Autonomous Agents and Multi-Agent Systems (2022) 36:1
https://doi.org/10.1007/s10458-021-09530-w

Evaluating approval-based multiwinner voting in terms of robustness to noise

Ioannis Caragiannis[1] · **Christos Kaklamanis**[2] · **Nikos Karanikolas**[3] · **George A. Krimpas**[3]

Accepted: 1 September 2021 / Published online: 9 November 2021

Abstract

Approval-based multiwinner voting rules have recently received much attention in the Computational Social Choice literature. Such rules aggregate approval ballots and determine a winning committee of alternatives. To assess effectiveness, we propose to employ new noise models that are specifically tailored for approval votes and committees. These models take as input a ground truth committee and return random approval votes to be thought of as noisy estimates of the ground truth. A minimum robustness requirement for an approval-based multiwinner voting rule is to return the ground truth when applied to profiles with sufficiently many noisy votes. Our results indicate that approval-based multiwinner voting can indeed be robust to reasonable noise. We further refine this finding by presenting a hierarchy of rules in terms of how robust to noise they are.

Keywords Computational social choice · Approval-based voting · Multiwinner voting rules · Noise models

---

A preliminary version of this paper appeared in Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI), pages 74–80, 2020.

✉ Ioannis Caragiannis
iannis@cs.au.dk

Christos Kaklamanis
kakl@ceid.upatras.gr

Nikos Karanikolas
nkaranik@ceid.upatras.gr

George A. Krimpas
krimpas@ceid.upatras.gr

1 Department of Computer Science, Aarhus University, Åbogade 34, 8200 Aarhus N, Denmark

2 Department of Computer Engineering and Informatics, Computer Technology Institute "Diophantus", University of Patras, 26504 Rion-Patras, Greece

3 Department of Computer Engineering and Informatics, University of Patras, 26504 Rion-Patras, Greece

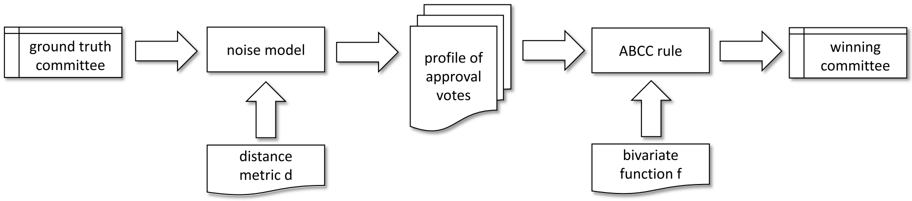

Fig. 1 Our evaluation framework

1 Introduction

Voting has received much attention from the AI and Multiagent Systems community recently, mostly due to its suitability for simple and effective decision making. One popular line of research, that originates from Arrow [1], has aimed to characterize voting rules in terms of the *social choice axioms* they satisfy. Another approach views voting rules as *estimators*. It assumes that there is an objectively correct choice, a *ground truth*, and votes are noisy estimates of it. Then, the main criterion for evaluating a voting rule is whether it can determine the ground truth as outcome when applied to noisy votes.

A typical scenario in studies that follow the second approach employs a hypothetical *noise model* that uses the ground truth as input and produces random votes. Then, a voting rule is applied on profiles of such random votes and is considered effective if it acts as a *maximum likelihood estimator* [11, 31] or if it has *low sample complexity* [9]. Since such evaluations are heavily dependent on the specifics of the noise model, it is often more informative to use relaxed effectiveness requirements. For example, the requirement for *accuracy in the limit* is sought after in a broad class of noise models [9].

We restrict our attention to *approval voting*, where ballots are simply sets of alternatives that are approved by the voters [21]. Furthermore, we consider *multiwinner voting rules* [14], which determine committees of alternatives as outcomes [3, 18]. In particular, we focus on approval-based counting choice rules (or, simply, *ABCC rules*), which were defined recently by Lackner and Skowron [20]. A famous rule in this category is known as multiwinner approval voting (AV). Each alternative gets a point every time it appears in an approval vote and the outcome consists of a fixed number of alternatives with the highest scores.

We consider noise models that are particularly tailored for approval votes and committees. These models use a committee as ground truth and produce random sets of alternatives as votes. We construct broad classes of noise models that share a particular structure, parameterized by *distance metrics* defined over sets of alternatives. In this way, we adapt to approval-based multiwinner voting the approach of Caragiannis et al. [9] for voting rules over rankings.

Figure 1 illustrates our evaluation framework. The noise model is depicted at the left. It takes as input the ground truth committee and its probability distribution over approval votes which is consistent to a distance metric d. Repeated executions of the noise model produce a profile of random approval votes. The ABCC rule (defined using a bivariate function f; see Sect. 2) is then applied on this profile and returns one or more winning committees. Our requirement for the ABCC rule is to be accurate in the limit (informally, on profiles with infinitely many votes, it must return the ground truth as the unique winning committee), not only for a single noise model, but for all models that belong to a sufficiently broad class. The breadth of this class quantifies the robustness of the ABCC rule to noise.

The details of our framework are presented in Sect. 2. Our results indicate that it indeed allows for a classification of ABCC rules in terms of their robustness to noise. In particular, we identify (in Sect. 3) the modal committee rule (MC) as the ultimately robust ABCC rule: MC is robust against all kinds of reasonable noise. AV follows in terms of robustness and seems to outperform other known ABCC rules (see Sect. 4). In contrast, the well-known approval Chamberlin-Courant (CC) rule is the least robust. On the other hand, all ABCC rules are robust if we restrict noise sufficiently (see Sect. 5). We conclude with a discussion on extensions and open problems in Sect. 6.

1.1 Further related work

Approval-based multiwinner voting rules have been studied in terms of their computational complexity [2, 29], axiomatic properties [3, 20, 27], as well as their applications [6]. In particular, axiomatic work has focused on three different principles that govern multiwinner rules: proportionality, diversity and individual excellence [14]. Lackner and Skowron [19] attempt a quantification of how close an approval-based multiwinner voting rule is to diversity and individual excellence. We remark that the primary focus of the current paper is on individual excellence, since a ground truth committee can be interpreted as the "excellent" choice in a very natural way.

The robustness of approval voting has been previously evaluated against noise models, using either the MLE [25] or the sample complexity [7] approach. These papers assume a ranking of the alternatives as ground truth, generate approval votes that consist of the top alternatives in rankings produced according to the noise model of Mallows [23], and assess how well approval voting recovers the ground truth ranking. We believe that our framework is fairer to approval votes, as recovering an underlying ranking when voters have very limited power to rank is very demanding. The robustness of (non-approval) multiwinner voting against noise has been studied by Procaccia et al. [26].

A different notion of robustness has been considered in a series of papers, starting with the work of Shiryaev et al. [28], who focused on single-winner voting. The main questions there are related to quantifying how small changes in the profile affect the voting outcome. Follow-up work in this direction studies the robustness of multi-winner voting rules, using the concept of the robustness-radius [5, 24]. Gawron and Faliszewski [16] adapt this concept to approval-based multiwinner voting, studying several ABCC rules.

Additional references related to specific ABCC rules are given in the next section. We remark that the modal committee (MC) rule is similar in spirit to the modal ranking rule considered by Caragiannis et al. [8] and the Perfectionist rule considered by Faliszewski et al. [15].

2 Preliminaries

Throughout the paper, we denote by A the set of alternatives. We use $m = |A|$ and denote the committee size by k. The term committee refers to a set of exactly k alternatives. We use n to denote the number of votes in a profile. We often use $[n]$ to represent the set of integers $\{1, \dots, n\}$.

2.1 Approval-based multiwinner voting

An approval vote is a subset $S_i \subseteq A$ of any size. An approval-based multiwinner voting rule takes as input a profile $\Pi = (S_i)_{i\in[n]}$ of n votes and returns one or more winning committees. We particularly consider voting rules that belong to the class of approval-based counting choice rules (or, simply, ABCC rules), introduced by Lackner and Skowron [20]. When applied on a profile of approval votes, such a rule computes a score for each committee (i.e., for every set of k alternatives). Then, winning committees are those having maximum score. ABCC rules compute the score of a committee as the sum of contributions of each approval vote. The contribution of an approval vote to the score of a committee depends on the size of the approval vote and the number of common alternatives between the approval vote and the committee. This is the general template that every ABCC rule follows.

More specifically, an ABCC rule is defined by a bivariate function f which defines the contribution of each approval vote to the committee score. For the ABCC rule f, $f(x, y)$ indicates the non-negative score a committee gets from an approval vote containing y alternatives, x of which are common with the committee. f is non-decreasing in its first argument. Formally, f is defined on the set $\mathcal{X}_{m,k}$, which consists of all pairs (x, y) of possible values of $|U \cap S|$ and $|S|$, given that U is k-sized (as a committee) and S can be any subset of the m alternatives of A (as an approval vote). Hence, $\mathcal{X}_{m,k}$ is the set consisting of the pairs of integers (x, y) with $y = 0, 1, \dots, m$ (the possible sizes of an approval vote) and $x = \max\{k + y - m, 0\}, \dots, \min\{y, k\}$ (the possible sizes of the intersection of an approval vote with y alternatives with a committee), i.e.,

$$\mathcal{X}_{m,k} = \{(x, y) : y = 0, 1, \dots, m, x = \max\{k + y - m, 0\}, \dots, \min\{y, k\}\}.$$

We extensively use the term "the ABCC rule f" to refer to the ABCC rule that uses the bivariate function f. We denote the score that an ABCC rule f assigns to the committee U given a profile $\Pi = (S_i)_{i\in[n]}$ of n votes by $\mathrm{sc}_f(U, \Pi) = \sum_{i=1}^{n} f(|U \cap S_i|, |S_i|)$. With some abuse of notation, we often use $\mathrm{sc}_f(U, S_i) = f(|U \cap S_i|, |S_i|)$ to refer to the score U gets from vote S_i. The set

$$\operatorname*{argmax}_{U\subseteq A : |U|=k} \mathrm{sc}_f(U, \Pi)$$

denotes the set of winning committees when applying the ABCC rule f on profile Π.

Well-known ABCC rules include:

- Multiwinner approval voting (AV), which uses the function $f_{\mathrm{AV}}(x, y) = x$.
- Approval Chamberlin-Courant (CC), which uses the function $f_{\mathrm{CC}}(x, y) = \min\{1, x\}$. The rule falls within a more general context considered by Chamberlin and Courant [10].
- Proportional approval voting (PAV), which uses the function $f_{\mathrm{PAV}}(x, y) = \sum_{i=1}^{x} 1/i$.

These rules belong to the class of rules that originate from the work of Thiele [30]. A Thiele rule uses a vector $\langle w_1, w_2, \dots, w_k \rangle$ of non-negative weights to define $f(x, y) = \sum_{i=1}^{x} w_i$. Other known Thiele rules include the p-Geometric rule [29] and Sainte Laguë approval voting [19].

A well-known non-Thiele rule is the satisfaction approval voting (SAV) rule that uses $f_{\mathrm{SAV}}(x, y) = x/y$ for $y > 0$ and $f(x, y) = 0$ otherwise [4]. Let us also introduce the *modal committee* (MC) rule which returns the committee (or committees) that has maximum number of appearances as approval votes in the profile. Recall that k is the committee size. Then, MC uses the bivariate function f defined by $f(k,k)=1$ and $f(x,y)=0$ for $(x,y)\neq(k,k)$. In words, a committee U gets one point from each approval vote that is identical to U. Clearly, MC is non-Thiele. Sometimes, we use the term *trivial* ABCC rule to refer to an ABCC rule that uses the fixed bivariate function $f(x,y)=c$ for every $(x,y)\in\mathcal{X}_{m,k}$. Clearly, in any profile, any k-sized set of alternatives is a winning committee under the trivial ABCC rule.

We remark that, formally, an ABCC rule is defined for a given number of alternatives m and a given committee size k. However, most of the ABCC rules we consider later in the paper, such as MC, AV, CC, or the trivial rule, are essentially *families* of rules defined for the whole range of m and k such that $1\leq k<m$. Our positive statements hold for general values of these parameters and, as such, they hold for families of ABCC rules. Some of our negative statements involving counter-examples are proved for specific values of m and k (this is enough to show that some statement is not true; e.g., see Example 1 and Theorem 5).

2.2 Noise models

We employ noise models to generate approval votes, assuming that the ground truth is a committee. Denoting the ground truth by $U\subseteq A$, a noise model $\mathcal{M}$ produces random approval votes according to a particular distribution that defines the probability $\Pr_{\mathcal{M}}[S|U]$ to generate the set $S\subseteq A$ when the ground truth is U.

Let us give the following noise model $\mathcal{M}_p$ as an example. $\mathcal{M}_p$ uses a parameter $p\in(1/2,1]$. Given a ground truth committee U, $\mathcal{M}_p$ generates a random set $S\subseteq A$ by selecting each alternative of U with probability p and each alternative in $A\backslash U$ with probability $1-p$.[1] Intuitively, the probability that a set will be generated depends on its "distance" from the ground truth: the higher this distance, the smaller this probability. To make this formal, let us define the notion of a *distance metric* over sets of alternatives.

Definition 1 A function $d:2^A\times 2^A\rightarrow\mathbb{R}_{\geq 0}$ is a distance metric over subsets of the set of alternatives A if

- $d(X,Y)=0$ if and only if $X=Y$ (*identity of indiscernibles*),
- $d(X,Y)=d(Y,X)$ for every pair of sets of alternatives $X,Y\subseteq A$ (*symmetry*), and
- $d(X,Y)\leq d(X,Z)+d(Z,Y)$ for every three sets of alternatives $X,Y,Z\subseteq A$ (*triangle inequality*).

[1] Even though it might look as a toy example of a noise model, a more careful look will reveal that $\mathcal{M}_p$ can be seen as the analog of the famous Mallows noise model [23] in the classical social choice setting when each voter provides a strict ranking of the alternatives instead of an approval set. Like in the Mallows model, the parameter p is the probability of "getting it right", i.e., of including in the random set alternatives of the gound truth committee, and leaving out alternatives not belonging to it. Interestingly, the ABCC rule AV turns out to be a maximum likelihood estimator for $\mathcal{M}_p$ (analogously to the fact that the well-known Kemeny rule is an MLE for Mallows; e.g., see [31]). As this is beyond the scope of the current paper, we present a proof in "Appendix".

Our first example of a distance metric is the *set difference*[2] distance metric d_Δ, defined as $d_\Delta(X,Y)=|X\setminus Y|+|Y\setminus X|$. The next lemma provides a strong relation between set difference d_Δ and the noise model $\mathcal{M}_p$.

Lemma 1 *For* $S\subseteq A$, $\Pr_{\mathcal{M}_p}[S|U]=p^m\cdot\left(\frac{1-p}{p}\right)^{d_\Delta(U,S)}$.

Proof Let U be a set of k alternatives. By the definition of the noise model $\mathcal{M}_p$, the set S is generated by the noise model $\mathcal{M}_p$ with ground truth committee U when each alternative in $S\cap U$ is selected (this happens with probability p, independently for each alternative of the set), each alternative in $S\setminus U$ is selected (probability $1-p$ each), each alternative in $U\setminus S$ is not selected (probability $1-p$ each), and each alternative in $A\setminus(S\cup U)$ is not selected (probability p each). Overall,

$$\begin{aligned}\Pr_{\mathcal{M}_p}[S|U]&=p^{|S\cap U|}\cdot(1-p)^{|S\setminus U|}\cdot(1-p)^{|U\setminus S|}\cdot p^{|A\setminus(S\cup U)|}\\&=p^{m-d_\Delta(U,S)}\cdot(1-p)^{d_\Delta(U,S)}=p^m\cdot\left(\frac{1-p}{p}\right)^{d_\Delta(U,S)}\end{aligned}$$

as desired. □

So, since $p>1/2$, the probability $\Pr_{\mathcal{M}_p}[S|U]$ is decreasing in $d_\Delta(U,S)$ and gets its maximum value when $S=U$. We will consider general noise models $\mathcal{M}$ with $\Pr_{\mathcal{M}}[S|U]$ depending on $d(U,S)$, where d is a distance metric defined over subsets of A.

Definition 2 Let d be a distance metric over sets of alternatives. A noise model $\mathcal{M}$ is called *d-monotonic* if for any two sets $S_1,S_2\subseteq A$, it holds $\Pr_{\mathcal{M}}[S_1|U]>\Pr_{\mathcal{M}}[S_2|U]$ if and only if $d(U,S_1)<d(U,S_2)$.

Thus, the notion of d-monotonicity requires that sets of alternatives that are closer to the ground truth according to the distance metric d have higher probability to be generated. Definition 2 implies that $\Pr_{\mathcal{M}}[S_1|U]=\Pr_{\mathcal{M}}[S_2|U]$ if and only if $d(U,S_1)=d(U,S_2)$. Also, together with Definition 1 (and, in particular, using the identity of indiscernibles property), we have that $\Pr_{\mathcal{M}}[U|U]>\Pr_{\mathcal{M}}[S|U]$, for every set $S\subseteq A$ different than U.

Besides the set difference metric used by $\mathcal{M}_p$, other well-known distance metrics (see [12]) are:

- the normalized set difference or *Jaccard* distance metric d_J, defined as $d_J(X,Y)=\frac{d_\Delta(X,Y)}{|X\cup Y|}$,
- the maximum difference or *Zelinka* metric d_Z, defined as $d_Z(X,Y)=\max\{|X\setminus Y|,|Y\setminus X|\}$, and
- the normalized maximum difference or *Bunke-Shearer* metric d_{BS}, defined as $d_{BS}(X,Y)=\frac{d_Z(X,Y)}{\max\{|X|,|Y|\}}$.

[2] Viewing sets as binary strings, the distance metric d_Δ is equivalent to the Hamming distance; see Deza and Deza [12].

The interested reader may verify that they (as well as set difference) are indeed distance metrics. Among them, the set difference and the Jaccard distance metrics are the most popular ones, with many applications in statistics and data analysis; e.g., see Sect. 3 in the book by Leskovec et al. [22].

In our proofs (e.g., in the proof of Theorem 1 and in Example 2), we often define simple distance metrics d by making sure that $d(X, Y) = 0$ if and only if $X = Y$, $d(X, Y) = d(Y, X)$ for every pair of sets of alternatives $X, Y \subseteq A$, and setting $d(X, Y)$ (arbitrarily) to either 1 or 2 if $X \neq Y$. To verify that d is always a distance metric, we need to verify that the triangle inequality $d(X, Y) \leq d(X, Z) + d(Z, Y)$ holds. This is obvious if $X = Y$. If $Z = X$ or $Z = Y$, the inequality holds with equality. If the three sets X, Y, and Z are different, the RHS of the inequality has value at least 2, while $d(X, Y) \leq 2$ by definition.

We remark that such a distance metric may satisfy, e.g., $d(\{a, b\}, \{a, c\}) \neq d(\{b, c\}, \{b, d\})$ for four alternatives a, b, c, and d. This cannot happen with the following class of distance metrics, where renaming the alternatives does not change the distance.

Definition 3 A distance metric d is called *alternative-independent* if $d(X, Y)$ depends only on the sizes of the sets X, Y, $X \setminus Y$, and $Y \setminus X$.

It can be easily seen that the four metric distances defined above are all alternative-independent. Our results apply to the most general definition of *alternative-dependent* distances, where $d(X, Y)$ can also depend on the contents of sets $X \setminus Y$, $Y \setminus X$, X, and Y.

2.3 Evaluating ABCC rules against noise models

We aim to evaluate the effectiveness of ABCC rules when applied to random profiles generated by large classes of noise models. To this end, we use *accuracy in the limit* as a measure.

Definition 4 *(accuracy in the limit)* An ABCC rule f is called *accurate in the limit* for a noise model $\mathcal{M}$ if for every $\varepsilon > 0$ there exists n_ε such that, for every profile with at least n_ε approval votes produced by $\mathcal{M}$ with ground truth U, f returns U as the unique winning committee with probability at least $1 - \varepsilon$.

Then, ABCC rules are evaluated in terms of *robustness* using the next definition.

Definition 5 *(robustness)* Let d be a distance metric over sets of alternatives. An ABCC rule f is *monotone robust* against d (or d-monotone robust) if it is accurate in the limit for all d-monotonic noise models.

We remark that even though we follow the standard definition according to which an ABCC rule may return more than one winning committees (see [20]), our definition of the accuracy in the limit (Definition 4) is particularly demanding and requires from the ABCC rule to return a *unique* committee with high probability. Our purpose here is to guarantee the maximum level of robustness.

3 MC is a uniquely robust ABCC rule

We begin our technical exposition by identifying the unique ABCC rule that is monotone robust against *all* distance metrics. Our proofs, in the current and subsequent sections, make extensive use of Lemma 3 below. In its proof, we use the following variant of the Hoeffding inequality.

Lemma 2 (Hoeffding [17]) *Let $X_1, X_2, \ldots, X_\ell$ be i.i.d. random variables with* $\mathbb{E}[X_i] = \mu$ and $X_i \in [a, b]$ for $i = 1, \ldots, \ell$, *and* $X = \sum_{i\in[\ell]} X_i$. *Then, for every* $t > 0$,

$$\Pr\left[|X - \ell\mu| \geq t\right] \leq 2\exp\left(-\frac{2t^2}{\ell(b-a)^2}\right).$$

We remark that the notation $S \sim \mathcal{M}(U)$ indicates that the random set S is drawn from the noise model $\mathcal{M}$ with ground truth U.

Lemma 3 *Let $\mathcal{M}$ be a noise model. An ABCC rule f is*

a. *accurate in the limit for* $\mathcal{M}$ if $\mathbb{E}_{S\sim\mathcal{M}(U)}[\mathrm{sc}_f(U,S) - \mathrm{sc}_f(V,S)] > 0$ *for every two different committees U and V.*
b. *not accurate in the limit for $\mathcal{M}$ if* $\mathbb{E}_{S\sim\mathcal{M}(U)}[\mathrm{sc}_f(U,S) - \mathrm{sc}_f(V,S)] < 0$ *for some pair of committees U and V.*

Proof We will need some additional general notation. Denote by a' and b' the minimum and maximum values of the quantity $f(x_1, y) - f(x_2, y)$ over all triplets of integers x_1, x_2, and y that define pairs $(x_1, y), (x_2, y) \in \mathcal{X}_{m,k}$. Also, for two k-sized sets of alternatives U and V, define $\mu(U, V) = \mathbb{E}_{S\sim\mathcal{M}(U)}[\mathrm{sc}_f(U,S) - \mathrm{sc}_f(V,S)]$.

Part (a). Define $\mu_{\min}$ as the minimum among all values $\mu(U, V)$ for all pairs of different k-sized sets of alternatives U and V. By the assumption of part (a) of the lemma, we have $\mu_{\min} > 0$. Let $\varepsilon > 0$ and

$$n_\varepsilon = \frac{(b'-a')^2}{2\mu_{\min}^2} \ln \frac{2m^k}{\varepsilon}.$$

We will prove part (a) by showing that for every profile $\Pi = (S_i)_{i\in[n]}$ with at least n_ε approval votes (i.e., $n \geq n_\varepsilon$) from the noise model $\mathcal{M}$ with ground truth U, the probability that rule f returns U as the unique winner is at least $1 - \varepsilon$.

First observe that

$$\mathrm{sc}_f(U,\Pi) - \mathrm{sc}_f(V,\Pi) = \sum_{i\in[n]} \left(\mathrm{sc}_f(U,S_i) - \mathrm{sc}_f(V,S_i)\right)$$

for every k-sized set of alternatives V. The quantity $\mathrm{sc}_f(U,S_i) - \mathrm{sc}_f(V,S_i)$ is a random variable following the distribution of $\mathrm{sc}_f(U,S) - \mathrm{sc}_f(V,S)$, where the set S is drawn randomly from the noise model $\mathcal{M}$ with ground truth U. Also, observe that the random variable $\mathrm{sc}_f(U,S) - \mathrm{sc}_f(V,S)$ takes values in $[a', b']$. Hence, the score difference $\mathrm{sc}_f(U,\Pi) - \mathrm{sc}_f(V,\Pi)$ is a sum of i.i.d random variables, each with expectation $\mu(U, V)$ and taking values in $[a', b']$.

Hence, we can apply Hoeffding inequality (Lemma 2) with $X = \mathrm{sc}_f(U, \Pi) - \mathrm{sc}_f(V, \Pi)$, $\ell = n$, $a = a'$, $b = b'$, $\mu = \mu(U, V)$, and $t = n\mu(U, V)$ to get that the probability that $\mathrm{sc}_f(U, \Pi) \leq \mathrm{sc}_f(V, \Pi)$ is

$$\begin{aligned}\Pr[\mathrm{sc}_f(U,\Pi) \leq \mathrm{sc}_f(V,\Pi)] &= \Pr[\mathrm{sc}_f(U,\Pi) - \mathrm{sc}_f(V,\Pi) \leq 0] \\ &\leq \Pr[|\mathrm{sc}_f(U,\Pi) - \mathrm{sc}_f(V,\Pi) - n\mu(U,V)| \geq n\mu(U,V)] \\ &\leq 2\exp\left(-\frac{2n\mu(U,V)^2}{(b'-a')^2}\right) \leq 2\exp\left(-\frac{2n_\varepsilon \mu_{\min}^2}{(b'-a')^2}\right) \leq \frac{\varepsilon}{m^k}.\end{aligned}$$

The second to last inequality follows since $n \geq n_\varepsilon$ and $\mu(U, V) \geq \mu_{\min}$ and the last one by the definition of n_ε.

So far, we have proved that the probability that the score of the k-sized set of alternatives U is not higher than the score of the k-sized set of alternatives V under f is at most ε/m^k. Hence, by applying a union bound, we have that the probability that the score of U is not higher than the score of any of the other at most m^k k-sized sets of alternatives is at most ε. In other words, U is the unique winning committee under f with probability at least $1-\varepsilon$ as the definition of the accuracy in the limit (Definition 4) requires. This completes the proof of part (a).

Part (b). We will again consider a profile $\Pi = (S_i)_{i\in[n]}$ of n approval votes from the noise model $\mathcal{M}$ with ground truth U, and show that, as n approaches infinity, the probability that the score of U under the ABCC rule f is at least as high as that of the k-sized set of alternatives V, with a probability that approaches 0. This is enough to show that f is not accurate in the limit for the noise model $\mathcal{M}$ with ground truth U. To see why, notice the following implication that clearly violates Definition 4: for $\varepsilon = 1/2$ and any definition of n_ε, there exist profiles with at least n_ε approval votes, so that the probability that U is the unique winning committee is strictly less than 1/2.

Indeed, by applying the Hoeffding inequality for the random variable $X = \mathrm{sc}_f(U, \Pi) - \mathrm{sc}_f(V, \Pi)$, using $\ell = n$, $a = a'$, $b = b'$, $\mu = \mu(U, V)$, and $t = -n\mu(U, V)$ (notice that $\mu(U, V) < 0$ now), we get

$$\begin{aligned}\Pr[\mathrm{sc}_f(U,\Pi) \geq \mathrm{sc}_f(V,\Pi)] &= \Pr[\mathrm{sc}_f(U,\Pi) - \mathrm{sc}_f(V,\Pi) \geq 0] \\ &\leq \Pr[|\mathrm{sc}_f(U,\Pi) - \mathrm{sc}_f(V,\Pi) - n\mu(U,V)| \geq -n\mu(U,V)] \\ &\leq 2\exp\left(-\frac{2n\mu(U,V)^2}{(b'-a')^2}\right),\end{aligned}$$

which approaches 0 as n approaches infinity. □

We are ready to present our first application of Lemma 3.

Theorem 1 *MC is the only ABCC rule that is monotone robust against any distance metric.*

Proof Let $\mathcal{M}$ be a noise model that is d-monotonic for some distance metric d. Let $U, V \subseteq A$ be any two different k-sized sets of alternatives. By the definition of MC, we have

$$\mathbb{E}_{S\sim\mathcal{M}(U)}[\mathrm{sc}_{\mathrm{MC}}(U,S) - \mathrm{sc}_{\mathrm{MC}}(V,S)] = \Pr_{\mathcal{M}}[U|U] - \Pr_{\mathcal{M}}[V|U] > 0.$$

By Lemma 3a, we obtain that MC is d-monotone robust.

We will now show that MC is the only ABCC rule that has this property. Clearly, the trivial ABCC rule is not monotone robust against any distance metric since, by its definition, it never returns a unique winning committee. The bivariate function f of any non-trivial ABCC rule can be assumed to be *normalized*. I.e., it can be assumed to satisfy $f(\max\{k+y-m,0\},y)=0$ (recall that $\max\{k+y-m,0\}$ is the smallest x so that (x, y) belongs to $\mathcal{X}_{m,k}$) for every integer $y=0,1,\dots,m$ and, furthermore,

$$\max_{(x,y)\in\mathcal{X}_{m,k}} f(x,y)=1.$$

This normalization assumption is valid due to the fact that for every constant $c>0$ and univariate function $e(\cdot)$ defined over the integers $0,1,\dots,m$, the ABCC rule defined

$$g(x,y)=c\cdot f(x,y)+e(y)$$

for $(x,y)\in\mathcal{X}_{m,k}$ is equivalent to the ABCC rule f, in the sense that both f and g return the same set of winning committees on every profile of approval votes [20]. Clearly, our definition of the ABCC rule MC uses a normalized bivariate function.

Now, let f be the normalized bivariate function of a non-trivial ABCC rule that is different than MC. This means that there exist integers x^* and y^* with $(x^*-1,y^*),(x^*,y^*)\in\mathcal{X}_{m,k}$, $(x^*,y^*)\neq(k,k)$, and $f(x^*,y^*)>f(x^*-1,y^*)$. We will construct a distance metric d and a d-monotonic noise model for which f is not accurate in the limit.[3]

Rename the alternatives of A as $a_1,a_2,\dots,a_m$ and let $U=\{a_1,a_2,\dots,a_k\}$, $V=\{a_2,\dots,a_{k+1}\}$, and $W=\{a_{k-x^*+2},\dots,a_{y^*+k-x^*+1}\}$. Notice that, by the definition of $\mathcal{X}_{m,k}$, $(x^*-1,y^*)\in\mathcal{X}_{m,k}$ implies that $\max\{y^*+k-m,0\}\le x^*-1$ and, hence, $y^*+k-x^*+1\le m$. Furthermore, $x^*\ge 1$. Also, $(x^*,y^*)\in\mathcal{X}_{m,k}$ implies that $x^*\le\max\{k,y^*\}$ and, hence, $k-x^*+2\ge 2$ and $x^*\le y^*$. Consequently, the set W is well-defined and has the following structure: It contains y^* alternatives. The x^* alternatives $a_{k-x^*+2},\dots,a_{k+1}$ appear also in set V and the x^*-1 alternatives $a_{k-x^*+2},\dots,a_k$ appear also in set U.

We define a distance metric d between subsets of A that has $d(X,Y)=0$ if $X=Y$, $d(X,Y)\in\{1,2\}$, otherwise, and in particular $d(U,V)=d(U,W)=1$ and $d(U,S)=2$ for every set of alternatives $S\subseteq A$ different than U, V, or W.

We are ready to define the d-monotonic noise model $\mathcal{M}$. For simplicity, we use $p_0=\Pr_{\mathcal{M}}[U|U]$, $p_1=\Pr_{\mathcal{M}}[V|U]=\Pr_{\mathcal{M}}[W|U]$, and $p_2=\Pr_{\mathcal{M}}[S|U]$ for every other set $S\subseteq A$ different than U, V, or W. For $0<\delta<\frac{1}{3(2^m-1)}$, we set $p_0=1/3$, $p_1=1/3-\delta$, and $p_2=\frac{2\delta}{2^m-3}$. The particular value of δ will be specified shortly; for the moment, the range of δ guarantees that $p_0>p_1>p_2$ so that $\mathcal{M}$ is indeed d-monotonic.

We now compute the quantity $\mathbb{E}_{S\sim\mathcal{M}(U)}[\operatorname{sc}_f(U,S)-\operatorname{sc}_f(V,S)]$; observe that $\operatorname{sc}_f(U,U)=\operatorname{sc}_f(V,V)=f(k,k)$, $\operatorname{sc}_f(U,V)=\operatorname{sc}_f(V,U)=f(k-1,k)$, $\operatorname{sc}_f(U,W)=f(x^*-1,y^*)$, and $\operatorname{sc}_f(V,W)=f(x^*,y^*)$. We obtain

[3] We remark that the distance metric d in the second part of the proof of Theorem 1 is alternative-dependent. This is necessary; see the discussion in Sect. 6.

$$
\begin{aligned}
&\mathbb{E}_{S\sim\mathcal{M}(U)}[\,\mathrm{sc}_f(U,S) - \mathrm{sc}_f(V,S)] \\
&= \mathrm{sc}_f(U,U)\cdot p_0 + \mathrm{sc}_f(U,V)\cdot p_1 + \mathrm{sc}_f(U,W)\cdot p_1 + \sum_{S\neq U,V,W} \mathrm{sc}_f(U,S)\cdot p_2 \\
&\quad - \mathrm{sc}_f(V,U)\cdot p_0 - \mathrm{sc}_f(V,V)\cdot p_1 - \mathrm{sc}_f(V,W)\cdot p_1 - \sum_{S\neq U,V,W} \mathrm{sc}_f(V,S)\cdot p_2 \\
&= f(k,k)\cdot p_0 + f(k-1,k)\cdot p_1 + f(x^*-1,y^*)\cdot p_1 + \sum_{S\neq U,V,W} f(|U\cap S|,|S|)\cdot p_2 \\
&\quad - f(k-1,k)\cdot p_0 - f(k,k)\cdot p_1 - f(x^*,y^*)\cdot p_1 - \sum_{S\neq U,V,W} f(|V\cap S|,|S|)\cdot p_2 \\
&\leq (p_0-p_1)\cdot(f(k,k)-f(k-1,k)) - p_1\cdot(f(x^*,y^*)-f(x^*-1,y^*)) + p_2\cdot \sum_{S\neq U,V,W} f(|U\cap S|,|S|) \\
&= \delta\cdot(f(k,k)-f(k-1,k)) - (1/3-\delta)\cdot(f(x^*,y^*)-f(x^*-1,y^*)) \\
&\quad + \frac{2\delta}{2^m-3}\cdot \sum_{S\neq U,V,W} f(|U\cap S|,|S|).
\end{aligned} \tag{1}
$$

Observe that the RHS of (1) is increasing in δ and approaches $-\frac{1}{3}(f(x^*,y^*)-f(x^*-1,y^*)) < 0$ as δ approaches 0. Hence, for a sufficiently small positive δ, we have

$$\mathbb{E}_{S\sim\mathcal{M}(U)}[\,\mathrm{sc}_f(U,S) - \mathrm{sc}_f(V,S)] < 0.$$

By Lemma 3b, f is not accurate in the limit for $\mathcal{M}$ and, hence, not monotone robust against the distance metric d. This completes the proof of the theorem. □

4 A characterization for AV

In this section, we identify the class of distance metrics against which AV is monotone robust. We will need some additional notation that will be useful in several proofs.

For a distance metric d and a set of alternatives U, let $\mathrm{span}(d,U)$ be the number of different non-zero values the quantity $d(U,\cdot)$ can take. We denote these different distance values by $\delta_1(d,U)$, $\delta_2(d,U)$,..., $\delta_{\mathrm{span}(d,U)}(d,U)$ and assume that $\delta_1(d,U) < \delta_2(d,U) < \ldots < \delta_{\mathrm{span}(d,U)}(d,U)$. We also use $\delta_0(d,U)=0$. For $t=0,1,\ldots,\mathrm{span}(d,U)$ and alternatives $a,b\in A$, we denote by $N^t_{a|b}(d,U)$ the class of sets S of alternatives that contain alternative a but not alternative b and satisfy $d(U,S)\leq \delta_t(d,U)$.

Definition 6 *(majority-concentricity)* A distance metric d is called *majority-concentric* if for every k-sized set of alternatives U, it holds $|N^t_{a|b}(d,U)| \geq |N^t_{b|a}(d,U)|$ for all alternatives $a\in U$ and $b\notin U$ and $t=0,1,\ldots,\mathrm{span}(d,U)$.

Majority-concentricity is similar in spirit with a property of distance metrics over rankings with the same name in [9]. Similarly to that paper, the term "concentric" comes from a visualization of the sets of alternatives as lying in concentric circles, with committee U at their center, and with the t-th circle from the center hosting the sets of alternatives at distance exactly $\delta_t(d,U)$ from U, for $t=1,2,\ldots,\mathrm{span}(d,U)$.

We are ready to prove our characterization for AV.

Theorem 2 *AV is d-monotone robust if and only if the distance metric d is majority-concentric.*

Proof Let $\mathcal{M}$ be a d-monotonic noise model for a majority concentric distance metric d. Let U and V be two different sets with k alternatives each. By Lemma 3a, in order to show that AV is accurate in the limit for $\mathcal{M}$ (and, consequently, d-monotone robust), it suffices to show that $\mathbb{E}_{S\sim\mathcal{M}(U)}[\text{ sc }_{\text{AV}}(U,S) - \text{ sc }_{\text{AV}}(V,S)] > 0$.

We will need some additional notation. For $t = 0, 1, \dots, \text{span}\,(d, U)$, we denote by $\bar{N}^t(d,U)$ the class of sets of alternatives S that satisfy $d(U,S) = \delta_t(d,U)$. For alternatives $a, b \in A$, we denote by $\bar{N}^t_a(d,U)$ the subclass of $\bar{N}^t(d,U)$ consisting of sets of alternatives that include a and by $\bar{N}^t_{a|b}(d,U)$ the subclass of $\bar{N}^t_a(d,U)$ consisting of sets that do not contain alternative b.

To simplify notation, we set $s = \text{span}\,(d,U)$. Also, we drop $(d,\ U)$ from notation (e.g., we use $N^t_{a|b}$ instead of $N^t_{a|b}(d,U)$) since it is clear from context. We have

$$\begin{aligned}\mathbb{E}_{S\sim\mathcal{M}(U)}[\text{ sc }_{\text{AV}}(U,S)] &= \sum_{S\subseteq A} \text{sc }_{\text{AV}}(U,S)\cdot \text{Pr}_{\mathcal{M}}[S|U] = \sum_{S\subseteq A} |U\cap S|\cdot \text{Pr}_{\mathcal{M}}[S|U] \\ &= \sum_{a\in U}\sum_{S\subseteq A:a\in S} \text{Pr}_{\mathcal{M}}[S|U] = \sum_{a\in U}\sum_{t=0}^{s}\sum_{S\in\bar{N}^t_a} \text{Pr}_{\mathcal{M}}[S|U].\end{aligned} \tag{2}$$

Now, observe that the probability $\text{Pr}_{\mathcal{M}}[S|U]$ is the same for all sets $S \in \bar{N}^t$, since all of them have the same distance $\delta_t(d,U)$ from U. In the following, we use $p_t = \text{Pr}_{\mathcal{M}}[S|U]$ for all $S \in \bar{N}^t$, for $t = 0, 1, \dots, s$. Hence, (2) becomes

$$\mathbb{E}_{S\sim\mathcal{M}(U)}[\text{ sc }_{\text{AV}}(U,S)] = \sum_{a\in U}\sum_{t=0}^{s} |\bar{N}^t_a|\cdot p_t.$$

Similarly, we have

$$\mathbb{E}_{S\sim\mathcal{M}(U)}[\text{ sc }_{\text{AV}}(V,S)] = \sum_{a\in V}\sum_{t=0}^{s} |\bar{N}^t_a|\cdot p_t,$$

and, by linearity of expectation,

$$\mathbb{E}_{S\sim\mathcal{M}(U)}[\text{ sc }_{\text{AV}}(U,S) - \text{ sc }_{\text{AV}}(V,S)] = \sum_{a\in U\setminus V}\sum_{t=0}^{s} |\bar{N}^t_a|\cdot p_t - \sum_{a\in V\setminus U}\sum_{t=0}^{s} |\bar{N}^t_a|\cdot p_t. \tag{3}$$

Let μ be a bijection that maps each alternative of $U\setminus V$ to a distinct alternative of $V\setminus U$. Then, (3) becomes

$$
\begin{aligned}
&\mathbb{E}_{S\sim\mathcal{M}(U)}[\,\mathrm{sc}_{\,\mathrm{AV}}(U,S)-\,\mathrm{sc}_{\,\mathrm{AV}}(V,S)]\\
&=\sum_{a\in U\setminus V}\sum_{t=0}^{s}|\bar{N}^t_a|\cdot p_t-\sum_{a\in U\setminus V}\sum_{t=0}^{s}|\bar{N}^t_{\mu(a)}|\cdot p_t=\sum_{a\in U\setminus V}\sum_{t=0}^{s}\left(|\bar{N}^t_{a|\mu(a)}|-|\bar{N}^t_{\mu(a)|a}|\right)\cdot p_t\\
&=\sum_{a\in U\setminus V}\left(|N^0_{a|\mu(a)}|-|N^0_{\mu(a)|a}|\right)\cdot p_0+\sum_{a\in U\setminus V}\sum_{t=1}^{s}\left(|N^t_{a|\mu(a)}|-|N^{t-1}_{a|\mu(a)}|-|N^t_{\mu(a)|a}|+|N^{t-1}_{\mu(a)|a}|\right)\cdot p_t\\
&=\sum_{a\in U\setminus V}\left(\sum_{t=0}^{s-1}\left(|N^t_{a|\mu(a)}|-|N^t_{\mu(a)|a}|\right)\cdot(p_t-p_{t+1})+\left(|N^s_{a|\mu(a)}|-|N^s_{\mu(a)|a}|\right)\cdot p_s\right)\\
&\geq\sum_{a\in U\setminus V}\left(|N^0_{a|\mu(a)}|-|N^0_{\mu(a)|a}|\right)\cdot(p_0-p_1)>0.
\end{aligned}
\tag{4}
$$

The third equality follows since $\bar{N}^0_{a|\mu(a)}=N^0_{a|\mu(a)}$, $\bar{N}^0_{\mu(a)|a}=N^0_{\mu(a)|a}$, and $\bar{N}^t_{a|\mu(a)}=N^t_{a|\mu(a)}\setminus N^{t-1}_{a|\mu(a)}$ and $\bar{N}^t_{\mu(a)|a}=N^t_{\mu(a)|a}\setminus N^{t-1}_{\mu(a)|a}$ for $t=1,\dots,s$. The first inequality follows since d is majority concentric and since $p_t>p_{t+1}$ and, thus, all differences in (4) are non-negative. The last inequality follows after observing that $|N^0_{a|\mu(a)}|=1$ and $|N^0_{\mu(a)|a}|=0$ for $a\in U\setminus V$ and since $p_0>p_1$. This completes the "if" part of the proof.

Let us now consider a non-majority concentric distance metric d that satisfies $|N^{t^*}_{a|b}(d,U)|<|N^{t^*}_{b|a}(d,U)|$ for the k-sized set of alternatives U, some alternatives $a\in U$ and $b\notin U$, and some $t^*\in\{1,2,\dots,\mathrm{span}\,(d,U)-1\}$. We show the "only if" part of the theorem by constructing a noise model $\mathcal{M}$ that satisfies $\mathbb{E}_{S\sim\mathcal{M}(U)}[\,\mathrm{sc}_{\,\mathrm{AV}}(U,S)-\,\mathrm{sc}_{\,\mathrm{AV}}(V,S)]<0$ for $V=(U\setminus\{a\})\cup\{b\}$.

Again, we use $p_t=\mathrm{Pr}_{\mathcal{M}}[S|U]$ for every set of alternatives $S\in\bar{N}^t(d,U)$ for $t=0,1,\dots,\mathrm{span}\,(d,U)$, $s=\mathrm{span}\,(d,U)$, and drop $(d,\ U)$ from notation. We define the model probabilities so that $\tau=p_0>p_1>\dots>p_{t^*}=\tau-\epsilon$ and $2\epsilon=p_{t^*+1}>\dots>p_s=\epsilon$. Notice that such a noise model exists for any arbitrarily small $\epsilon>0$. Since there are 2^m sets of alternatives and τ is the probability that $\mathcal{M}$ returns the ground truth ranking, it must be $\tau>1/2^m$. We now apply equality (4). Observe that, since $V=(U\setminus\{a\})\cup\{b\}$, we have $\mu(a)=b$. We obtain

$$
\begin{aligned}
&\mathbb{E}_{S\sim\mathcal{M}(U)}[\,\mathrm{sc}_{\,\mathrm{AV}}(U,S)-\,\mathrm{sc}_{\,\mathrm{AV}}(V,S)]=\sum_{t=0}^{s-1}\left(|N^t_{a|b}|-|N^t_{b|a}|\right)\cdot(p_t-p_{t+1})+\left(|N^s_{a|b}|-|N^s_{b|a}|\right)\cdot p_s\\
&=\sum_{t=0}^{t^*-1}\left(|N^t_{a|b}|-|N^t_{b|a}|\right)\cdot(p_t-p_{t+1})+\left(|N^{t^*}_{a|b}|-|N^{t^*}_{b|a}|\right)\cdot(p_{t^*}-p_{t^*+1})\\
&\quad+\sum_{t=t^*+1}^{s-1}\left(|N^t_{a|b}|-|N^t_{b|a}|\right)\cdot(p_t-p_{t+1})+\left(|N^s_{a|b}|-|N^s_{b|a}|\right)\cdot p_s.
\end{aligned}
$$

Now, observe that for $t\neq t^*$, it holds $|N^t_{a|b}|-|N^t_{b|a}|\leq 2^m$ (the total number of sets of alternatives) and $p_t-p_{t+1}\leq\epsilon$. Also, $|N^{t^*}_{a|b}|-|N^{t^*}_{b|a}|\leq -1$ and $p_{t^*}-p_{t^*+1}=\tau-3\epsilon$. Setting specifically $\epsilon=\frac{1}{s8^m}$, we obtain that

$$
\mathbb{E}_{S\sim\mathcal{M}(U)}[\,\mathrm{sc}_{\,\mathrm{AV}}(U,S)-\,\mathrm{sc}_{\,\mathrm{AV}}(V,S)]\leq s2^m\epsilon-(\tau-3\epsilon)\leq\frac{1}{4^m}-\frac{1}{2^m}+\frac{3}{s\cdot 8^m},
$$

which is negative for $m\geq 2$ since $s\geq 1$. The proof of the "only if" part of Theorem 2 now follows by Lemma 3b. ∎

It is tempting to conjecture that AV and MC are the only ABCC rules that are monotone robust against all majority concentric distance metrics. However, this is not true as the next example, which uses a different ABCC rule, shows.

Example 1 Let $A = \{a, b, c\}$ and $k = 2$. Consider the ABCC rule f with $f(1, 1) = 1$, $f(2, 2) = 2$, and $f(x, y) = 0$ otherwise. Despite its similarity with AV, the rule f is different since $f(1, 2) = 0$. Clearly, f is different than MC as well.

We will show that f is d-monotone robust against any majority-concentric distance metric d. To do so, by Lemma 3a, it suffices to show that $\mathbb{E}_{S\sim\mathcal{M}(U)}[\, \mathrm{sc}_f(U, S) - \mathrm{sc}_f(V, S)] > 0$ for every majority-concentric distance metric d, any d-monotonic noise model $\mathcal{M}$, and any pair of k-sized sets of alternatives U and V.

Without loss of generality, let us assume that $U = \{a, b\}$ and $V = \{a, c\}$. Observe that, by the definition of the ABCC rule f, the quantity $\mathrm{sc}_f(U, S) - \mathrm{sc}_f(V, S)$ is equal to 0 when $S = \emptyset, \{a\}, \{b, c\}, \{a, b, c\}$, 1 when $S = \{b\}$, -1 when $S = \{c\}$, 2 when $S = \{a, b\}$, and -2 when $S = \{a, c\}$. Let p_{ab}, p_{ac}, p_b, and p_c be abbreviations for the probabilities $\Pr_{\mathcal{M}}[S|U]$ for $S = \{a, b\}, \{a, c\}, \{b\}$, and $\{c\}$, respectively. Then, for the noise model $\mathcal{M}$, we have

$$\begin{aligned}\mathbb{E}_{S\sim\mathcal{M}(U)}[\, \mathrm{sc}_f(U, S) - \mathrm{sc}_f(V, S)] &= 2p_{ab} - 2p_{ac} + p_b - p_c\\ &= (p_{ab} - p_{ac}) + (p_{ab} - \max\{p_{ac}, p_c\}) + (p_b - \min\{p_{ac}, p_c\}).\end{aligned}$$

Clearly, $p_{ab} > \max\{p_{ac}, p_c\}$ implying that the two leftmost parentheses above are positive. We will show that majority-concentricity implies that $p_b \geq \min\{p_{ac}, p_b\}$ and the third parenthesis is non-negative; this will immediately yield $\mathbb{E}_{S\sim\mathcal{M}(U)}[\, \mathrm{sc}_f(U, S) - \mathrm{sc}_f(V, S)] > 0$ and d-monotone robustness of f will follow using Lemma 3a.

It remains to show that majority-concentricity implies $p_b \geq \min\{p_{ac}, p_b\}$. Assume otherwise that $p_b < \min\{p_{ac}, p_b\}$. Equivalently, we would have $d(U, \{b\}) > \max\{d(U, \{a, c\}), d(U, \{c\})\}$. Setting t such that $\delta_t(d, U) = \max\{d(U, \{a, c\}), d(U, \{c\})\}$, we get that $N^t_{b|c}(d, U)$ consists only of the set of alternatives $\{a, b\}$ and $N^t_{c|b}(d, U)$ contains the sets of alternatives $\{a, c\}$ and $\{c\}$. Hence, $|N^t_{b|c}(d, U)| < |N^t_{c|b}(d, U)|$, contradicting the majority-concentricity of d. □

5 Robustness of other ABCC rules

Before presenting the results in this section, let us give some context. In Sect. 3, we characterized the ABCC rules that are monotone robust against all distance metrics and concluded that MC is the only such rule. In Sect. 4, we characterized the class of majority-concentric distance metrics for which AV is monotone robust. However, this is not a characterization of the ABCC rules that are monotone robust against all majority-concentric distance metrics though; Example 1 indicates that there are other ABCC rules besides AV and MC that have this property.

In this section, we present characterization results for two large subclasses of majority concentric distance metrics, called *natural* and *similarity* distances. The characterization of the former excludes some well-known ABCC rules such as the CC rule, while the characterization of the latter involves all non-trivial ABCC rules. Interestingly, even the (narrowest among the two) class of similarity metrics contains the four prominent distance metrics presented in Sect. 2.

The definitions, statements, and proofs that we present in this section use appropriately defined bijections on sets of alternatives.

Definition 7 Given two different sets of alternatives U and V with $|U| = |V|$, a (U, V)-*bijection* $\mu : 2^A \rightarrow 2^A$ is defined as $\mu(S) = \{\mu'(a) : a \in S\}$, where $\mu' : A \rightarrow A$ is such that $\mu'(a) = a$ for every alternative $a \in U \cap V$ or $a \notin U \cup V$, $\mu'(a)$ is a distinct alternative in $V \setminus U$ for $a \in U \setminus V$, and $\mu'(a)$ is a distinct alternative in $U \setminus V$ for $a \in V \setminus U$.

It is easy to see that a (U, V)-bijection μ has the following properties.

Lemma 4 *Let $U, V \subseteq A$ with $|U| = |V|$ and let μ be a (U, V)-bijection. For every $S \subseteq A$, it holds $|S| = |\mu(S)|$, $|U \cap S| = |V \cap \mu(S)|$, and $|U \cap \mu(S)| = |V \cap S|$.*

Proof The proof follows easily by the definition of the (U, V)-bijection μ. The equality $|S| = |\mu(S)|$ holds because the function μ' maps each alternative of S to a distinct alternative. To prove the second equality, observe that the function μ' maps each alternative in $U \cap V$ to itself, each alternative in $U \setminus V$ to a distinct alternative in $V \setminus U$, and each alternative not belonging to U to an alternative not belonging to V. Hence, the alternatives in $U \cap S$ (and no other alternative in S) are mapped to distinct alternatives of V and, hence, $|U \cap S| = |V \cap \mu(S)|$. The proof of the third equality is symmetric. □

We are now ready to proceed with the presentation of our last set of results. In particular, our results for ABCC rules different than MC and AV involve two classes of distance metrics. We define the first one here.

Definition 8 *(natural distance)* A distance metric d is called *natural* if for every three sets of alternatives U, V, and S with $|U| = |V|$ such that $|U \cap S| > |V \cap S|$, it holds that $d(U, S) \leq d(V, S)$.

The next observation follows easily by the definitions.

Lemma 5 *Every natural distance metric is majority-concentric.*

Proof Let d be a natural distance, U a k-sized set of alternatives, and $a, b \in A$ with $a \in U$ and $b \notin U$. We will show that $|N^t_{a|b}(d, U)| \geq |N^t_{b|a}(d, U)|$ for $t = 0, 1, \ldots, \mathrm{span}(d, U)$. For $t = 0$, this is clearly true since $N^0_{a|b}(d, U) = \{U\}$ and $N^0_{b|a}(d, U) = \emptyset$.

Let $V = (U \setminus \{a\}) \cup \{b\}$ and μ be any (U, V)-bijection on sets of alternatives. For $t \geq 1$, let $S \in N^t_{b|a}(d, U)$. By the definition of the sets U and V and of the bijection μ, we have $\mu(S) = (S \setminus \{b\}) \cup \{a\}$. Then, $|\mu(S) \cap U| = |S \cap U| + 1 > |S \cap U|$. Due to the naturality of d, we get $d(\mu(S), U) \leq d(S, U)$; this follows by applying Definition 8 with sets S and $\mu(S)$ playing the role of U and V and set U playing the role of S there. By this inequality on distances and by the fact that $\mu(S)$ contains alternative a but not alternative b, we conclude that $\mu(S) \in N^t_{a|b}(d, U)$. Since μ is a bijection (the sets of $N^t_{b|a}(d, U)$ are mapped to distinct sets in $N^t_{a|b}(d, U)$), we get $|N^t_{a|b}(d, U)| \geq |N^t_{b|a}(d, U)|$, as desired. □

The opposite is not true as the next example illustrates.

Example 2 Let $A = \{a, b, c\}$ and consider the distance metric with $d(X, Y) = 0$ for every pair of sets with $X = Y$, $d(X, Y) = 1$ if $Y = A \setminus X$, and $d(X, Y) = 2$, otherwise. It can be easily seen that the distance is majority-concentric; it suffices to observe that, within distance 1 from any set, each alternative appears in exactly one set. To see that is not natural, consider $U = \{a, b\}$, $V = \{a, c\}$ and $S = \{b\}$. We have $|U \cap S| > |V \cap S|$ but $d(U, S) = 2 > 1 = d(V, S)$. □

The next lemma provides a sufficient condition so that an ABCC rule is monotone robust against a natural distance metric. It will play a crucial role later in proving our characterizations.

Lemma 6 *An ABCC rule f is d-monotone robust against a natural distance metric d if for every two different committees U and V there exists a* (U, V)*-bijection* μ *on sets of alternatives and a set* $S \subseteq A$ *with* $\mathrm{sc}_f(U, S) > \mathrm{sc}_f(V, S)$ *and* $d(U, S) < d(U, \mu(S))$.

Proof Let U and V be two different sets with k alternatives each. Let $\mathcal{S}_+$, $\mathcal{S}_-$, and $\mathcal{S}_0$ be the classes of sets of alternatives S with $|U \cap S| > |V \cap S|$, $|U \cap S| < |V \cap S|$, and $|U \cap S| = |V \cap S|$, respectively. Also, let $\mathcal{M}$ be a d-monotonic noise model for the natural distance metric d. Using this notation, we have

$$\begin{aligned}
\mathbb{E}_{S \sim \mathcal{M}(U)}[\mathrm{sc}_f(U, S) - \mathrm{sc}_f(V, S)] &= \sum_{S \subseteq A} \left(\mathrm{sc}_f(U, S) - \mathrm{sc}_f(V, S)\right) \cdot \Pr_{\mathcal{M}}[S|U] \\
&= \sum_{S \in \mathcal{S}_+} \left(\mathrm{sc}_f(U, S) - \mathrm{sc}_f(V, S)\right) \cdot \Pr_{\mathcal{M}}[S|U] \\
&\quad + \sum_{S \in \mathcal{S}_0} \left(\mathrm{sc}_f(U, S) - \mathrm{sc}_f(V, S)\right) \cdot \Pr_{\mathcal{M}}[S|U] \\
&\quad + \sum_{S \in \mathcal{S}_-} \left(\mathrm{sc}_f(U, S) - \mathrm{sc}_f(V, S)\right) \cdot \Pr_{\mathcal{M}}[S|U]
\end{aligned} \tag{5}$$

We will now transform the third sum in the RHS of (5) to one running over the sets of $\mathcal{S}_+$ like the first sum.

Let μ be a (U, V)-bijection on sets of alternatives; by Lemma 4, μ maps every set of $\mathcal{S}_-$ to a set of $\mathcal{S}_+$ and vice-versa. Hence, instead of enumerating sets of $\mathcal{S}_-$, we could enumerate sets of $\mathcal{S}_+$ and apply the bijection μ on them. The third sum in the RHS of (5) then becomes

$$\begin{aligned}
\sum_{S \in \mathcal{S}_-} \left(\mathrm{sc}_f(U, S) - \mathrm{sc}_f(V, S)\right) \cdot \Pr_{\mathcal{M}}[S|U] &= \sum_{S \in \mathcal{S}_+} \left(\mathrm{sc}_f(U, \mu(S)) - \mathrm{sc}_f(V, \mu(S))\right) \cdot \Pr_{\mathcal{M}}[\mu(S)|U] \\
&= \sum_{S \in \mathcal{S}_+} \left(\mathrm{sc}_f(V, S) - \mathrm{sc}_f(U, S)\right) \cdot \Pr_{\mathcal{M}}[\mu(S)|U]
\end{aligned} \tag{6}$$

The second equality follows since, by Lemma 4, $\mathrm{sc}_f(U, \mu(S)) = f(|U \cap \mu(S)|, |\mu(S)|) = \mathrm{sc}_f(|V \cap S|, |S|) = \mathrm{sc}_f(V, S)$ and, similarly, $\mathrm{sc}_f(V, \mu(S)) = \mathrm{sc}_f(U, S)$.

Now observe that $\mathrm{sc}_f(U, S) = \mathrm{sc}_f(V, S)$ when $S \in \mathcal{S}_0$. Hence, the second sum in the RHS of (5) is equal to 0. By combining (5) and (6), we get

$$\mathbb{E}_{S \sim \mathcal{M}(U)}[\mathrm{sc}_f(U, S) - \mathrm{sc}_f(V, S)] = \sum_{S \in \mathcal{S}_+} \left(\mathrm{sc}_f(U, S) - \mathrm{sc}_f(V, S)\right) \cdot (\Pr_{\mathcal{M}}[S|U] - \Pr_{\mathcal{M}}[\mu(S)|U]). \tag{7}$$

Finally, observe that the RHS of (7) is always non-negative. This is due to the fact that $S \in \mathcal{S}_+$ which implies that $\mathrm{sc}_f(U,S) = f(|U \cap S|, |S|) \geq f(|V \cap S|, |S|) = \mathrm{sc}_f(V,S)$ since f is non-decreasing in its first argument and $d(U,S) \leq d(U,\mu(S))$ (and, consequently, $\Pr_{\mathcal{M}}[S|U] \geq \Pr_{\mathcal{M}}[\mu(S)|U]$) since d is natural and, by Lemma 4, $|U \cap S| > |V \cap S| = |U \cap \mu(S)|$. The RHS of (7) is strictly positive if there exists a set $S \in \mathcal{S}_+$ that further satisfies $d(U,S) < d(U,\mu(S))$ (and, consequently, $\Pr_{\mathcal{M}}[S|U] > \Pr_{\mathcal{M}}[\mu(S)|U]$) and $\mathrm{sc}_f(U,S) > \mathrm{sc}_f(V,S)$. The lemma then follows by Lemma 3a. □

We now present two applications of Lemma 6.

Theorem 3 *An ABCC rule f is monotone robust against any natural distance metric if and only if $f(k,k) > f(k-1,k)$.*

Proof For the proof of the "if" part, we use Lemma 6. Let d be a natural distance metric. Consider any pair of different k-sized sets of alternatives U and V and any (U, V)-bijection μ. For $S = U$, we have $\mu(S) = V$ and, hence, $d(U,U) < d(U,V) = d(U,\mu(S))$. Furthermore, by the definition of f, we have $\mathrm{sc}_f(U,U) = f(k,k) > f(k-1,k) \geq f(|U \cap V|, |V|) = \mathrm{sc}_f(U,V)$. Hence, the conditions of Lemma 6 are satisfied, and f is d-monotone robust.

For the proof of the "only if" part, assume that $f(k,k) = f(k-1,k)$ and consider the natural distance metric d with $d(X,Y) = 0$ if $X = Y$ and $d(X,Y) = 1$ otherwise. Let $U, V \subseteq A$ be different sets with k alternatives each such that $|U \cap V| = k-1$. Let $U \backslash V = \{a\}$ and $V \backslash U = \{b\}$.

Unfortunately, as we will see, we are in the case $\mathbb{E}_{S \sim \mathcal{M}(U)}[\mathrm{sc}_f(U,S) - \mathrm{sc}_f(V,S)] = 0$ and, hence, we cannot use Lemma 3 to complete the proof. We will instead prove that $\mathrm{sc}_f(U,S) - \mathrm{sc}_f(V,S)$ follows a symmetric distribution (i.e., $\Pr_{S \sim \mathcal{M}(U)}[\mathrm{sc}_f(U,S) - \mathrm{sc}_f(V,S) = t] = \Pr_{S \sim \mathcal{M}(U)}[\mathrm{sc}_f(U,S) - \mathrm{sc}_f(V,S) = -t]$ for every $t > 0$). Then, the random variable $\mathrm{sc}_f(U,\Pi) - \mathrm{sc}_f(V,\Pi)$, where Π is a random profile of approval votes drawn from the noisy model $\mathcal{M}$ with ground truth U will be symmetric with expectation zero as well. Hence, the probability that $\mathrm{sc}_f(U,\Pi) - \mathrm{sc}_f(V,\Pi)$ is strictly positive (which is a necessary condition so that U is the unique winner) is at most 1/2, and, hence, f is not accurate in the limit for $\mathcal{M}$.

First observe that $\mathrm{sc}_f(U,S) - \mathrm{sc}_f(V,S) = 0$ when S is equal to U or V, or contains both alternatives a and b, or contains neither a nor b. Indeed, we have $\mathrm{sc}_f(U,U) - \mathrm{sc}_f(V,U) = f(k,k) - f(k-1,k) = 0$ and $\mathrm{sc}_f(U,V) - \mathrm{sc}_f(V,V) = f(k-1,k) - f(k,k) = 0$, by our assumption. Furthermore, if S contains both a and b or none of them, we have that $|U \cap S| = |V \cap S|$ and, hence, $\mathrm{sc}_f(U,S) = \mathrm{sc}_f(V,S)$.

Denote by $\mathcal{S}$ the collection of the remaining sets of alternatives, i.e.,

$$\mathcal{S} = \{S : S \neq U, V \text{ and } |S \cap \{a,b\}| = 1\},$$

and partition $\mathcal{S}$ into the subcollections $\mathcal{S}_{a|b}$ and $\mathcal{S}_{b|a}$ consisting of sets that include alternative a (but not alternative b) and alternative b (but not alternative a, respectively). Now, consider the (unique) (U, V)-bijection μ and observe that for every set S in $\mathcal{S}_{a|b}$, $\mu(S)$ is a distinct set of $\mathcal{S}_{b|a}$ and vice-versa. Furthermore, notice that, by Lemma 4, we have

$$
\begin{aligned}
\mathrm{sc}_f(U,S) - \mathrm{sc}_f(V,S) &= f(|U\cap S|,|S|) - f(|V\cap S|,|S|)\\
&= f(|V\cap \mu(S)|,|\mu(S)|) - f(|U\cap \mu(S)|,|\mu(S)|)\\
&= -(\mathrm{sc}_f(U,\mu(S)) - \mathrm{sc}_f(V,\mu(S))).
\end{aligned}
$$

The proof completes by observing that, by the definition of the distance metric d, the noise model $\mathcal{M}$ with ground truth U returns each set of $\mathcal{S}$ (and, consequently, S and $\mu(S)$) equiprobably. ◻

Notice that most popular ABCC rules from Sect. 2 satisfy the condition of Theorem 3. CC is an exception. The proof of Theorem 3 implies that CC is not monotone robust for the natural distance metric d defined as $d(X,Y)=0$ if $X=Y$ and $d(X,Y)=1$, otherwise.

Our second application of Lemma 6 involves all non-trivial ABCC rules and an important subclass of natural distances.

Definition 9 *(similarity distance)* A natural distance metric d is a *similarity* distance metric if for every three sets of alternatives U, V, and S with $|U|=|V|$ such that $|U\cap S|>|V\cap S|$, it holds that $d(U,S)<d(V,S)$.

We remark that the only difference between the definitions of natural and similarity distance metrics is in the inequality on distances which becomes strict for similarity distance metrics. Interestingly, this minor difference leads to a notable difference in our characterizations in Theorems 3 and 4 .

Theorem 4 *Any non-trivial ABCC rule is monotone robust against any similarity distance metric.*

Proof We apply Lemma 6 assuming a non-trivial ABCC rule f and a similarity distance metric d. Non-triviality of f implies that for every two different sets U and V with k alternatives each, there is a set S such that $\mathrm{sc}_f(U,S)>\mathrm{sc}_f(V,S)$. This yields $|U\cap S|>|V\cap S|=|U\cap \mu(S)|$, where μ is any (U, V)-bijection (see Lemma 4), and implies that $d(U,S)<d(U,\mu(S))$ since d is a similarity distance. ◻

We can easily show that the four distance metrics that we defined in Sect. 2 (namely, the set difference, Jaccard, Zelinka, and Bunke-Shearer distances) are all similarity distance metrics. Using this observation and Theorem 4, we obtain the next statement.

Corollary 1 *Any non-trivial ABCC rule is monotone robust against the set difference, Jaccard, Zelinka, and Bunke-Shearer distance metrics.*

6 Discussion and open problems

We believe that our approach complements nicely the axiomatic and quantitative analysis of approval-based multiwinner voting. Our results define a hierarchy of the ABCC rules in terms of robustness, depending on the breadth of the class of distance metrics against which a rule is monotone robust. MC is at the top of this hierarchy and the trivial ABCC rule is at the bottom, with the rest of the ABCC rules lying in between. An important problem that we leave open is whether there are ABCC rules that lie between rules MC and AV in terms of robustness.

This question is related to a subtle issue that involves alternative-independent distance metrics. In Sect. 3, we showed that the ABCC rule MC is the only one that is monotone robust against any distance metric. In the second part of the proof of Theorem 1, we proved that for every ABCC rule f different than MC, there exists an alternative-dependent distance metric d and a d-monotonic noise model, against which f is not accurate in the limit. Would that part of the proof work using an alternative-independent distance metric instead? Or is alternative-dependence really necessary?

In the following we show that this is indeed the case by presenting an ABCC rule, which is different than MC, and is monotone robust against any alternative-independent distance metric. Consider the case with $m = 4$, $A = \{a, b, c, d\}$, and $k = 2$, and the ABCC rule f_{AV2} defined as follows: for $(x, y) \in \mathcal{X}_{4,2}$, it is $f_{\mathrm{AV2}}(x, y) = x$ if $y = 2$ and $f_{\mathrm{AV2}}(x, y) = 0$ otherwise.

Theorem 5 *The ABCC rule f_{AV2} is monotone robust against any alternative-independent distance metric.*

Proof Let $A = \{a, b, c, d\}$. We consider an alternative-independent distance metric d and the ground truth $U = \{a, b\}$. For a set of alternatives $S \subseteq A$, the distance $d(U, S)$ can be thought of as a function that depends only on $|U \cap S|$ and $|S|$. The d-monotonic noise model $\mathcal{M}$ produces the approval set S with probability $\Pr_{\mathcal{M}}[S|U] = p(|U \cap S|, |S|)$ when the ground truth U is used. The different probability values that the noise model uses correspond to those pairs (x, y) that belong to the set $\mathcal{X}_{4,2} = \{(0,0), (0,1), (1,1), (0,2), (1,2), (2,2), (1,3), (2,3), (2,4)\}$.

We will show that f_{AV2} is accurate in the limit for $\mathcal{M}$. By Lemma 3 and due to the symmetry implied by alternative-independence, we need to show that $\mathbb{E}_{S\sim\mathcal{M}(U)}[\,\mathrm{sc}_{\mathrm{AV2}}(U, S) - \mathrm{sc}_{\mathrm{AV2}}(V, S)] > 0$ for $V = \{a, c\}$ and $V = \{c, d\}$.

For $V = \{a, c\}$, we have

$$\begin{aligned}
&\mathbb{E}_{S\sim\mathcal{M}(U)}[\,\mathrm{sc}_{\mathrm{AV2}}(U, S) - \mathrm{sc}_{\mathrm{AV2}}(V, S)\\
&\quad = \sum_{S\subset A} f_{\mathrm{AV2}}(|\{a, b\} \cap S|, |S|) - f_{\mathrm{AV2}}(|\{a, c\} \cap S|, |S|) \cdot p(|\{a, b\} \cap S|, |S|)\\
&\quad = p(2, 2) - p(1, 2) > 0.
\end{aligned}$$

In the second equality, we have used the fact that the contribution of all sets S of size different than 2 as well as of the sets $\{a, d\}, \{b, c\}$ to the sum is zero. In addition, the contribution of the sets $\{a, b\}, \{a, c\}, \{b, d\}$, and $\{c, d\}$ is $p(2, 2)$, $-p(1, 2)$, $p(1, 2)$, and $-p(1, 2)$. The inequality follows since $p(2, 2)$ is the probability with which $\mathcal{M}$ returns the ground truth.

For $V = \{c, d\}$, we have

$$\begin{aligned}
&\mathbb{E}_{S\sim\mathcal{M}(U)}[\,\mathrm{sc}_{\mathrm{AV2}}(U, S) - \mathrm{sc}_{\mathrm{AV2}}(V, S)]\\
&\quad = \sum_{S\subset A} f_{\mathrm{AV2}}(|\{a, b\} \cap S|, |S|) - f_{\mathrm{AV2}}(|\{c, d\} \cap S|, |S|) \cdot p(|\{a, b\} \cap S|, |S|)\\
&\quad = 2 \cdot p(2, 2) - 2 \cdot p(0, 2) > 0.
\end{aligned}$$

The contribution to the sum of all sets besides $\{a, b\}$ and $\{c, d\}$ is zero. The contribution of the sets $\{a, b\}$ and $\{c, d\}$ is $2 \cdot p(2, 2)$ and $-2 \cdot p(0, 2)$ respectively. Again, the inequality follows since $p(2, 2)$ is the probability with which $\mathcal{M}$ returns the ground truth. □

Clearly, if we restrict our attention to alternative-independent distance metrics, the rule f_{AV2} is, in addition to MC, another ABCC rule that is strictly more robust than AV. We can furthermore show that there are rules with intermediate robustness. For example, in the same setting with $m = 4$ and $k = 2$, one such rule $f_{\mathrm{AV+}}$ is defined as $f_{\mathrm{AV+}}(x, y) = f_{\mathrm{AV}}(x, y)$ if $y \neq 2$ and $f_{\mathrm{AV+}}(x, 2) = 2f_{\mathrm{AV}}(x, 2)$. It can be shown (the proof is omitted) to be strictly more robust than AV and strictly less robust than MC (and f_{AV2}). It would be interesting to obtain general results (for general values of the parameters m and k) and characterize all ABCC rules that lie between MC and AV in terms of robustness in alternative-independent distance metrics only. Furthermore, applying our framework to non-ABCC rules deserves investigation.

Beyond assessing the effects of noise in the limit, studying the sample complexity of approval-based multiwinner voting is important. This will require the design of concrete noise models like the $\mathcal{M}_p$ model that we presented in Sect. 2. To make this problem concrete, consider the following question regarding AV and the noise model $\mathcal{M}_p$. Given $\epsilon > 0$, determine an integer n_ϵ so that, when applied on a profile with at least n_ϵ votes generated by the noise model $\mathcal{M}_p$ with ground truth committee U, AV returns U as the unique winner with probability at least $1 - \epsilon$. By slightly modifying our argument in the proof of Lemma 3, we can show that the sample complexity bound n_ϵ is at most $O\left(k \ln m + \ln \frac{1}{\epsilon}\right)$. Is this bound tight? What about the sample complexity of other ABCC rules for this noise model? And, more importantly, what about other noise models?

In particular, models that simulate user behaviour in crowdsourcing platforms will be useful for evaluating approval-based voting in such environments. Even though the $\mathcal{M}_p$ model is very simple, we expect that implementation issues will emerge for more elaborate noise models. For example, consider a noise model $\mathcal{M}_p^Z$ that uses a parameter $p \in (1/2, 1]$ like the noise model $\mathcal{M}_p$, but generates the approval vote S with probability proportional to $\left(\frac{1-p}{p}\right)^{d_Z(U,S)}$ instead of $\left(\frac{1-p}{p}\right)^{d_\Delta(U,S)}$ when it uses the ground truth committee U. Can this random selection be implemented in polynomial time? Similar issues in the implementation of the Mallows ranking model [23] have triggered much non-trivial work; see, e.g., Doignon et al. [13].

Appendix

Theorem 6 *The ABCC rule AV is a maximum likelihood estimator for the noise model $\mathcal{M}_p$.*

Proof Let $\Pi = (S_i)_{i \in [n]}$ be a profile with n approval votes. We need to show that the profile Π has maximum probability to have been produced by the noise model $\mathcal{M}_p$ with a set of k alternatives of maximum AV score from the votes of Π as the ground truth committee.

Indeed, the probability that Π has been produced by the noise model $\mathcal{M}_p$ with ground truth committee U is

$$\prod_{i \in [n]} \Pr_{\mathcal{M}_p}[S_i | U] = \prod_{i \in [n]} p^m \cdot \left(\frac{1-p}{p}\right)^{d_\Delta(S_i,U)} = p^{mn} \cdot \left(\frac{1-p}{p}\right)^{\sum_{i \in [n]} d_\Delta(S_i,U)}.$$

Since $p > 1/2$, the above expression is maximized by minimizing the quantity $\sum_{i \in [n]} d_\Delta(S_i, U)$. Now, we can express this quantity in terms of the number of agents n, the committtee size k, the total size of approval votes in profile Π, and the AV score of committee U from the votes of Π, using the following derivation:

$$\begin{aligned}\sum_{i\in[n]} d_\Delta(U,S_i) &= \sum_{i\in[n]} \left(|U\backslash S_i| + |S_i\backslash U|\right) = \sum_{i\in[n]} \left(|U| + |S_i| - |U\cap S_i|\right)\\ &= nk + \sum_{i\in[n]} |S_i| - \sum_{i\in[n]} \text{sc}_{\text{AV}}(U,S_i) = nk + \sum_{i\in[n]} |S_i| - \text{sc}_{\text{AV}}(U,\Pi).\end{aligned}$$

Hence, the probability that the profile Π is generated by a noise model $\mathcal{M}_p$ is maximized for the ground truth committee U of maximum score $\text{sc}_{\text{AV}}(U,\Pi)$. □

Acknowledgements This research is co-financed by Greece and the European Union (European Social Fund) through the Operational Programme "Human Resources Development, Education and Lifelong Learning 2014-2020" (project MIS 5047146).

References

1. Arrow, K. J. (1951). *Social choice and individual values*. Cowles Foundation.
2. Aziz, H., Gaspers, S., Gudmundsson, J., Mackenzie, S., Mattei, N., & Walsh, T. (2015). Computational aspects of multi-winner approval voting. In *Proceedings of the 14th international conference on autonomous agents and multiagent systems (AAMAS)* (pp. 107–115).
3. Aziz, H., Brill, M., Conitzer, V., Elkind, E., Freeman, R., & Walsh, T. (2017). Justified representation in approval-based committee voting. *Social Choice and Welfare, 48*(2), 461–485.
4. Brams, S. J., & Kilgour, D. M. (2014). Satisfaction approval voting. In R. Fara, D. Leech, & M. Salles (Eds.), *Voting power and procedures: Essays in honour of Dan Felsenthal and Moshé Machover* (pp. 323–346). Springer.
5. Bredereck, R., Faliszewski, P., Kaczmarczyk, A., Niedermeier, R., Skowron, P., & Talmon, N. (2017). Robustness among multiwinner voting rules. In *Proceedings of the 10th international symposium on algorithmic game theory (SAGT)* (pp. 80–92).
6. Brill, M., Laslier, J., & Skowron, P. (2017). Multiwinner approval rules as apportionment methods. In *Proceedings of the 31st AAAI conference on artificial intelligence (AAAI)* (pp. 414–420).
7. Caragiannis, I., & Micha, E. (2017). Learning a ground truth ranking using noisy approval votes. In *Proceedings of the 26th international joint conference on artificial intelligence (IJCAI)* (pp. 149–155).
8. Caragiannis, I., Procaccia, A. D., & Shah, N. (2014). Modal ranking: A uniquely robust voting rule. In *Proceedings of the 28th AAAI conference on artificial intelligence (AAAI)* (pp. 616–622).
9. Caragiannis, I., Procaccia, A. D., & Shah, N. (2016). When do noisy votes reveal the truth? *ACM Transactions on Economics and Computation, 4*(3), 15:1-15:30.
10. Chamberlin, J. R., & Courant, P. N. (1983). Representative deliberations and representative decisions: Proportional representation and the Borda rule. *American Political Science Review, 77*(3), 718–733.
11. Conitzer, V., & Sandholm, T. (2005). Common voting rules as maximum likelihood estimators. In *Proceedings of the 21st conference on uncertainty in artificial intelligence (UAI)* (pp. 145–152).
12. Deza, M. M., & Deza, E. (2016). *Encyclopedia of distances* (4th ed.). Springer.
13. Doignon, J. P., Pakeč, A., & Regenwetter, M. (2004). The repeated insertion model for rankings: Missing link between two subset choice models. *Psychometrika, 69*(1), 33–54.
14. Faliszewski, P., Skowron, P., Slinko, A. M., & Talmon, N. (2017). Multiwinner voting: A new challenge for social choice theory. In U. Endriss (Ed.), *Trends in computational social choice, AI access* (pp. 27–47).
15. Faliszewski, P., Skowron, P., Slinko, A., & Talmon, N. (2018). Multiwinner analogues of the plurality rule: Axiomatic and algorithmic perspectives. *Social Choice and Welfare, 51*(3), 513–550.
16. Gawron, G., & Faliszewski, P. (2019). Robustness of approval-based multiwinner voting rules. In *Proceedings of the 6th international conference on algorithmic decision theory (ADT)* (pp. 17–31).
17. Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association, 58*(301), 13–30.
18. Kilgour, D. M. (2010). Approval balloting for multi-winner elections. In J. F. Laslier & M. R. Sanver (Eds.), *Handbook on approval voting* (pp. 105–124). Springer.

19. Lackner, M., & Skowron, P. (2019). A quantitative analysis of multi-winner rules. In *Proceedings of the 28th international joint conference on artificial intelligence (IJCAI)* (pp. 407–413).
20. Lackner, M., & Skowron, P. (2021). Consistent approval-based multi-winner rules. *Journal of Economic Theory, 192,* 105173.
21. Laslier, J. F., & Sanver, M. R. (Eds.). (2010). *Handbook on approval voting*. Springer.
22. Leskovec, J., Rajaraman, A., & Ullman, J. D. (2014). *Mining of massive datasets* (2nd ed.). Cambridge University Press.
23. Mallows, C. L. (1957). Non-null ranking models. *Biometrika, 44*, 114–130.
24. Misra, N., & Sonar, C. (2019). Robustness radius for chamberlin-courant on restricted domains. In *Proceedings of the 45th international conference on current trends in theory and practice of computer science (SOFSEM)* (vol. 11376, pp. 341–353). Springer.
25. Procaccia, A. D., & Shah, N. (2015). Is approval voting optimal given approval votes? In *Proceedings of the 29th annual conference on neural information processing systems (NIPS)* (pp. 1801–1809).
26. Procaccia, A. D., Reddi, S. J., & Shah, N. (2012). A maximum likelihood approach for selecting sets of alternatives. In *Proceedings of the 28th conference on uncertainty in artificial intelligence (UAI)* (pp. 695–704).
27. Sánchez-Fernández, L., Elkind, E., Lackner, M., García, N. F., Arias-Fisteus, J., Basanta-Val, P., & Skowron, P. (2017). Proportional justified representation. In *Proceedings of the 31st AAAI conference on artificial intelligence (AAAI)* (pp. 670–676).
28. Shiryaev, D., Yu, L., & Elkind, E. (2013). On elections with robust winners. In *Proceedings of the 12th international conference on autonomous agents and multi-agent systems (AAMAS)* (pp. 415—422).
29. Skowron, P., Faliszewski, P., & Lang, J. (2016). Finding a collective set of items: From proportional multirepresentation to group recommendation. *Artificial Intelligence, 241*, 191–216.
30. Thiele, T. N. (1895). On multiple choice (in Danish). *Bulletin of the Royal Danish Academy of Sciences and Letters,* 415–441.
31. Young, H. P. (1988). Condorcet's theory of voting. *American Political Science Review, 82*(4), 1231–1244.