i)|$, and if there exists a voter j such that $|W_1 \cap A(j)| > |W_2 \cap A(j)|$.*

An ABC rule $\mathcal{R}$ satisfies efficiency if for each profile $A \in \mathcal{A}$ and each committee size k there exists no committee $W \in \mathcal{S}_k(C)$ that dominates each committee in $\mathcal{R}(A, k)$.

We start aith the rather surprising observation that seq-Phragmén does not satisfy efficiency.

Example 2. *Consider the set of 36 voters, and five candidates, $c_1, \dots, c_5$. By $N(c)$ we denote the set of voters who approve c. Assume that:*

$$N(c_1) = \{1, \dots 20\}; \qquad N(c_2) = \{11, \dots 28\}; \qquad N(c_3) = \{1, \dots 10, 29, \dots, 36\};$$
$$N(c_4) = \{21, \dots 36\}; \qquad N(c_5) = \{1, \dots 19\}.$$

The sequential Phragmén's rule will select c_1 first, c_4 second, and c_5 third, yet committee $\{c_1, c_4, c_5\}$ is dominated by $\{c_1, c_2, c_3\}$. This example also works for the Open d'Hondt method, which can be viewed as another variant of the Phragmén's rule [33].

We note that the violation of efficiency is not an artifact of the rule being sequential (and so, in some sense "suboptimal"). Indeed, consider the optimal Phragmén's rule, which is the variant where the committee members and their associated load distributions are not chosen sequentially, but rather simultaneously in a single step. Similarly, as in the case of its sequential counterpart, the goal of the optimal Phragmén's rule is to find a committee and an associated load distribution that minimizes the load of the voter with the highest load (for more details on this rule we refer the reader to the work of Brill et al. [10]). The following example shows that the optimal Phragmén's rule does not satisfy efficiency. The same example shows that the Monroe rule does not satisfy efficiency.

Example 3. *Consider 24 voters, and four candidates, $c_1, \dots, c_4$, with the following preferences:*

$$N(c_1) = \{3, \dots 22\}; \qquad N(c_2) = \{1, 2, 23, 24\};$$
$$N(c_3) = \{2, \dots 12\}; \qquad N(c_4) = \{13, \dots 23\}.$$

The optimal Phragmén's and the Monroe's rule would select $\{c_3, c_4\}$, which is dominated by $\{c_1, c_2\}$.

Greedy Monroe, seq-CC, and seq-PAV do not satisfy efficiency either. Intuitively, this is due to their sequential nature.

Example 4. *Consider the following profile with 20 voters and 4 candidates, where:*

$$N(c_1) = \{2, \dots 10\}; \qquad N(c_2) = \{11, \dots 19\};$$
$$N(c_3) = \{6, \dots 15\}; \qquad N(c_4) = \{2, 3, 4, 16, 17, 18, 19\}.$$

For this profile and for $k = 2$ the greedy Monroe rule first picks c_3, who is approved by 10 voters, will remove these 10 voters, and will pick c_4. However, committee $\{c_3, c_4\}$ is dominated by $\{c_1, c_2\}$. The same example shows that seq-CC and seq-PAV do not satisfy efficiency.

All the remaining rules that we consider satisfy efficiency.

Proposition 8. *AV, CC, PAV, and p-geometric satisfy efficiency.*

6 Conclusion and Future Work

Our work demonstrates the flow of ideas from theoretical computer science to theoretical economics, in particular to social choice. We designed new tools that can be used to assess the level of diversity and individual excellence provided by certain rules. Our results help to understand the landscape of multiwinner rules, specifically how they behave with respect to two contradictory goals.

Our work can be extended in several directions. First, we have focused on approval-based multi-winner rules—a natural next step is to perform a similar analysis for multi-winner rules that take rankings over candidates as input. Second, we have excluded some interesting voting rules from our analysis, in particular reverse-sequential PAV [37] and Minimax Approval Voting [8]; it is unclear how they compare to rules considered in this paper. Finally, we have chosen AV and CC as extreme notions that represent diversity and individual excellence. Another natural approach would be to take a proportional rule (such as PAV) as a standard and see how well others rules approximate it.

Acknowledgements

Martin Lackner was supported by the European Research Council (ERC) under grant number 639945 (ACCORD) and by the Austrian Science Foundation FWF, grant P25518 and Y698. Piotr Skowron was supported by a Humboldt Research Fellowship for Postdoctoral Researchers.

References

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

[2] H. Aziz, E. Elkind, P. Faliszewski, M. Lackner, and P. Skowron. The Condorcet principle for multiwinner elections: From shortlisting to proportionality. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI-2017)*, pages 84–90, 2017.

[3] H. Aziz, E. Elkind, S. Huang, M. Lackner, L. Sánchez-Fernández, and P. Skowron. On the complexity of extended and proportional justified representation. In *Proceedings of the 32nd Conference on Artificial Intelligence (AAAI-2018)*, 2018. To appear.

[4] H. Aziz, S. Gaspers, J. Gudmundsson, S. Mackenzie, N. Mattei, and T. Walsh. Computational aspects of multi-winner approval voting. In *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems*, May 2015.

[5] M. Balinski and H. P. Young. *Fair Representation: Meeting the Ideal of One Man, One Vote*. Yale University Press, 1982. (2nd Edition [with identical pagination], Brookings Institution Press, 2001).

[6] S. Barberà and D. Coelho. How to choose a non-controversial list with k names. *Social Choice and Welfare*, 31(1):79–96, 2008.

[7] D. Baumeister, T. Böhnlein, L. Rey, O. Schaudt, and A. Selker. Minisum and minimax committee election rules for general preference types. In *Proceedings of the 22nd European Conference on Artificial Intelligence*, pages 1656–1657, 2016.

[8] S. J. Brams, D. M. Kilgour, and M. R. Sanver. A minimax procedure for electing committees. *Public Choice*, 132(3–4):401–420, 2007.

[9] R. Bredereck, P. Faliszewski, A. Kaczmarczyk, R. Niedermeier, P. Skowron, and N. Talmon. Robustness among multiwinner voting rules. In *Proceedings of the 10th International Symposium on Algorithmic Game Theory (SAGT-2017)*, pages 80–92, 2017.

[10] M. Brill, R. Freeman, S. Janson, and M. Lackner. Phragmén's voting methods and justified representation. In *Proceedings of the 31st Conference on Artificial Intelligence (AAAI-2017)*, pages 406–413, 2017.

[11] M. Brill, J. Laslier, and P. Skowron. Multiwinner approval rules as apportionment methods. In *Proceedings of the 31st Conference on Artificial Intelligence (AAAI-2017)*, pages 414–420, 2017.

[12] I. Caragiannis, S. Nath, A. D. Procaccia, and N. Shah. *Journal of Artificial Intelligence Research*, 58:123–152, 2017.

[13] B. Chamberlin and P. Courant. Representative deliberations and representative decisions: Proportional representation and the Borda rule. *American Political Science Review*, 77(3):718–733, 1983.

[14] M. Dummett. *Voting Procedures*. Oxford University Press, 1984.

[15] C. Dwork, R. Kumar, M. Naor, and D. Sivakumar. Rank aggregation methods for the web. In *Proceedings of the 10th International World Wide Web Conference*, pages 613–622. ACM Press, Mar. 2001.

[16] E. Elkind, P. Faliszewski, J. F. Laslier, P. Skowron, A. Slinko, and N. Talmon. What do multiwinner voting rules do? An experiment over the two-dimensional euclidean domain. In *Proceedings of the 31st Conference on Artificial Intelligence (AAAI-2017)*, 2017. 494–501.

[17] E. Elkind, P. Faliszewski, P. Skowron, and A. Slinko. Properties of multiwinner voting rules. *Social Choice and Welfare*, 48(3):599–632, 2017.

[18] P. Faliszewski, P. Skowron, A. Slinko, and N. Talmon. Multiwinner voting: A new challenge for social choice theory. In U. Endriss, editor, *Trends in Computational Social Choice*. AAAI Press, 2017.

[19] M. Feldman, A. Fiat, and I. Golomb. On voting and facility location. In *Proceedings of the 17th ACM Conference on Economics and Computation (ACM-EC-2016)*, pages 269–286, 2016.

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

[21] Y. Koriyama, J.-F. Laslier, A. Macé, and R. Treibich. Optimal Apportionment. *Journal of Political Economy*, 121(3):584–608, 2013.

[22] M. Lackner and P. Skowron. Consistent approval-based multi-winner rules. Technical Report arXiv:1704.02453v1 [cs.GT], arXiv.org, Apr. 2017.

[23] J.-F. Laslier and K. Van der Straeten. Strategic voting in multi-winners elections with approval balloting: a theory for large electorates. *Social Choice and Welfare*, 2016. Forthcoming.

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

[25] N. Mattei and T. Walsh. Preflib: A library for preferences http: //www.preflib.org. In *Proceedings of the 3rd International Conference on Algorithmic Decision Theory (ADT-2013)*, pages 259–270, 2013.

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

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

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

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

[30] A. Procaccia, S. Reddi, and N. Shah. A maximum likelihood approach for selecting sets of alternatives. In *Proceedings of the 28th Conference in Uncertainty in Artificial Intelligence (UAI-2012)*, pages 695–704, 2012.

[31] A. Rapoport, D. Felsenthal, and Z. Maoz. Proportional representation: An empirical evaluation of single-stage, non-ranked voting procedures. *Public Choice*, 59(2):151–165, 1988.

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

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

[34] P. Skowron. What do we elect committees for? a voting committee model for multi-winner rules. pages 1141–1148, 2015.

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

[36] P. Skowron, P. Faliszewski, and A. Slinko. Achieving fully proportional representation: Approximability result. *Artificial Intelligence*, 222:67–103, 2015.

[37] P. Skowron, M. Lackner, M. Brill, D. Peters, and E. Elkind. Proportional rankings. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI-2017)*, pages 409–415, 2017.

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

[39] K. Van der Straeten, R. Lachat, and J.-F. Laslier. *Strategic voting in multi-winner elections with approval balloting: An application to the 2011 regional government election in Zurich*. CBS, 2017. forthcoming.

Martin Lackner
Database and Artificial Intelligence Group
TU Wien
Austria
Email: lackner@dbai.tuwien.ac.at

Piotr Skowron
Department of Mathematics, Informatics and Mechanics
University of Warsaw
Poland
Email: p.skowron@mimuw.edu.pl

A Proofs Omitted from the Main Text

Proposition 2. *The AV-guarantee of CC and sequential CC is $1/k$.*

Proof. For an approval profile A let W_{cc} and W_{av} be committees winning according to CC and AV, respectively. We will first prove that $\mathrm{sc}_{\mathrm{av}}(A, W_{\mathrm{cc}}) \geq \frac{\mathrm{sc}_{\mathrm{av}}(A, W_{\mathrm{av}})}{k}$. If it was not the case, then by the pigeonhole principle, there would exists a candidate $c \in W_{\mathrm{av}}$ such that $\mathrm{sc}_{\mathrm{av}}(A, W_{\mathrm{cc}}) < \mathrm{sc}_{\mathrm{av}}(A, \{c\})$. However, this means that a committee that consists of c and any $k-1$ candidates has a higher CC-score than W_{cc}, a contradiction. Thus, the AV-guarantee of CC is at least $1/k$. For seq-CC, the same argument by contradiction applies as this candidate c would have been chosen in the first round.

To see that this guarantee cannot be higher than $1/k$ consider the following profile: assume there are x voters (x is a large integer) who approve candidates $c_1, \ldots, c_k$. Further, for each candidate $c_{k+1}, \ldots, c_{2k}$ there is a single voter who approves only her. The CC-winning committee is $\{c_1, c_{k+1} \ldots, c_{2k-1}\}$ with the AV-score of $x + k - 1$. However, the AV-score of committee $\{c_1, \ldots c_k\}$ is xk, and for large enough x the ratio $\frac{x+k-1}{xk}$ can be made arbitrarily close to $1/k$. □

Proposition 4. *The CC-guarantee of a rule that satisfies lower quota is at most $\frac{3}{4} + \frac{3}{8k-4}$.*

Proof. Let $\mathcal{R}$ be a rule that satisfies lower quota. Consider a profile A with $n = 2kx$ voters for some $x \geq 1$. Each from the first kx voters approves candidates $X = \{x_1, \ldots, x_k\}$. The other voters are divided into k equal-size groups, each approving a different candidate from the set $Y = \{y_1, \ldots, y_k\}$. Lower quota ensures that at least $k/2$ candidates need to be chosen from X. Thus, the CC-score of a committee selected by $\mathcal{R}$ is at most equal to $kx + \frac{kx}{2}$. By selecting one candidate from X and $k-1$ candidates from Y we get a CC-score of $2kx - x$. Thus, the CC-guarantee is at most equal to:

$$\frac{kx + \frac{kx}{2}}{2kx - x} = \frac{3k}{4k-2} = \frac{3}{4} + \frac{3}{8k-4}.$$

□

Proposition 5. *The AV-guarantee of Greedy Monroe and Monroe is $1/k$.*

Proof. First, let us consider the greedy Monroe rule. To see the lower bound of $1/k$, let A be an approval profile and let $\bar{c}$ denote the candidate who is approved by most voters. For the sake of clarity we assume that k divides n; the proof can be generalized to hold for arbitrary n. Clearly, for any committee W it holds that $\mathrm{sc}_{\mathrm{av}}(A, W) \leq k|N(\bar{c})|$. If $|N(\bar{c})| \leq \frac{n}{k}$, then the greedy Monroe rule in the first step will select $\bar{c}$. Otherwise, it will select some candidate approved by at least $\frac{n}{k}$ voters, and will remove $\frac{n}{k}$ of them from A. By a similar reasoning we can infer that in the second step the rule will pick a candidate who is approved by at least $\min\left(\frac{n}{k}, |N(\bar{c})| - \frac{n}{k}\right)$ voters; and in general, that in the i-th step the rule will pick the candidate who is approved by at least $\min\left(\frac{n}{k}, |N(\bar{c})| - \frac{n(i-1)}{k}\right)$ voters. As a result, we infer that number of voters that have at least one approved candidate in the chosen committee is at least

$$\sum_{i=1}^{k} \min\left(\frac{n}{k}, |N(\bar{c})| - \frac{n(i-1)}{k}\right) = |N(\bar{c})|.$$

Hence the AV-guarantee of Greedy Monroe is at least $1/k$.

To see that the same lower bound holds for the Monroe rule, we distinguish two cases; let W be a winning committee. If $\bar{c} \in W$, then $\mathrm{sc}_{CC}(A, W) \geq |N(\bar{c})|$ and we are done. If $\bar{c} \notin W$ and $\mathrm{sc}_{CC}(A, W) < |N(\bar{c})|$, then there is a committee with a higher Monroe-score that contains $\bar{c}$; a contradiction.

Now, consider the following instance witnessing that the AV-guarantee of Greedy Monroe is at most $\frac{1}{k}$. Let $n = k \cdot (x+1)$ and let A be a profile with n voters. Let $W \subseteq C$ with $|W| = k$ and

$c_1, \ldots, c_k \notin W$. We define profile A as follows: we have x voters that approve $W \cup \{c_1\}$ and one voter that approves only $\{c_1\}$, we have x voters that approve $W \cup \{c_2\}$ and one voter that approves only $\{c_2\}$, etc. This defines in total $k \cdot (x+1)$ voters. AV selects the committee W with an AV-score of xk^2; Greedy Monroe selects the committee $\{c_1 \ldots, c_k\}$ with an AV-score of $(x+1)k$. We have a ratio of $\frac{(x+1)}{xk}$, which converges to $\frac{1}{k}$ for $x \to \infty$. The same instance shows that the AV-guarantee of the Monroe rule is at most $\frac{1}{k}$. $\square$

Proposition 6. *The CC-guarantee of Monroe and greedy Monroe is between* $\frac{1}{2}$ *and* $\frac{1}{2} + \frac{1}{k-1}$.

Proof. First, for the sake of contradiction let us assume that there exists a profile A where the CC-guarantee of Greedy Monroe is below $\frac{1}{2}$. Let W_{cc} and W_M be the committees winning in A according to CC and Greedy Monroe, respectively. Let ϕ be an assignment of the voters to the committee members obtained during the construction of W_M; we say that a voter is represented if it is assigned to a member of W_M who she approves of. Since $\mathrm{sc}_{\mathrm{cc}}(A, W_M) < \frac{1}{2} \cdot \mathrm{sc}_{\mathrm{cc}}(A, W_{\mathrm{cc}})$, by the pigeonhole principle we infer that there exists a candidate $c \in W_{\mathrm{cc}} \setminus W_M$ who is approved by x unrepresented voters, where:

$$x \geq \frac{\mathrm{sc}_{\mathrm{cc}}(A, W_{\mathrm{cc}}) - \mathrm{sc}_{\mathrm{cc}}(A, W_M)}{k} \geq \frac{2\mathrm{sc}_{\mathrm{cc}}(A, W_M) - \mathrm{sc}_{\mathrm{cc}}(A, W_M)}{k} = \frac{\mathrm{sc}_{\mathrm{cc}}(A, W_M)}{k}.$$

Similarly, by the pigeonhole principle we can infer that there exists a candidate $c' \in W_M$ who is represented by at most $\frac{\mathrm{sc}_{\mathrm{cc}}(A, W_M)}{k}$ voters. Thus, Greedy Monroe would select c rather than c', a contradiction. A similar argument can be made to show that the CC-guarantee of the Monroe rule is $\geq \frac{1}{2}$.

Now, consider the following approval profile. There are $2k+1$ candidates, $c_1, \ldots, c_{2k+1}$, and $2k$ disjoint equal-size groups of voters, $N_1, \ldots, N_{2k}$. For each $i \in [2k]$, candidate c_i is approved by all voters from N_i. Candidate c_{2k+1} is approved by all voters from $N_1 \cup \ldots \cup N_k$. One of the winning committees according to the Monroe and Greedy Monroe rule is $\{c_1, \ldots, c_{k-2}, c_{k+1}, c_{2k+1}\}$, which has a CC-score of $\frac{n}{k} + (k-1)\frac{n}{2k}$. On the other hand, $\{c_{k+1}, \ldots, c_{2k-1}, c_{2k+1}\}$ has a CC-score of $n - \frac{n}{k}$. Thus, the CC-guarantee of Monroe and Greedy Monroe is at most:

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

This completes the proof. $\square$

Theorem 2. *The CC-guarantee of PAV is between* $\frac{1}{2}$ *and* $\frac{1}{2} + \frac{1}{4k-2}$.

Proof. We first prove a lower bound of $1/2$ for the CC-guarantee of PAV. Consider an approval-based profile A and a PAV winning committee W_{pav}. Similarly as in the proof of Theorem 1, for each voter $i \in N$ we set $w_i = |A(i) \cap W_{\mathrm{pav}}|$. Let W_{cc} be a committee winning according to the Chamberlin–Courant rule For each two candidates, $c \in W_{\mathrm{pav}}$ and $c' \in W_{\mathrm{cc}}$, let $\Delta(c', c)$ denote the change of the PAV-score of W_{pav} due to replacing c with c'. By Inequality (1), we have:

$$\Delta(c', c) \geq \sum_{i \in N(c')} \frac{1}{w_i + 1} - \sum_{i \in N(c)} \frac{1}{w_i}.$$

Let us now consider an arbitrary bijection $\tau \colon W_{\mathrm{pav}} \to W_{\mathrm{cc}}$, matching members of W_{pav} with the members of W_{cc}. We compute the sum:

$$\begin{aligned}
\sum_{c \in W_{\text{pav}}} \Delta(\tau(c), c) &\geq \sum_{c' \in W_{\text{cc}}} \sum_{i \in N(c')} \frac{1}{w_i + 1} - \sum_{c \in W_{\text{pav}}} \sum_{i \in N(c)} \frac{1}{w_i} \\
&= \sum_{i \in N(W_{\text{cc}})} \underbrace{\sum_{c' \in W_{\text{cc}} \cap A(i)} \frac{1}{w_i + 1}}_{\geq \frac{1}{w_i+1}} - \sum_{i \in N(W_{\text{pav}})} \underbrace{\sum_{c \in W_{\text{pav}} \cap A(i)} \frac{1}{w_i}}_{=1} \\
&\geq \sum_{i \in N(W_{\text{cc}})} \frac{1}{w_i + 1} - |N(W_{\text{pav}})| \geq \sum_{i \in N(W_{\text{cc}}) \setminus N(W_{\text{pav}})} 1 - |N(W_{\text{pav}})| \\
&\geq |N(W_{\text{cc}}) \setminus N(W_{\text{pav}})| - |N(W_{\text{pav}})| \\
&\geq |N(W_{\text{cc}})| - |N(W_{\text{pav}})| - |N(W_{\text{pav}})| \\
&= |N(W_{\text{cc}})| - 2|N(W_{\text{pav}})|.
\end{aligned} \tag{3}$$

Since W_{pav} is an PAV-optimal committee, we know that for each $c \in W_{\text{pav}}$, it holds that $\Delta(\tau(c), c) \leq 0$. Consequently, $\sum_{c \in W_{\text{pav}}} \Delta(\tau(c), c) \leq 0$, and so we get that $|N_{W_{\text{cc}}}| - 2|N_{W_{\text{pav}}}| \leq 0$, Consequently, we get that $|N_{W_{\text{pav}}}| \geq \frac{|N_{W_{\text{cc}}}|}{2}$, which shows that the CC-guarantee of PAV is at least equal to $1/2$.

Now, we will prove the upper bound using the following construction. Let n, the number of voters, be divisible by $2k$. The set of candidates is $X \cup Y$ with $X = \{x_1, \ldots, x_k\}$ and $Y = \{y_1, \ldots, y_k\}$. There are $n/2$ voters who approve X. Further, for each $i \in [k]$, there are $\frac{n}{2k}$ voters who approve candidate y_i. All committees that contain at least $k-1$ candidates from X are winning according to PAV, among them X itself. Committee X has a CC-score of $n/2$. The optimal CC committee consists of a single candidate from X and $(k-1)$ candidates from Y—this would give a CC-score of $\frac{n}{2} + (k-1) \cdot \frac{n}{2k} = n \cdot \frac{2k-1}{2k}$. Thus, the CC-guarantee of PAV is at most equal to:

$$\frac{2k}{4k-2} = \frac{1}{2} + \frac{1}{4k-2}.$$

This completes the proof. □

Theorem 3. *The AV-guarantee of sequential PAV is between $\frac{1}{2\sqrt{k}}$ and $\frac{2}{\lfloor\sqrt{k}\rfloor} - \frac{1}{k}$.*

Proof. Since sequential PAV satisfies lower quota [11], the upper bound of $\frac{2}{\lfloor\sqrt{k}\rfloor} - \frac{1}{k}$ follows from Proposition 1. In the remaining part of the proof we will prove the lower-bound.

For $k = 1$, seq-PAV is AV and hence the Av-guarantee is 1. For $k = 2$, in the first step the AV-winner is chosen and hence we have an AV-guarantee for $k = 2$ is $\frac{3}{4} \geq \frac{1}{2\sqrt{2}}$. Now assume that $k \geq 3$. Let $W_{\text{pav}}^{(j)}$ denote the first j candidates selected by sequential PAV; in particular, $W_{\text{pav}}^{(0)} = \emptyset$. Let w_j denote the candidate selected by sequential PAV in the jth step, thus w_j is the single candidate in the set $W_{\text{pav}}^{(j)} \setminus W_{\text{pav}}^{(j-1)}$. Let $x_{i,j} = |W_{\text{pav}}^{(j)} \cap A(i)|$. Next, let W_{av} be the optimal committee according to Approval Voting, and let $s_{\text{av}} = \text{sc}_{\text{av}}(W_{\text{av}})$.

If at some step j of the run of sequential PAV, it happens that the AV-score of $W_{\text{pav}}^{(j)}$, which is $\sum_{i \in N} x_{i,j}$, is greater or equal than $\frac{s_{\text{av}}}{2\sqrt{k}}$, then our hypothesis is clearly satisfied. Thus, from now on, we assume that for each j we have that $\sum_{i \in N} x_{i,j} < \frac{s_{\text{av}}}{2\sqrt{k}}$. Also, this means that in each step there exists a candidate c from $W_{\text{av}} \setminus W_{\text{pav}}$ who is approved by $n_c \geq \frac{s_{\text{av}} - \frac{s_{\text{av}}}{2\sqrt{k}}}{k} \geq \frac{s_{\text{av}}}{k}(1 - \frac{1}{2\sqrt{3}})$ voters. Let $n_c = |N(c)|$.

Let Δp_j denote the increase of the PAV-score due to adding w_{j+1} to $W_{\text{pav}}^{(j)}$. Using the inequality between harmonic and arithmetic mean, we have that:

$$\Delta p_j = \sum_{i \in N(c)} \frac{1}{x_{i,j} + 1} \geq \frac{n_c^2}{\sum_{i \in N(c)} x_{i,j} + n_c} > \frac{n_c^2}{\frac{s_{\text{av}}}{2\sqrt{k}} + n_c}$$

$$\geq \frac{\left(\frac{s_{\mathrm{av}}}{k}(1-\frac{1}{2\sqrt{3}})\right)^2}{\frac{s_{\mathrm{av}}}{2\sqrt{k}}+\frac{s_{\mathrm{av}}}{k}(1-\frac{1}{2\sqrt{3}})} \geq \frac{\left(\frac{s_{\mathrm{av}}}{k}(1-\frac{1}{2\sqrt{3}})\right)^2}{\frac{s_{\mathrm{av}}}{\sqrt{k}}(\frac{1}{2}+\frac{1}{\sqrt{3}}-\frac{1}{6})}$$
$$= \frac{s_{\mathrm{av}}}{k\sqrt{k}} \cdot \underbrace{\frac{\left(1-\frac{1}{2\sqrt{3}}\right)^2}{\frac{1}{2}+\frac{1}{\sqrt{3}}-\frac{1}{6}}}_{\approx 0.56} > \frac{s_{\mathrm{av}}}{2k\sqrt{k}}.$$

Since this must hold in each step of sequential PAV, we get that the total PAV-score of $W_{\mathrm{pav}}^{(k)}$ must be at least equal to $k \cdot \frac{s_{\mathrm{av}}}{2k\sqrt{k}} = \frac{s_{\mathrm{av}}}{2\sqrt{k}}$. Since the AV-score is at least equal to the PAV-score of any committee, we obtain a contradiction and conclude that $\mathrm{sc}_{\mathrm{av}}(A, W_{\mathrm{pav}}^{(k)}) \geq \frac{s_{\mathrm{av}}}{2\sqrt{k}}$. □

Proposition 7. *The CC-guarantee of sequential PAV is at least equal to* $\frac{1}{\log(k)+2}$.

Proof. Consider an approval profile A and let W_{spav} and W_{cc} denote the winning committees in A according to seq-PAV and CC, respectively. Let $n_{\mathrm{spav}} = \mathrm{sc}_{\mathrm{cc}}(A, W_{\mathrm{spav}})$ and $n_{\mathrm{cc}} = \mathrm{sc}_{\mathrm{cc}}(A, W_{\mathrm{cc}})$. The total PAV-score of W_{spav} is at most equal to $n_{\mathrm{spav}}\mathrm{H}(k) \leq n_{\mathrm{spav}}(\log(k)+1)$. Thus, at some step sequential PAV selected a committee member who improved the PAV-score by at most $\frac{n_{\mathrm{spav}}(\log(k)+1)}{k}$. On the other hand, by the pigeonhole principle, we know that at each step of seq-PAV there exists a not-selected candidate whose selection would improve the PAV-score by at least $\frac{n_{\mathrm{cc}}-n_{\mathrm{spav}}}{k}$. Consequently, we get that

$$\frac{n_{\mathrm{spav}}(\log(k)+1)}{k} \geq \frac{n_{\mathrm{cc}}-n_{\mathrm{spav}}}{k}.$$

After reformulation we have that $n_{\mathrm{spav}} \geq \frac{n_{\mathrm{cc}}}{\log(k)+2}$, which completes the proof. □

Theorem 4. *The AV-guarantee of the p-geometric rule is between:*

$$\frac{\mathrm{W}(k\log(p))}{k\log(p)+\mathrm{W}(k\log(p))} \quad \text{and} \quad \frac{2\mathrm{W}(k\log(p))}{k\log(p)}+\frac{1}{k}.$$

Proof. We use the same notation as in the proof of Theorem 1 with a difference that instead of W_{pav} (denoting a PAV winning committee) we will use $W_{\text{p-geom}}$, denoting a committee winning according to the p-geometric rule. By repeating the reasoning from the proof of Theorem 1 instead of Inequality (2) we would obtain:

$$\sum_{c' \in W_{\text{p-geom}}} \Delta(c,c') = k\sum_{i \in N(c)}\left(\frac{1}{p}\right)^{w_i+1} - \sum_{i\in N}\sum_{c' \in W_{\mathrm{pav}} \cap A(i)}\left(\frac{1}{p}\right)^{w_i}$$
$$= k\sum_{i \in N(c)}\left(\frac{1}{p}\right)^{w_i+1} - \sum_{i\in N} w_i\left(\frac{1}{p}\right)^{w_i}$$

By using Jensen's inequality we get that $\sum_{i\in N(c)} \frac{1}{n_c}\cdot\left(\frac{1}{p}\right)^{w_i+1} \geq \left(\frac{1}{p}\right)^{\frac{\sum_{i\in N(c)} w_i + n_c}{n_c}}$. Thus:

$$\sum_{c' \in W_{\text{p-geom}}} \Delta(c,c') = kn_c\left(\frac{1}{p}\right)^{\frac{\sum_{i\in N} w_i}{n_c}+1} - \sum_{i\in N} w_i\left(\frac{1}{p}\right)^{w_i}$$
$$\geq kn_c\left(\frac{1}{p}\right)^{\frac{\sum_{i\in N} w_i}{n_c}+1} - \frac{1}{p}\sum_{i\in N} w_i$$

Since we know that $\sum_{c' \in W_{\text{p-geom}}} \Delta(c, c') \leq 0$, we have that:

$$\frac{1}{p} \sum_{i \in N} w_i \geq k n_c \left(\frac{1}{p}\right)^{\frac{\sum_{i \in N} w_i}{n_c} + 1}$$

Let us set $r = \frac{k n_c}{\sum_{i \in N} w_i}$, and observe (similarly as in the proof of Theorem 1) that $\frac{\mathrm{sc}_{\mathrm{av}}(A, W_{\mathrm{av}})}{\mathrm{sc}_{\mathrm{av}}(A, W_{\text{p-geom}})} \leq 1 + r$. We have that $p^{\frac{k}{r}} \geq r$. The equation $p^{\frac{k}{r}} = r$ has only one solution, $r = \frac{k \log(p)}{\mathrm{W}(k \log(p))}$. This gives $r \leq \frac{k \log(p)}{\mathrm{W}(k \log(p))}$ and proves that the AV-guarantee is at least equal to $\frac{\mathrm{W}(k \log(p))}{k \log(p) + \mathrm{W}(k \log(p))}$.

Now, let us prove the upper bound on the AV-guarantee. Let $z = \frac{k \log(p)}{\mathrm{W}(k \log(p))}$; in particular, by the properties of the Lambert function we have that $z = p^{\frac{k}{z}}$. Consider the following instance. Let x be a large integer so that $\lfloor x \cdot z \rfloor \approx xz$. (Formally, we choose an increasing sequence $\bar{x}$ so that $z\bar{x} - \lfloor z\bar{x} \rfloor \to 0$.) Assume there are $\lfloor x \cdot z \rfloor$ voters who approve candidates $B = \{c_1, \ldots, c_k\}$. Additionally, for each candidate $c \in D = \{c_{k+1}, \ldots, c_{2k}\}$ there are x distinct voters who approve c. For this instance the p-geometric rule selects at most $\lceil \frac{k}{z} \rceil$ members from B: if more candidates from B were selected, then replacing one candidate from B with a candidate from D would increase the p-geometric-score by more than

$$\frac{x}{p} - \lfloor x \cdot z \rfloor \cdot \left(\frac{1}{p}\right)^{\lceil \frac{k}{z} \rceil + 1} > \frac{x}{p} - \frac{x}{p} \cdot z \cdot \left(\frac{1}{p}\right)^{\frac{k}{z}} = \frac{x}{p} - \frac{x}{p} \cdot z \cdot \left(\frac{1}{z}\right) = 0,$$

a contradiction. Thus, the AV-score of the committee selected by the p-geometric rule would be smaller than $x \cdot z \cdot \left(1 + \frac{k}{z}\right) + kx = xz + 2kx$. Thus, we get that the AV-guarantee of the p-geometric rule is at most equal to:

$$\frac{2kx + xz}{xzk} = \frac{1}{k} + \frac{2}{z} = \frac{1}{k} + \frac{2\mathrm{W}(k \log(p))}{k \log(p)}.$$

□

Theorem 5. *The CC-guarantee of the p-geometric rule is between* $\frac{p-1}{p}$ *and* $\frac{p}{p + \frac{k}{k+2}}$.

Proof. Let A be an approval profile and let W_{cc} and $W_{\text{p-geom}}$ be two committees winning according to the Chamberlin–Courant and p-geometric rule, respectively. Let $n_{\text{p-geom}} = \mathrm{sc}_{\mathrm{cc}}(A, W_{\text{p-geom}})$ and $n_{\mathrm{cc}} = \mathrm{sc}_{\mathrm{cc}}(A, W_{\mathrm{cc}})$. We observe that:

$$\mathrm{sc}_{\text{p-geom}}(A, W_{\text{p-geom}}) \leq n_{\text{p-geom}} \left(\frac{1}{p} + \frac{1}{p^2} + \ldots\right) \leq n_{\text{p-geom}} \cdot \frac{1}{p} \cdot \frac{1}{1 - \frac{1}{p}}$$

and that:

$$\mathrm{sc}_{\text{p-geom}}(A, W_{\mathrm{cc}}) \geq n_{\mathrm{cc}} \cdot \frac{1}{p}.$$

Consequently, from $\mathrm{sc}_{\text{p-geom}}(A, W_{\text{p-geom}}) \geq \mathrm{sc}_{\text{p-geom}}(A, W_{\mathrm{cc}})$ we get that:

$$n_{\text{p-geom}} \cdot \frac{1}{1 - \frac{1}{p}} \geq p \cdot \mathrm{sc}_{\text{p-geom}}(A, W_{\text{p-geom}}) \geq p \cdot \mathrm{sc}_{\text{p-geom}}(A, W_{\mathrm{cc}}) \geq n_{\mathrm{cc}},$$

which gives the lower bound on the CC-guarantee.

Now, let us prove the upper bound. Fix a rational number p and some large integer x such that px is integer. First, let k be even with $k = 2k'$. Let the set of candidates be $\{x_1, \ldots, x_k\} \cup \{y_1, \ldots, y_{k'}\}$.

There are k' groups of voters who consists of px voters; in each group voters approve some two distinct candidates from $\{x_1, \ldots, x_k\}$. Additionally, there are k' groups consisting of x voters who approve some distinct candidate from $\{y_1, \ldots, y_{k'}\}$. It is easy to see that for such instances the CC-guarantee is at most equal to $\frac{k'px}{k'px+k'x} = \frac{p}{1+p}$.

Now, let k be odd with $k = 2k' + 1$; the set of candidates is $\{x_1, \ldots, x_{2k'+2}\} \cup \{y_1, \ldots, y_{k'}\}$. There are $k' + 1$ groups of voters who consists of px voters; in each group voters approve some two distinct candidates from $\{x_1, \ldots, x_{2k'+2}\}$. Additionally, there are k' groups consisting of x voters who approve some distinct candidate from $\{y_1, \ldots, y_{k'}\}$. Now, we see that the for such instances the CC-guarantee is at most equal to

$$\frac{(k'+1)px}{(k'+1)px + k'x} = \frac{p}{p + 1 - \frac{1}{k'+1}} = \frac{p}{p + 1 - \frac{2}{k+2}} = \frac{p}{p + \frac{k}{k+2}}.$$

The upper bound for the odd case is larger and hence prevails. $\square$

Theorem 6. *The AV-guarantee of seq-Phragmén is between* $\frac{1}{5\sqrt{k}+1}$ *and* $\frac{2}{\lfloor\sqrt{k}\rfloor} - \frac{1}{k}$.

Proof. First, we will prove the lower bound of $\frac{1}{5\sqrt{k}+1}$. Consider an approval profile A, and let W_{phrag} and W_{av} be committees winning according to seq-Phragmén and AV, respectively. W.l.o.g., we assume that $W_{\mathrm{phrag}} \neq W_{\mathrm{av}}$. For each iteration t we will use the following notation:

(1) Let $w_{\mathrm{phrag}}^{(t)}$ be the candidate selected by seq-Phragmén in the t-th iteration. Further, let $w_{\mathrm{av}}^{(t)}$ be a candidate with the highest AV-score in $W_{\mathrm{av}} \setminus \{w_{\mathrm{phrag}}^{(1)}, \ldots, w_{\mathrm{phrag}}^{(t-1)}\}$.

(2) Let $n_{\mathrm{phrag}}^{(t)} = |N(w_{\mathrm{phrag}}^{(t)})|$, and $n_{\mathrm{av}}^{(t)} = |N(w_{\mathrm{av}}^{(t)})|$.

(3) Let $\ell_j(t)$ denote the total load assigned to voter j until t. The *maximum load* in iteration t is $\max_{j \in N} \ell_j(t)$.

(4) Let $\ell_{\mathrm{av}}^{(t)}$ denote the total load distributed to the voters from $N(w_{\mathrm{av}}^{(t)})$ until iteration t, and let $m_{\mathrm{av}}^{(t)}$ denote the maximum load assigned to a voter from $N(w_{\mathrm{av}}^{(t)})$ until t, i.e., $m_{\mathrm{av}}^{(t)} = \max_{j \in N(w_{\mathrm{av}}^{(t)})} \ell_j(t)$.

We will use an argument based on a potential function $\Phi\colon [0,t] \to \mathbb{R}$, which we maintain during each iteration of seq-Phragmén. Let $\Phi(0) = 0$. In iteration t, we increase the potential function by $\big(5\sqrt{k}+1\big) \cdot n_{\mathrm{phrag}}^{(t)}$ and decrease it by $n_{\mathrm{av}}^{(t)}$, i.e.,

$$\Phi(t) = \Phi(t-1) + \big(5\sqrt{k}+1\big) \cdot n_{\mathrm{phrag}}^{(t)} - n_{\mathrm{av}}^{(t)}.$$

Our goal is to show that $\Phi(k) \geq 0$. If we know that $\Phi(k) > 0$, we can infer that

$$\sum_{t=1}^{k} \big(5\sqrt{k}+1\big) \cdot n_{\mathrm{phrag}}^{(t)} - \sum_{c \in W_{\mathrm{av}}} |N(c)| \geq \sum_{t=1}^{k} \big(5\sqrt{k}+1\big) \cdot n_{\mathrm{phrag}}^{(t)} - \sum_{t=1}^{k} n_{\mathrm{av}}^{(t)} = \Phi(k) \geq 0.$$

and hence the AV-guarantee of seq-Phragmén is lower-bounded by $\frac{1}{5\sqrt{k}+1}$.

Let s be the first iteration where $\ell_{\mathrm{av}}^{(s)} > 3\sqrt{k}$; if $\ell_{\mathrm{av}}^{(t)} \leq 3\sqrt{k}$ for all $t \in [k]$ then we set $s = k+1$.

First, let us consider iterations $t < s$ and show that $\Phi(t) \geq \Phi(t-1) + n_{\mathrm{phrag}}^{(t)} \cdot 2\sqrt{k}$. If $w_{\mathrm{phrag}}^{(t)} = w_{\mathrm{av}}^{(t)}$, then $\Phi(t) = \Phi(t-1) + \big(5\sqrt{k}\big) \cdot n_{\mathrm{phrag}}^{(t)}$. Let us assume $w_{\mathrm{phrag}}^{(t)} \neq w_{\mathrm{av}}^{(t)}$. We first show that $m_{\mathrm{av}}^{(t)} \leq \frac{\ell_{\mathrm{av}}^{(t)}+1}{n_{\mathrm{av}}^{(t)}}$. For the sake of contradiction assume that t is the first iteration after which $m_{\mathrm{av}}^{(t)} > \frac{\ell_{\mathrm{av}}^{(t)}+1}{n_{\mathrm{av}}^{(t)}}$. First note that this is only possible if indeed $w_{\mathrm{av}}^{(t)} \neq w_{\mathrm{phrag}}^{(t)}$. However, by selecting $w_{\mathrm{av}}^{(t)}$ instead of $w_{\mathrm{phrag}}^{(t)}$, it can be ensured that the load does not increase above $\frac{\ell_{\mathrm{av}}^{(t)}+1}{n_{\mathrm{av}}^{(t)}}$, so seq-Phragmén would have chosen $w_{\mathrm{av}}^{(t)}$, a contradiction. Next, observe that after $w_{\mathrm{phrag}}^{(t)}$ has been selected, the largest load assigned in total to a voter is at least equal to $1/n_{\mathrm{phrag}}^{(t)}$. Yet, if $w_{\mathrm{av}}^{(t)}$ were selected, then the largest total load assigned to a voter would be at most equal to $\frac{\ell_{\mathrm{av}}^{(t)}+1}{n_{\mathrm{av}}^{(t)}}$. Thus, it must hold that $\frac{\ell_{\mathrm{av}}^{(t)}+1}{n_{\mathrm{av}}^{(t)}} \geq 1/n_{\mathrm{phrag}}^{(t)}$, which is equivalent to $n_{\mathrm{av}}^{(t)} \leq n_{\mathrm{phrag}}^{(t)}(\ell_{\mathrm{av}}^{(t)} + 1)$. It follows that $n_{\mathrm{av}}^{(t)} \leq n_{\mathrm{phrag}}^{(t)}(3\sqrt{k}+1)$. Consequently, we have that

$$\Phi(t) \geq \Phi(t-1) + \left(5\sqrt{k}+1\right) \cdot n_{\mathrm{phrag}}^{(t)} - n_{\mathrm{av}}^{(t)} \tag{4}$$

$$\geq \left(5\sqrt{k}+1\right) \cdot n_{\mathrm{phrag}}^{(t)} - \left(3\sqrt{k}+1\right) \cdot n_{\mathrm{phrag}}^{(t)} = n_{\mathrm{phrag}}^{(t)} \cdot 2\sqrt{k}. \tag{5}$$

Now, we bound $\Phi(s-1)$. Let $w = w_{\mathrm{av}}^{(s+1)}$, i.e., let w be a candidate with the highest AV-score contained in $W_{\mathrm{av}} \setminus \{w_{\mathrm{phrag}}^{(1)}, \ldots, w_{\mathrm{phrag}}^{(s)}\}$; let $n_w = |N(w)|$. Here, we divide our reasoning into the following sequence of claims:

(1) Observe that in step s, a candidate other than w is selected by seq-Phragmén and selecting candidate w would increase the maximum load by at most $1/n_w$. As a consequence, in each iteration $t \leq s$, the maximum load increased by at most $1/n_w$.

(2) We will show that the following holds: if the maximum load in $N(w)$ increases by at least $2/n_w$ between two iterations t_1 and $t_2 \leq s$, then the AV-score from voters in $N(w)$ increased between these two iterations by at least $\frac{n_w}{2}$. Towards a contradiction, assume that this is not the case, i.e., that between t_1 and t_2 the maximum load from voter in $N(w)$ increases by at least $2/n_w$, and the load of more than $n_w/2$ voters in $N(w)$ does not increase. Without loss of generality, assume that t_2 is the first iteration for which our assumption holds. Then, if in t_2 we selected w and distributed its load among these more than $n_w/2$ voters whose load has not yet increased, then the maximum load would increase by less than $2/n_w$. This contradicts the fact that seq-Phragmén does not choose w (by definition of w).

(3) Let us group the iterations of seq-Phragmén before s into blocks. The i-th block starts after the $(i-1)$-th block ends (the first block starts with the first iteration). Further, each block ends right after the first iteration which increases the maximum load assigned to a voter from $N(w)$ by at least $2/n_w$ since the moment the block has started (thus, the last iterations may not be part of a block). Thus, in each block the maximum load assigned to a voter from $N(w)$ increases by at least $2/n_w$. Since in one step the load can increase by no more than $1/n_w$, in each block the maximum load assigned to a voter from $N(w)$ increases by at most $2/n_w + 1/n_w = 3/n_w$. Consequently, since $\ell_{\mathrm{av}}^{(s)} > 3\sqrt{k}$ (and so, by the pigeonhole principle, some voter from $N(w)$ is assigned the load at least equal to $\frac{3\sqrt{k}}{n_w}$), until s there are at least $\sqrt{k}$ blocks. By the previous point, the total AV-score of voters increases in each block by at least $n_w/2$. Since there are at least $\sqrt{k}$ blocks, we have that

$$\sum_{t=1}^{s-1} n_{\mathrm{phrag}}^{(t)} \geq \sqrt{k} \cdot n_w/2.$$

By Equation (5), we have that

$$\Phi(s-1) \geq \sqrt{k} \cdot n_w/2 \cdot 2\sqrt{k} = kn_w.$$

By choice of w, candidates not contained in W_{phrag} are approved by at most n_w voters and hence $\Phi(k) - \Phi(s-1) \geq -kn_w$. Hence $\Phi(k) \geq 0$. This concludes the lower bound proof.

For the upper bound we observe that seq-Phragmén satisfies the lower quota property [11] and use Proposition 3. This completes the proof. $\square$

Theorem 7. *The CC-guarantee of seq-Phragmén is between $\frac{1}{2}$ and $\frac{1}{2} + \frac{1}{4k-2}$.*

Proof. We first prove the lower bound on the CC-guarantee of seq-Phragmén. Consider an approval profile A, and let W_{phrag} be a committee selected by seq-Phragmén for A; let W_{cc} be a committee maximizing the CC-score for A. Further, for each i, $1 \leq i \leq k$, by $W_{\text{phrag}}^{(i)}$ we denote the first i candidates selected by seq-Phragmén. We set $n_{\text{cc}} = |N(W_{\text{cc}})|$ and $n_{\text{phrag}}^{(i)} = |N(W_{\text{phrag}}^{(i)})|$.

We will show by induction that for each i it holds that $n_{\text{phrag}}^{(i)} \geq \frac{i \cdot n_{\text{cc}}}{k+i}$. For $i = 0$, the base step of the induction is trivially satisfied. Now, assume that for some i we have $n_{\text{phrag}}^{(i)} \geq \frac{i \cdot n_{\text{cc}}}{k+i}$, and we consider the $(i+1)$-th step of seq-Phragmén. Observe that there exists a not-yet selected candidate c who is supported by at least $\frac{n_{\text{cc}} - n_{\text{phrag}}^{(i)}}{k}$ voters who do not have yet a representative in $W_{\text{phrag}}^{(i)}$. Consider the following two cases:

Case 1: c is not selected in the $(i+1)$-th step. After this step the maximum load assigned to a voter is at least equal to $\frac{i+1}{n_{\text{phrag}}^{(i+1)}}$, which is the number of chosen candidates divided by the number of voters that share their load. By selecting c the load would increase to no more than $\frac{k}{n_{\text{cc}} - n_{\text{phrag}}^{(i+1)}}$. Consequently, we have that $\frac{k}{n_{\text{cc}} - n_{\text{phrag}}^{(i+1)}} \geq \frac{i+1}{n_{\text{phrag}}^{(i+1)}}$. This is equivalent to $n_{\text{phrag}}^{(i+1)} \geq \frac{(i+1)n_{\text{cc}}}{k+i+1}$.

Case 2: c is selected in the $(i+1)$-th step. Then, $n_{\text{phrag}}^{(i+1)} \geq n_{\text{phrag}}^{(i)} + \frac{n_{\text{cc}} - n_{\text{phrag}}^{(i)}}{k}$. After reformulating:

$$n_{\text{cc}} - n_{\text{phrag}}^{(i+1)} \leq n_{\text{cc}} - n_{\text{phrag}}^{(i)} - \frac{n_{\text{cc}} - n_{\text{phrag}}^{(i)}}{k} = (n_{\text{cc}} - n_{\text{phrag}}^{(i)}) \cdot \frac{k-1}{k}.$$

By the inductive assumption we have $n_{\text{cc}} - n_{\text{phrag}}^{(i)} \leq n_{\text{cc}} - \frac{n_{\text{cc}} i}{k+i} = \frac{n_{\text{cc}} k}{k+i}$ and

$$n_{\text{cc}} - n_{\text{phrag}}^{(i+1)} \leq \frac{n_{\text{cc}} k}{k+i} \cdot \frac{k-1}{k} = \frac{n_{\text{cc}}(k-1)}{k+i}.$$

Consequently,

$$n_{\text{phrag}}^{(i+1)} \geq n_{\text{cc}} - \frac{n_{\text{cc}}(k-1)}{k+i} = \frac{n_{\text{cc}}(i+1)}{k+i} \geq \frac{n_{\text{cc}}(i+1)}{k+i+1}.$$

In both cases the inductive step is satisfied, which shows that our hypothesis holds. In particular, for $i = k$, we have that $n_{\text{phrag}}^{(k)} \geq \frac{k n_{\text{cc}}}{k+k} = \frac{n_{\text{cc}}}{2}$. This proves the lower bound on the CC-guarantee of seq-Phragmén.

For the upper bound we use the same construction and argument as in the proof of Theorem 2. $\square$

Proposition 8. *AV, CC, PAV, and p-geometric satisfy efficiency.*

Proof. Let $\mathcal{R} \in \{\text{AV}, \text{CC}, \text{PAV}, p\text{-geometric}\}$. For the sake of contradiction let us assume that there exists $k \in \mathbb{N}$, profile $A \in \mathcal{A}$, and a committee $W \in \mathcal{S}_k(C)$ such that W dominates each committee from $\mathcal{R}(A, k)$. In particular, this means that W has strictly lower score than some committee $W_{\text{opt}} \in \mathcal{R}(A, k)$. Thus, there exists a voter $i \in N$ that assigns to W_{opt} a higher score than to W. However, this is not possible since for each of the considered rules the score that i assigns to a committee W' is an increasing function of $|W' \cap A(i)|$, a contradiction. $\square$