CrossMark

ORIGINAL PAPER

Multiwinner analogues of the plurality rule: axiomatic and algorithmic perspectives

Piotr Faliszewski[1] · **Piotr Skowron[2]** · **Arkadii Slinko[3]** · **Nimrod Talmon[4]**

Abstract We characterize the class of committee scoring rules that satisfy the fixed-majority criterion. We argue that rules in this class are multiwinner analogues of the single-winner Plurality rule, which is uniquely characterized as the only single-winner scoring rule that satisfies the simple majority criterion. We define top-k-counting committee scoring rules and show that the fixed-majority consistent rules are a subclass of the top-k-counting rules. We give necessary and sufficient conditions for a top-k-counting rule to satisfy the fixed-majority criterion. We show that, for many top-k-counting rules, the complexity of winner determination is high (formally, we show that the problem of deciding if there exists a committee with at least a given score is NP-hard), but we also show examples of rules with polynomial-time winner determination procedures. For some of the computationally hard rules, we provide either exact FPT algorithms or approximate polynomial-time algorithms.

✉ Piotr Skowron
  p.k.skowron@gmail.com

[1] AGH University, Kraków, Poland

[2] University of Warsaw, Warsaw, Poland

[3] University of Auckland, Auckland, New Zealand

[4] Ben-Gurion University of the Negev, Be'er Sheva, Israel

Springer

1 Introduction

A multiwinner voting rule is a formal procedure for selecting a subset of predetermined size from the available candidates in accord with the preferences of an electorate (such a subset of candidates is usually referred to as a *committee*). Parliamentary elections constitute one of the most classic examples where multiwinner rules are regularly used. For country-wide elections, societies typically use the district-based First-Past-the-Post rule, or a party-list system, or some mixture of the two [nonetheless, some countries use other rules for this purpose, such as SNTV or STV (Lijphart and Aitkin 1994)]. In smaller-scale elections, where it is possible for the voters to rank all the candidates, many other rules become available; for example, the k-Borda rule (Debord 1992) selects committees where each member receives broad support from the electorate, the Chamberlin–Courant rule (Chamberlin and Courant 1983) finds committees with diverse membership, and the Monroe rule (Monroe 1995) is designed to achieve proportional representation.

Apart from political elections, multiwinner rules are useful for many other purposes: to shortlist candidates for a job interview (Barberà and Coelho 2008; Elkind et al. 2017), to determine the locations of public facilities (Zanjirani and Hekmatfar 2009), in a wide range of scenarios where resources need to be selected and assigned to the agents for their (shared) use (Skowron et al. 2016), in segmentation problems (Kleinberg et al. 2004), or even in search strategies of genetic algorithms (Faliszewski et al. 2017; Sawicki et al. 2017). In business applications, company strategists deciding which sets of products to advertise on the front pages of their websites implicitly use multiwinner elections to make their choices (Lu and Boutilier 2011, 2015). Since these tasks are very different in spirit, one may presume that not all rules are equally suitable for all scenarios. This makes the question of comparing different rules, and of understanding their nature and their shortcomings (including their computational difficulty), very relevant.

One approach to advance our understanding of the nature of multiwinner rules is to view them as extensions of certain well-understood single-winner ones. For example, Single Non-Transferable Vote (SNTV) can clearly be viewed as an extension of Plurality, because it selects the k candidates with the k highest numbers of the first-place votes. However, this is not the only point of view that one can take in generalizing Plurality. For instance, one could argue that a voter's most preferred committee consists exactly of those k candidates that this voter ranks in top k positions, and, so, if under the Plurality rule a voter gives a point only to his or her most preferred candidate, then under a multiwinner Plurality a voter should give points only to those candidates that belong to his or her most preferred committee. In fact, this is exactly how the Bloc rule works, and one can argue that Bloc is an extension of Plurality to the multiwinner setting as well. Naturally, there are also many other rules that would qualify for this title. Our goal in this paper is to seek and study such rules.

Our goal requires some justification. It is widely acknowledged that the single-winner Plurality rule has only one advantage: simplicity. Apart from this, it is considered a very bad rule—for instance, during the "Voting Power in Practice" workshop, held in 2010 at the Chateau du Baffy, Normandy, the participants (who were experts in voting procedures) were asked to rank election rules. Laslier (2012) reports

that Plurality was considered the worst. One of the most serious drawbacks of Plurality is that voters are pressured to vote for one of the two candidates they predict are most likely to win, even if their true most preferred candidate is neither of them; they do that from the fear of casting a 'wasted vote' (Dummett 1984). However, in the multiwinner setting this pressure becomes milder, because there are more candidates to be elected. We view this as one reason why multiwinner analogues of Plurality are worth investigating.

We seek multiwinner analogues of Plurality within the family of committee scoring rules, recently introduced by Elkind et al. (2017). This is a natural choice because Plurality belongs to the class of positional scoring rules and committee scoring rules generalize this class to the multiwinner setting. (However, looking for such rules beyond the class of committee scoring rules would not be unthinkable.) Further, we take the following axiomatic approach. We note that Plurality is the only single-winner scoring rule that satisfies the *simple majority* criterion,[1] which stipulates that a candidate ranked first by more than half of the voters must be the unique winner of the election. The *fixed-majority criterion*, introduced by Debord (1993), extends this notion to the world of multiwinner elections by requiring that, if there is a simple majority of voters, each of whom ranks the same k candidates in the top k positions (perhaps in a different order), then these k candidates should form the unique winning committee.[2] Thus, all in all, we seek committee scoring rules that satisfy the fixed-majority criterion.

One can verify that SNTV fails the fixed-majority criterion for all $k > 1$, but that Bloc does satisfy it. Yet, Bloc is not the only fixed-majority consistent rule within the class of committee scoring rules. In fact, our approach led us to the discovery of a new class of voting rules, which includes all committee scoring rules satisfying the fixed-majority criterion. We call them *top-k-counting* rules. As in the case of Bloc, they take only the top k preferences of the voters into consideration. Specifically, under a top-k-counting rule, each voter awards points to every committee on the basis of the number of this voter's top k candidates that are members of the committee; the committee with the most points collected from all the voters wins. The function that determines the score of a committee based on the number of committee members ranked in the top k positions by a voter will be called the *counting function*. As it turns out, the nature of this function (e.g., whether it is convex or concave) has very strong impact on both axiomatic and computational properties of the voting rule it defines.

We provide an (almost) full characterization[3] of fixed-majority consistent committee scoring rules and we analyze the computational complexity of their winner determination problems. More specifically, we obtain the following results:

1. We prove that all committee scoring rules that satisfy the fixed-majority criterion are top-k-counting rules and we establish a condition on the counting function that is necessary and sufficient for the corresponding top-k-counting rule to satisfy the

---

[1] In the literature, simple majority is often referred to as majority. However, we write 'simple majority' to clearly distinguish it from qualified majority and from fixed-majority.

[2] In Sect. 3.1 we argue that, indeed, this is a natural extension of the simple majority property.

[3] We consider the case where there are at least twice as many candidates as the size of the committee. We are not sure whether this restriction can be dropped.

fixed-majority criterion. This condition is a fairly mild relaxation of the classic notion of convexity; in particular, if the counting function is convex then the corresponding top-k-counting rule satisfies the fixed-majority criterion.

2. We show that a number of top-k-counting rules are NP-hard to compute[4] (for example, we show an example of a rule that closely resembles the Bloc rule and is hard even to approximate). There are, however, some polynomial-time computable ones (for example, the Bloc and the Perfectionist rules; the latter one is introduced in this paper).

3. We show that if the counting function is concave, then the corresponding top-k-counting rule fails the fixed-majority criterion, but the rule seems to be computationally easier than in the convex case. Specifically, for top-k-counting rules defined via concave counting functions we present a polynomial-time $(1 - \frac{1}{e})$-approximation algorithm and an exact fixed-parameter tractable algorithm (parameterized by the number of voters) for the problem of computing the highest-scoring committees.

All in all, there is no unique multiwinner analogue of Plurality, even if we restrict ourselves to polynomial-time computable committee scoring rules, but there is a rich class of such rules that deserves further investigation.

2 Preliminaries

An election is a pair $E = (C, V)$, where $C = \{c_1, \ldots, c_m\}$ is a set of candidates and $V = (v_1, \ldots, v_n)$ is a collection of voters. Throughout the paper, we reserve the symbol m to denote the number of candidates. Each voter v_i is associated with a preference order $\succ_i$ in which v_i ranks the candidates from his or her most desirable one to his or her least desirable one (we assume the unrestricted domain, i.e., each voter is free to choose any preference order). If X and Y are two (disjoint) subsets of C, then by $X \succ_i Y$ we mean that for each $x \in X$ and each $y \in Y$ it holds that $x \succ_i y$. For a positive integer t, we denote the set $\{1, \ldots, t\}$ by $[t]$.

Single-Winner Voting Rules A single-winner voting rule $\mathcal{R}$ is a function that given an election $E = (C, V)$, outputs a subset $\mathcal{R}(E) \subseteq C$ of candidates that are called (tied) winners of this election. There is quite a variety of single-winner voting rules, but for this paper it suffices to consider scoring rules only. Given a voter v and a candidate c, we write $\mathrm{pos}_v(c)$ to denote the position of c in v's preference order (for example, if v ranks c first then $\mathrm{pos}_v(c) = 1$). A scoring function for m candidates is a function $\gamma_m \colon [m] \to \mathbb{R}_+$ such that for each $i \in [m-1]$ we have $\gamma_m(i) \geq \gamma_m(i+1)$ (by $\mathbb{R}_+$ we mean the set of nonnegative real numbers). Each family of scoring functions $\gamma = (\gamma_m)_{m \in \mathbb{N}}$ (one function for each possible choice of m) defines a voting rule $\mathcal{R}_\gamma$ as follows. Let $E = (C, V)$ be an election with m candidates. Under $\mathcal{R}_\gamma$, each candidate $c \in C$ receives $\mathrm{score}(c) := \sum_{v \in V} \gamma_m(\mathrm{pos}_v(c))$ points and the candidate with the highest number of points wins. (If there are several such candidates, then they all tie as winners; the term *single-winner voting rule* refers to the fact that we use the rule to

---

[4] See Remark [4] in Sect. [4] for an exact explanation of this statement.

fill-in a single position, and not to indicate that the rule is resolute.) We often refer to the value score(c) as the γ-score of c.

The following scoring functions and scoring rules are particularly interesting. The t-Approval scoring function α_t is defined as $\alpha_t(i) := 1$ for $i \leq t$ and $\alpha_t(i) := 0$ otherwise. (If t is fixed, then the definition of α_t does not depend on m; in such cases, α_t can both be viewed as a scoring function and as a family of scoring functions.) For example, Plurality is $\mathcal{R}_{\alpha_1}$, the t-Approval rule is $\mathcal{R}_{\alpha_t}$, and the Veto rule is $\mathcal{R}_{(\alpha_{m-1})_{m \in \mathbb{N}}}$. The Borda scoring function (for m candidates), β_m, is defined as $\beta_m(i) := m - i$, and $\mathcal{R}_\beta$ is the Borda rule, where $\beta = (\beta_m)_{m \in \mathbb{N}}$. This notation for these scoring functions will be used throughout the paper.

Multiwinner Voting Rules A multiwinner voting rule $\mathcal{R}$ is a function that given an election $E = (C, V)$ and a number k representing the size of the desired committee, outputs a family $\mathcal{R}(E, k)$ of size-k subsets of C; the sets in this family are the committees that tie as winners. As in the case of single-winner voting rules, one may need a tie-breaking rule to get a unique winning committee, but we ignore this aspect in the current paper.

We focus on the class of committee scoring rules, introduced by Elkind et al. (2017) (we remark that the conference version of their paper was published in 2014). Consider an election $E = (C, V)$ and some committee S of a given size k. Let v be some voter in V. By pos$_v(S)$ we mean the sequence $(i_1, \ldots, i_k)$ that results from sorting the set $\{\text{pos}_v(c) : c \in S\}$ in a strictly increasing order. For example, if $C = \{a, b, c, d, e\}$, the preference order of v is $a \succ b \succ c \succ d \succ e$, and $S = \{a, c, d\}$, then pos$_v(S) = (1, 3, 4)$. If $I = (i_1, \ldots, i_k)$ and $J = (j_1, \ldots, j_k)$ are two strictly increasing sequences of integers, then we say that I *(weakly) dominates* J (denoted $I \succeq J$) if $i_t \leq j_t$ for each $t \in [k]$. For positive integers m and k, $k \leq m$, by $[m]_k$ we mean the set of all strictly increasing size-k sequences of integers from $[m]$.

Definition 1 (Elkind et al. 2017) A committee scoring function for a multiwinner election with m candidates, where we seek a committee of size k, is a function $f_{m,k} : [m]_k \to \mathbb{R}_+$ such that for each two sequences $I, J \in [m]_k$ it holds that if $I \succeq J$ then $f(I) \geq f(J)$.

Intuitively, the function $f_{m,k}$ from Definition 1 assigns to each sequence I of k positions the number of points that a committee C gets from a voter v when the members of C stand on exactly the positions of I in the preference order of v.

A committee scoring rule is defined by a family of committee scoring functions $f = (f_{m,k})_{k \leq m}$, which contains one function for each possible choice of m and k. Analogously to the case of single-winner scoring rules, we will denote such a multiwinner rule by $\mathcal{R}_f$. Let $E = (C, V)$ be an election with m candidates and let k, $k \leq m$, be the size of the desired committee. Under the committee scoring rule $\mathcal{R}_f$, every committee $S \subseteq C$ with $|S| = k$ receives score$(S) := \sum_{v \in V} f_{m,k}(\text{pos}_v(S))$ points (for this notation, the election $E = (C, V)$ will always be clear from the context). The committee with the highest score wins. (If there are several such committees, then they all tie as winners.)

Many well-known multiwinner voting rules are, in fact, committee scoring rules. Consider the following examples (we will use them throughout the paper to illustrate various points):

1. The SNTV, Bloc, and k-Borda rules pick k candidates with the highest Plurality, k-Approval, and Borda scores, respectively, and so they are defined through the following scoring functions:

$$
\begin{aligned}
f_{m,k}^{\text{SNTV}} (i_1, \ldots, i_k) &:= \sum_{t=1}^{k} \alpha_1(i_t) = \alpha_1(i_1), \\
f_{m,k}^{\text{Bloc}} (i_1, \ldots, i_k) &:= \sum_{t=1}^{k} \alpha_k(i_t), \\
f_{m,k}^{k\text{-Borda}}(i_1, \ldots, i_k) &:= \sum_{t=1}^{k} \beta_m(i_t).
\end{aligned}
$$

Note that $f_{m,k}^{\text{SNTV}}$ is defined as a sum of functions that do not depend on either m or k, $f_{m,k}^{\text{Bloc}}$ is defined as a sum of functions that depend on k but not m, and $f_{m,k}^{k\text{-Borda}}$ is defined as a sum of functions that depend on m but not k.[5]

2. The two versions of the Chamberlin–Courant rule that we consider are defined through the following committee scoring functions, respectively:

$$
\begin{aligned}
f_{m,k}^{\beta\text{-CC}} (i_1, \ldots, i_k) &:= \beta_m(i_1), \\
f_{m,k}^{\alpha_k\text{-CC}}(i_1, \ldots, i_k) &:= \alpha_k(i_1).
\end{aligned}
$$

The first one defines the classical Chamberlin-Courant rule (Chamberlin and Courant 1983) and the second one defines what we refer to as the k-Approval Chamberlin–Courant rule [approval-based variants of the Chamberlin–Courant rule were first mentioned by Thiele (1895) and recently they were recalled by Procaccia et al. (2008); they were studied further, for example, by Betzler et al. (2013), Aziz et al. (2017), and Skowron and Faliszewski (2017)]. For brevity, we sometimes refer to the k-Approval Chamberlin–Courant rule as the α_k-CC rule.

Intuitively, under the Chamberlin–Courant rules, each voter is represented by the committee member that this voter ranks highest; the Chamberlin–Courant rule chooses a committee S that maximizes the sum of the scores that the voters give to their representatives in S (which characterizes the total satisfaction of the society with the assignment of representatives to the voters).

3. We introduce the Perfectionist rule. This rule is defined through scoring functions of the form:

$$
f_{m,k}^{\text{Perf}}(i_1, \ldots, i_k) := \alpha_k(i_k).
$$

In other words, a voter gives score of 1 to a committee only if its members occupy the top k positions of his or her vote. The rule is not necessarily very appealing, but

---

[5] One could define the Borda scoring function so that $\beta_m(i) = -i$, removing the dependence on m. However, the traditional definition, $\beta_m(i) = m - i$, is much more common and, on the formal ground, we require the values of the scoring functions to be nonnegative.

it has interesting features that will illustrate several points that we make throughout our discussion.

Below we provide an example election where SNTV, Bloc, k-Borda, β-CC, α_k-CC, and Perfectionist give different outcomes (with the exception that the results of Bloc and α_k-CC are the same).

Example 1 Let us consider the set of candidates $C = \{a, b, c, d, e, f, g, h\}$ and eight voters with the following preference orders:

$$v_1 : a \succ f \succ c \succ g \succ h \succ e \succ b \succ d, \quad v_2 : c \succ e \succ g \succ h \succ a \succ f \succ b \succ d,$$
$$v_3 : a \succ f \succ c \succ h \succ g \succ e \succ b \succ d, \quad v_4 : d \succ e \succ h \succ g \succ a \succ f \succ b \succ c,$$
$$v_5 : b \succ c \succ g \succ h \succ a \succ e \succ f \succ d, \quad v_6 : e \succ g \succ d \succ h \succ a \succ b \succ f \succ c,$$
$$v_7 : b \succ d \succ h \succ g \succ a \succ e \succ f \succ c, \quad v_8 : f \succ h \succ d \succ g \succ a \succ b \succ e \succ c.$$

Let the committee size k be 2. It is easy to compute the winners under the SNTV and Bloc rules. For the former, the unique winning committee is $\{a, b\}$ (these are the only two candidates that are ranked in the top positions twice), and for the latter it is $\{e, f\}$ (these are the only two candidates that are ranked among top two positions three times; all the other candidates are ranked there at most twice). A somewhat tedious calculation shows that the unique k-Borda winning committee is $\{g, h\}$, which follows since the Borda scores of the candidates a, b, c, d, e, f, g, h are, respectively:

$$32, \ 22, \ 23, \ 23, \ 28, \ 26, \ 35, \ 35.$$

Further calculations show that under the (classical) Chamberlin–Courant rule, the unique winning committee is $\{c, d\}$. (While it is tedious to compute these results by hand, and indeed we used a computer to find them, the intuition for the k-Borda and Chamberlin–Courant winners is as follows: g and h are always ranked in the middle of each vote, or slightly above, so that they get high total Borda score, whereas c and d are ranked so that one of them is (almost) always ahead of g and h, whereas the other one is in the last position. This way, as representatives, c and d get higher scores than g and h, even though their total Borda score is lower.)

On the other hand, it is relatively easy to verify that under α_k-CC, the winning committee is $\{e, f\}$ (its α_k-CC score is six; there is no other committee whose members are ranked among the top two positions of six or more voters).

Finally, let us consider the Perfectionist rule. It assigns two points to committee $\{a, f\}$, one point to each of $\{b, c\}$, $\{b, d\}$, $\{c, e\}$, $\{d, e\}$, $\{e, g\}$, and $\{f, h\}$, and zero points to all the other committees. Thus, $\{a, f\}$ is the unique winning committee.

All the above rules are examples of OWA-based committee scoring rules, i.e., their committee scoring functions can be expressed as ordered weighted averages (OWAs) of single-winner scores. Formally, an OWA operator Λ of dimension k is a sequence $\Lambda = (\lambda^1, \dots, \lambda^k)$ of nonnegative reals[6] and the class of OWA-based rules (due to Skowron et al. 2016) is defined as follows.

---

[6] We slightly generalize the notion and, unlike Yager (1988), we do not require that $\lambda^1 + \cdots + \lambda^k = 1$.

Definition 2 Let $\Lambda = (\Lambda_{m,k})_{k \leq m}$ be a family of OWA operators such that $\Lambda_{m,k} = (\lambda_{m,k}^1, \ldots, \lambda_{m,k}^k)$ has dimension k (one size-k vector for each pair m, k). Let $\gamma = (\gamma_{m,k})_{k \leq m}$ be a family of (single-winner) scoring functions (one scoring function for each pair m, k). Then γ together with Λ define a family of committee scoring functions $f = f_{m,k}(\Lambda, \gamma)$ such that for each $(i_1, \ldots, i_k) \in [m^k]$ we have:

$$f_{m,k}(i_1, \ldots, i_k) = \sum_{t=1}^k \lambda_{m,k}^t \gamma_{m,k}(i_t).$$

The committee scoring rule $\mathcal{R}_f$ corresponding to the family f is called *OWA-based*.

Intuitively, the OWA operators specify to what extent the voters care about each member of the committee, depending on how this member is ranked among the other ones. For example, rules with OWA operators of the form $(1, \ldots, 1)$, such as SNTV, Bloc, or k-Borda, care about all the committee members equally, whereas rules with OWA operators of the form $(1, 0, \ldots, 0)$, such as our two versions of the Chamberlin–Courant rule, care about the top-ranked committee members only. Rules of the first type are called *weakly separable*, and those of the second type are called *representation focused* (Elkind et al. 2017; Faliszewski et al. 2016). Naturally, there are also many other choices of OWA operators. For example, the t-Approval variant of the Proportional Approval Voting rule (α_t-PAV) uses OWA operators of the form $(1, \frac{1}{2}, \ldots, \frac{1}{k})$, indicating the decreasing attention the voters pay to their lower-ranked committee members; the Perfectionist rule uses the OWA operator $(0, \ldots, 0, 1)$. For more discussions regarding the OWA-based rules, we refer the reader to the works of Skowron et al. (2016), Aziz et al. (2017, 2015), and Lackner and Skowron (2017) (the latter ones include a more detailed discussion of PAV; see also the work of Kilgour (2010) for a description of this rule).

Remark 1 We note that in most cases the OWA vectors $\Lambda_{m,k}$ used to define OWA-based rules do not depend on m. Yet, formally, we allow for such a dependence in order to build the relation between our general framework in which committee scoring functions $f_{m,k}$ might depend on m in any, even not very intuitive, way, and the world of OWA-based rules.

Naturally, there are also committee scoring rules that are not OWA-based. For example, Faliszewski et al. (2017) study the family of ℓ_p-Borda rules, with committee scoring functions of the following form ($p \geq 1$):

$$f_{m,k}^{\ell_p\text{-Borda}}(i_1, \ldots, i_k) := \sqrt[p]{\sum_{t=1}^k \beta_m^p(i_t)}.$$

In particular, they discuss how the ℓ_p-Borda rules (for $p > 1$) are, in a certain sense, between the k-Borda rule (which is simply the ℓ_1-Borda rule) and the classical Chamberlin–Courant rule (which, with slight abuse of notation, could be referred to as ℓ_∞-Borda).

3 Fixed-majority consistent rules

We are ready to start our quest for finding committee scoring rules that can be seen as multiwinner analogues of Plurality. We begin by describing the fixed-majority criterion that, in our view, encapsulates the idea of "closeness" to Plurality. Then, we provide a class of committee scoring rules—the class of top-k-counting rules—that contains all the rules which satisfy the fixed-majority criterion. Finally, we provide an almost complete characterization of those top-k-counting rules that have the fixed-majority property.

3.1 Initial remarks

One of the features that distinguishes Plurality among all the other scoring rules is the fact that it satisfies the simple majority criterion.

Definition 3 A single-winner voting rule $\mathcal{R}$ satisfies the *simple majority criterion* if for every election $E = (C, V)$ where more than half of the voters rank some candidate c first, it holds that $\mathcal{R}(E) = \{c\}$.

Importantly, the simple majority criterion indeed characterizes Plurality within the class of single-winner scoring rules. The result is a part of folklore (we provide the proof for the sake of completeness).

Proposition 1 *Let $\gamma = (\gamma_m)_{m \in \mathbb{N}}$ be a family of single-winner scoring functions that defines a scoring rule $\mathcal{R}_\gamma$. Then, $\mathcal{R}_\gamma$ satisfies the simple majority criterion if and only if for each m it holds that $\gamma_m(1) > \gamma_m(2) = \cdots = \gamma_m(m)$ (that is, if and only if $\mathcal{R}_\gamma$ coincides with Plurality).*

Proof It is straightforward to verify that if for each m we have $\gamma_m(1) > \gamma_m(2) = \cdots = \gamma_m(m)$ then $\mathcal{R}_\gamma$ satisfies the simple majority criterion. For the other direction, assume that $\mathcal{R}_\gamma$ satisfies the simple majority criterion. This immediately implies that for each $m \geq 2$ we have $\gamma_m(1) > \gamma_m(m)$ (otherwise all the candidates would always tie as winners). Hence for $m = 2$ the result follows.

Let us fix $m \geq 3$. For each positive integer n, define the election $E_n = (C, V_n)$ with the candidate set $C = \{c_1, \ldots, c_m\}$ and with V_n containing:

$$n + 1 \text{ voters with preference order } c_1 \succ c_2 \succ \cdots \succ c_m \text{ and}$$
$$n \text{ voters with preference order } c_2 \succ c_3 \succ \cdots \succ c_m \succ c_1.$$

Since $\mathcal{R}_\gamma$ satisfies the simple majority criterion, it must be the case that c_1 is the unique $\mathcal{R}_\gamma$-winner for each E_n. Further, for a given value of n, the difference between the scores of c_1 and c_2 in E_n is:

$$\text{score}(c_1) - \text{score}(c_2) = \big((n+1)\gamma_m(1) + n\gamma_m(m)\big) - \big((n+1)\gamma_m(2) + n\gamma_m(1)\big)$$
$$= \gamma_m(1) - \gamma_m(2) + n\big(\gamma_m(m) - \gamma_m(2)\big).$$

Thus, if it held that $\gamma_m(2) > \gamma_m(m)$, then—for large enough value of n—candidate c_1 would not be a winner of E_n. This implies that $\gamma_m(2) = \cdots = \gamma_m(m)$. Since $\gamma_m(1) > \gamma_m(m)$, we reach the conclusion that $\gamma_m(1) > \gamma_m(2) = \cdots = \gamma_m(m)$. $\qquad\square$

There are at least two ways of generalizing the simple majority criterion to the multiwinner setting. We choose perhaps the simplest one, the *fixed-majority criterion* introduced by Debord (1993) (other notions of majority studied by Debord are variants of the Condorcet principle and are incompatible with Plurality and scoring rules in general).

Definition 4 A multiwinner voting rule $\mathcal{R}$ satisfies the *fixed-majority criterion* for m candidates and committee size k if for every election $E = (C, V)$ with m candidates the following holds: if there is a committee W of size k such that more than half of the voters rank all the members of W above the non-members of W (equivalently: put the candidates from W on top), then $\mathcal{R}(E, k) = \{W\}$. We say that $\mathcal{R}$ satisfies the fixed-majority criterion if it satisfies it for all choices of m and k (with $k \leq m$).

Remark 2 Another possible way of extending the simple majority criterion to the multiwinner case would be to say that if a committee W is such that for each $c \in W$ a majority of voters rank c among their top k positions (possibly a different majority for each c), then W must be a winning committee. However, consider the following votes over the candidate set $\{a, b, c\}$:

$$v_1 : a > b > c, \quad v_2 : a > c > b, \quad v_3 : b > c > a.$$

For $k = 2$, all three committees, $\{a, b\}$, $\{a, c\}$, and $\{b, c\}$ have majority support in the sense just described. We feel that this is against the spirit of the simple majority criterion (since at most one candidate can be ranked on the top position by more than half of the voters, we feel that there should be at most one committee that can claim to have the majority support). Thus, and since we have not found any other convincing ways of generalizing the simple majority criterion to the multiwinner setting, we focus on Debord's fixed-majority notion.

It seems that the fixed-majority criterion is far more important for the multiwinner setting than the simple majority criterion is for the single-winner one. For example, one can verify that the Bloc rule satisfies the fixed-majority criterion and, in fact, this property is crucial in explaining its inner workings (we characterize the Bloc rule as the unique committee scoring rule that is noncrossing monotone and that satisfies the fixed-majority criterion[7]). This is important as in practice Bloc is among the most commonly used multiwinner rules. Further, the fixed-majority property may be useful when arguing that a given voting rule is appropriate for a setting where the selected committee needs strong legitimization: If a rule fails the fixed-majority property, then it is possible that even though a majority of the voters agree which committee is the

---

[7] For a discussion of noncrossing monotonicity that leads to this characterization, see the work of Faliszewski et al. (2016).

best, a different committee is elected (whose legitimacy might be questioned by this majority).[8]

While the Bloc rule satisfies the fixed-majority criterion, the SNTV rule does not (it will follow formally from our further discussion). This means that in our axiomatic sense, Bloc is closer to Plurality than SNTV. This is quite interesting since one's first idea of generalizing Plurality would likely be to think of SNTV. Yet, Bloc is certainly not the only committee scoring rule that satisfies our criterion. For example, the Perfectionist rule satisfies the fixed-majority criterion and, indeed, closely resembles Plurality. The following remark strongly highlights this similarity.

Remark 3 Consider a situation where the voters extend their rankings of candidates to rankings of committees in some natural way (see, e.g., the work of Barberà et al. (2004) for an overview of how this may be done). Then, for each voter, the best committee would consist of his or her k best candidates. As a result, running Plurality on the profile of preferences over the committees would give the same result as running Perfectionist over the profile of preferences over the candidates.

Naturally, not all committee scoring rules satisfy the fixed-majority criterion. For example, neither k-Borda nor the Chamberlin–Courant rule do. To see this, it suffices to note that for $k = 1$ they both become the single-winner Borda rule, which fails the simple majority criterion.

3.2 Top-k-counting rules

To characterize the committee scoring rules that satisfy the fixed-majority criterion, we introduce the class of scoring functions that depend only on the number of committee members ranked in the top k positions.

Definition 5 We say that a committee scoring function $f_{m,k} \colon [m]_k \to \mathbb{R}_+$ is *top-k-counting* if there is a function $g_{m,k} \colon \{0, \ldots, k\} \to \mathbb{R}_+$ such that $g_{m,k}(0) = 0$ and for each $(i_1, \ldots, i_k) \in [m]_k$ we have $f_{m,k}(i_1, \ldots, i_k) = g_{m,k}(|\{t \in [k] \colon i_t \leq k\}|)$. We refer to $g_{m,k}$ as the *counting function* for $f_{m,k}$. We say that a committee scoring rule $\mathcal{R}_f$ is *top-k-counting* if it can be defined through a family of top-k-counting scoring functions $f = (f_{m,k})_{k \leq m}$.

Both Bloc and Perfectionist are top-k-counting rules. The former uses the linear counting function $g_{m,k}(x) = x$, while the latter uses the counting function $g_{m,k}$ which is a step-function: $g_{m,k}(x) = 0$ for $x < k$ and $g_{m,k}(k) = 1$. Another example of a top-k-counting rule is the α_k-CC rule, which uses the counting function $g_{m,k}$ such that $g_{m,k}(0) = 0$ and $g_{m,k}(x) = 1$ for all $x \in [k]$.

Top-k-counting rules have a number of interesting features. First, their counting functions have to be nondecreasing. Second, every top-k-counting rule is OWA-based. Third, every committee scoring rule that satisfies the fixed-majority criterion is top-k-counting. We express these facts in the following two propositions and in Theorem 4.

---

[8] Naturally, on its own, the fact that a rule is fixed-majority consistent is not sufficient to claim that this rule is good for such a setting.

For the rest of the paper we make the assumption that $m \geq 2k$; this assumption is technical as our arguments are greatly simplified by the fact that we can form two disjoint committees of size k. Further, it is also quite natural: one could say that if we were to choose a committee consisting of more than half of the candidates, then perhaps we should rather be voting for who should *not* be in the elected committee. We are not sure whether this assumption can be dropped.

Proposition 2 *Let $m \geq 2k$ and let $f_{m,k} \colon [m]_k \to \mathbb{R}_+$ be a top-k-counting scoring function defined through a counting function $g_{m,k}$. Then, $g_{m,k}$ is nondecreasing.*

Proof Let $t \in \{0, \ldots, k\}$ be a number. Consider the sequences $I_t = (1, \ldots, t, k + 1, \ldots, k+(k-t))$ and $I_{t+1} = (1, \ldots, t+1, k+1, \ldots, k+(k-t-1))$ from $[m]_k$. (Note that we need $m \geq 2k$ for defining I_0.) Since $I_{t+1} \succeq I_t$, we have that $f_{m,k}(I_{t+1}) \geq f_{m,k}(I_t)$. By the definition, however, we have that $f_{m,k}(I_{t+1}) = g_{m,k}(t+1)$ and $f_{m,k}(I_t) = g_{m,k}(t)$. Hence, $g_{m,k}(t+1) \geq g_{m,k}(t)$. □

Without the assumption that $m \geq 2k$, Proposition 2 would have to be phrased more cautiously, and would speak only of the existence of a nondecreasing counting function. (For example, for $m = k$, the function $g_{m,k}$ could be arbitrary.)

Proposition 3 *Every top-k-counting rule is OWA-based.*

Proof Let us consider a top-k-counting rule $\mathcal{R}_f$, where $f = (f_{m,k})_{k \leq m}$ is the corresponding family of top-k-counting functions defined by a family of counting functions $(g_{m,k})_{k \leq m}$. Let us consider one function $f_{m,k}$ from this family. We know that $f_{m,k} \colon [m]_k \to \mathbb{R}_+$ is a top-k-counting scoring function defined through a counting function $g_{m,k}$ so that $f_{m,k}(i_1, \ldots, i_k) = g_{m,k}(s)$, where $s = |\{t \in [k] \colon i_t \leq k\}|$. As $g_{m,k}(0) = 0$, we have

$$f_{m,k}(i_1, \ldots i_k) = g_{m,k}(s) - g_{m,k}(0) = \sum_{t=1}^{s}(g_{m,k}(t) - g_{m,k}(t-1))$$
$$= \sum_{t=1}^{k} \alpha_k(i_t) \cdot (g_{m,k}(t) - g_{m,k}(t-1)),$$

from which we see that $\mathcal{R}_f$ is OWA-based through the family of OWA operators:

$$\Lambda_{m,k} = (g_{m,k}(1) - g_{m,k}(0), g_{m,k}(2) - g_{m,k}(1), \ldots, g_{m,k}(k) - g_{m,k}(k-1)),$$

and the family of k-Approval scoring functions $(\gamma_{m,k} = \alpha_k)$. □

In the next theorem (and in many further theorems) we speak of a committee scoring rule $\mathcal{R}_f$ defined through a family of committee scoring functions $f = (f_{m,k})_{2k \leq m}$. We use this notation as a shorthand for the assumption that the theorem is restricted to the cases where $2k \leq m$.

Theorem 4 *Let $f = (f_{m,k})_{2k \leq m}$ be a family of committee scoring functions. Then, if $\mathcal{R}_f$ satisfies the fixed-majority criterion, then $\mathcal{R}_f$ is top-k-counting.*

Proof Let us fix two numbers m and k such that $2k \leq m$. Consider an election with m candidates, where a committee of size k is to be elected. For each positive integer t such that $0 \leq t \leq k$ we define the following two sequences from $[m]_k$:

1. $I_t = (1, \ldots, t, k+1, \ldots, k+k-t)$ is a sequence of positions of the candidates where the first t candidates are ranked in the top t positions and the remaining $k - t$ candidates are ranked just below the kth position.
2. $J_t = (k - (t-1), \ldots, k, m - ((k-t) - 1), \ldots, m)$ is a sequence of positions where the first t candidates are ranked just above (and including) the kth position, whereas the remaining $k - t$ candidates are ranked at the bottom.

Among these, $I_k = (1, \ldots, k)$ is the highest-scoring sequence of positions and $J_k = (m - (k-1), \ldots, m)$ is the lowest-scoring sequence. Further, for every t we have $I_t \succeq J_t$ and, in effect, $f_{m,k}(I_t) \geq f_{m,k}(J_t)$.

We claim that if there exists some $t \in \{0, \ldots, k\}$ such that $f_{m,k}(I_t) > f_{m,k}(J_t)$ then $\mathcal{R}_f$ does not have the fixed-majority property. For the sake of contradiction, assume that there is some t such that $f_{m,k}(I_t) > f_{m,k}(J_t)$. Let $E = (C, V)$ be an election with m candidates and $2n + 1$ voters. The set of candidates is $C = X \cup Y \cup Z \cup D$, where $X = \{x_1, \ldots, x_t\}$, $Y = \{y_{t+1}, \ldots, y_k\}$, $Z = \{z_{t+1}, \ldots, z_k\}$, and D is a set of sufficiently many dummy candidates so that $|C| = m$. We focus on two committees, $M = X \cup Y$ and $N = X \cup Z$. The first $n + 1$ voters have preference order $X \succ Y \succ Z \succ D$, and the next n voters have preference order $Z \succ X \succ D \succ Y$. Note that the fixed-majority criterion requires that M be the unique winning committee.

Committee M receives the total score of $(n + 1) f_{m,k}(I_k) + n f_{m,k}(J_t)$, whereas committee N receives the total score of $(n + 1) f_{m,k}(I_t) + n f_{m,k}(I_k)$. The difference between these values is:

$$(n + 1) f_{m,k}(I_k) + n f_{m,k}(J_t) - (n + 1) f_{m,k}(I_t) - n f_{m,k}(I_k)$$
$$= f_{m,k}(I_k) + n f_{m,k}(J_t) - (n + 1) f_{m,k}(I_t) =$$
$$= f_{m,k}(I_k) - f_{m,k}(I_t) + n\big(f_{m,k}(J_t) - f_{m,k}(I_t)\big),$$

which, for a large enough value of n, is negative (since, by assumption, we know that $f_{m,k}(J_t) < f_{m,k}(I_t)$ and so $f_{m,k}(J_t) - f_{m,k}(I_t)$ is negative). That is, for large enough n, committee M does not win the election and $\mathcal{R}_f$ fails the fixed-majority criterion.
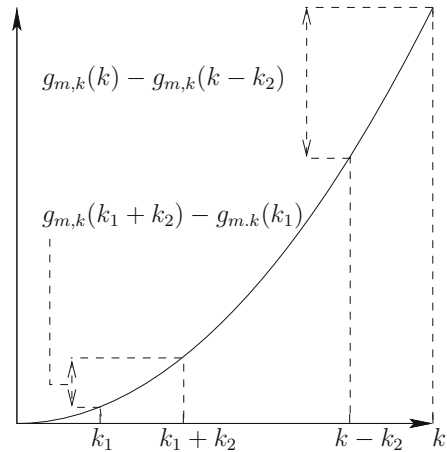
So, if $\mathcal{R}_f$ satisfies the fixed-majority criterion, then for every $t \in \{0, \ldots, k\}$ we have that $f_{m,k}(I_t) = f_{m,k}(J_t)$. This, however, means that $f_{m,k}$ is a top-k-counting scoring function. To see this, consider some sequence of positions $L = (\ell_1, \ldots, \ell_k) \in [m]_k$ where exactly the first t entries are smaller than or equal to k. Clearly, we have that $I_t \succeq L \succeq J_t$ and so $f_{m,k}(I_t) = f_{m,k}(L) = f_{m,k}(J_t)$, which means that $f_{m,k}(i_1, \ldots, i_k)$ depends only on the cardinality of the set $\{t \in [k] : i_t \leq k\}$. Since m and k were chosen arbitrarily (with $2k \leq m$), this completes the proof. $\qquad\square$

Unfortunately, the converse of Theorem 4 does not hold: α_k-CC, for example, is a top-k-counting rule that fails the fixed-majority criterion.

Example 2 Consider an election $E = (C, V)$ with $C = \{a, b, c, d\}$, $V = (v_1, v_2, v_3)$, and $k = 2$. Let the preference orders of the voters be:

$$v_1 : a \succ b \succ c \succ d, \qquad v_2 : a \succ b \succ c \succ d, \qquad v_3 : c \succ d \succ a \succ b.$$

The fixed-majority criterion requires $\{a, b\}$ to be the only winning committee, while under α_k-CC, other committees, such as $\{a, c\}$, have strictly higher scores. (Incidentally, this example also witnesses that SNTV fails the fixed-majority criterion; this is hardly surprising since SNTV is not a top-k-counting rule.)

3.3 Criterion for fixed-majority consistency

In this section, we provide a formal characterization of those top-k-counting rules that satisfy the fixed-majority criterion. Together with Theorem 4, this gives an almost full characterization of committee scoring rules with this property.

Theorem 5 *Let $f = (f_{m,k})_{2k \leq m}$ be a family of committee scoring functions with the corresponding family $(g_{m,k})_{2k \leq m}$ of counting functions. Then, $\mathcal{R}_f$ satisfies the fixed-majority criterion if and only if for every $k, m \in \mathbb{N}$, $2k \leq m$, it holds that:*

  (i) *$g_{m,k}$ is not constant, and*
 (ii) *for each pair of nonnegative integers k_1, k_2 with $k_1 + k_2 \leq k$, we have that:*

$$g_{m,k}(k) - g_{m,k}(k - k_2) \geq g_{m,k}(k_1 + k_2) - g_{m,k}(k_1).$$

(Condition (ii) in Theorem 5 is a relaxation of the convexity property for function $g_{m,k}$ and is illustrated in Fig. 1; We discuss this in more detail after the proof of the theorem.)

Proof of Theorem 5 Let $f_{m,k}$ be one of the committee scoring functions and $g_{m,k}$ be its corresponding counting function. By Proposition 2, $g_{m,k}$ is nondecreasing so the fact that it is non-constant is equivalent to $g_{m,k}(k) > g_{m,k}(0)$. Moreover, we note that conditions (i) and (ii) imply that for each k' with $0 \leq k' \leq k - 1$, we have $g_{m,k}(k) > g_{m,k}(k')$. To see this we take $k_2 = 1$ and note that for each k_1 it holds that $g_{m,k}(k) - g_{m,k}(k - 1) \geq g_{m,k}(k_1 + 1) - g_{m,k}(k_1)$. As $g_{m,k}(k) > g_{m,k}(0)$, for

some k_1 we have that $g_{m,k}(k_1 + 1) - g_{m,k}(k_1) > 0$. Thus, $g_{m,k}(k) > g_{m,k}(k - 1)$. Since $g_{m,k}$ is nondecreasing, it is also true that $g_{m,k}(k - 1) \geq g_{m,k}(k')$. It follows that $g_{m,k}(k) > g_{m,k}(k')$.

Let us now show that if for each m and k, $g_{m,k}$ satisfies (ii), then $\mathcal{R}_f$ has the fixed-majority property. Let $E = (C, V)$ be an election with n voters and m candidates for which there is a size-k committee M such that a majority of the voters rank all members of M in the top k positions, but M loses to some committee $S \neq M$ (also of size k). That is, we have score$(S) \geq$ score(M). Let ξ be a rational number, $\frac{1}{2} < \xi \leq 1$, such that exactly ξn voters rank all the members of M in the top k positions; we will refer to these voters as M-voters and to the others as non-M-voters.

Without loss of generality, we can assume that all the non-M-voters have identical preference orders. Indeed, if it were the case that $f_{m,k}(\text{pos}_{v_i}(S)) - f_{m,k}(\text{pos}_{v_i}(M)) > f_{m,k}(\text{pos}_{v_j}(S)) - f_{m,k}(\text{pos}_{v_j}(M))$ for some two non-M-voters v_i and v_j, then we could replace the preference order of v_j with that of v_i and increase the advantage of S over M. If for all non-M-voters this difference were the same, then we could simply pick the preference order of one of them and assign it to all the other ones.

Let k_1, k_2, k_3, and k_4 be four numbers such that:

1. k_1 is the number of candidates from $S \cap M$ that the non-M-voters rank among their top k positions,
2. k_2 is the number of candidates from $S \setminus M$ that the non-M-voters rank among their top k positions,
3. k_3 is the number of candidates from $C \setminus (S \cup M)$ that the non-M-voters rank among their top k positions, and
4. k_4 is the number of candidates from $M \setminus S$ that the non-M-voters rank among their top k positions.

Without loss of generality, we can assume that $k_4 = 0$ and that $|S \setminus M| = k_2$ (since $m \geq 2k$, we can replace all members of $M \setminus S$ with candidates from $C \setminus M$, and, similarly, we can ensure that all members of $S \setminus M$ are ranked among the top k positions by non-M-voters; these changes never decrease the score of S relative to that of M). In effect, we have that $k_1 + k_2 + k_3 = k$ and, since $|S \cap M| + |S \setminus M| = k$, we have that $|S \cap M| = k - k_2$. We can assume that $k_2 > 0$ as otherwise we would have $S = M$. Given this notation, the difference between the scores of M and S is:

$$
\begin{aligned}
\text{score}(M) - \text{score}(S) &= \xi n \cdot g_{m,k}(k) + (1 - \xi)n \cdot g_{m,k}(k_1) - \xi n \cdot g_{m,k}(k - k_2) \\
&\quad - (1 - \xi)n \cdot g_{m,k}(k_1 + k_2) \\
&= \xi n \cdot \big(g_{m,k}(k) - g_{m,k}(k - k_2)\big) - (1 - \xi)n \cdot \big(g_{m,k}(k_1 + k_2) - g_{m,k}(k_1)\big) \\
&> 0,
\end{aligned}
$$

where the second equality holds due to rearranging of terms, and the final inequality is an immediate consequence of the assumptions regarding the value of ξ and the properties of $g_{m,k}$ (namely, that $g_{m,k}(k) - g_{m,k}(k - k_2) \geq g_{m,k}(k_1 + k_2) - g_{m,k}(k_1)$ and that $g_{m,k}(k) - g_{m,k}(k - k_2) > 0$). This, however, contradicts the assumption that score$(S) \geq$ score(M) and, so, $\mathcal{R}_f$ satisfies the fixed-majority criterion.

We now consider the other direction. For the sake of contradiction, let us assume that $\mathcal{R}_f$ satisfies the fixed-majority criterion but that there exist m and k such that it is not the case that conditions (i) and (ii) are both satisfied. If condition (i) is not satisfied and $g_{m,k}$ is a constant function, then $\mathcal{R}_f$ fails the fixed-majority criterion because it always outputs all the subsets of size k, independently of the voters' preferences. Thus we assume that $g_{m,k}$ is not constant. Thus, suppose that condition (ii) does not hold and there exist k_1 and k_2 with $k_1 + k_2 \leq k$ such that $g_{m,k}(k) - g_{m,k}(k - k_2) < g_{m,k}(k_1 + k_2) - g_{m,k}(k_1)$. We form an election with m candidates, $c_1, \ldots, c_m$, and $2n + 1$ voters (we describe the choice of n later). The first $n + 1$ voters have preference order:

$$c_1 \succ c_2 \succ \cdots \succ c_m,$$

and the remaining n voters have preference order:

$$c_1 \succ \cdots \succ c_{k_1} \succ c_m \succ c_{m-1} \succ \cdots \succ c_{k_1+1}.$$

Since $\mathcal{R}_f$ satisfies the fixed-majority criterion, in this election it outputs the unique winning committee $M = \{c_1, \ldots, c_k\}$. However, consider committee S:

$$S = \{c_1, \ldots, c_{k_1+k_2}, c_m, \ldots, c_{m-(k-k_1-k_2)+1}\}.$$

Since $m \geq 2k$, the difference between the scores of M and S is:

$$
\begin{aligned}
&\text{score}(M) - \text{score}(S) \\
&= (n+1)g_{m,k}(k) + ng_{m,k}(k_1) - (n+1)g_{m,k}(k_1+k_2) - ng_{m,k}(k-k_2) \\
&= n\big(g_{m,k}(k) - g_{m,k}(k-k_2)\big) + g_{m,k}(k) \\
&\quad - n\big(g_{m,k}(k_1+k_2) - g_{m,k}(k_1)\big) - g_{m,k}(k_1+k_2).
\end{aligned}
$$

Since $g_{m,k}(k) - g_{m,k}(k - k_2) < g_{m,k}(k_1 + k_2) - g_{m,k}(k_1)$, we observe that for large enough n the difference $\text{score}(M) - \text{score}(S)$ becomes negative. This is a contradiction showing that (ii) holds. □

Let us take a step back and consider what condition (ii) from Theorem 5 means (recall Fig. 1). Intuitively, it resembles the convexity condition, but 'focused' on $g_{m,k}(k)$ (see the explanation below).

Definition 6 Let $g_{m,k}$ be a counting function for some top-k-counting function $f_{m,k}\colon [m]_k \to \mathbb{R}_+$. We say that $g_{m,k}$ is *convex* if for each k' such that $2 \leq k' \leq k$, it holds that:

$$g_{m,k}(k') - g_{m,k}(k' - 1) \geq g_{m,k}(k' - 1) - g_{m,k}(k' - 2).$$

On the other hand, we say that g is *concave* if for each k' with $2 \leq k' \leq k$ it holds that:

$$g_{m,k}(k') - g_{m,k}(k' - 1) \leq g_{m,k}(k' - 1) - g_{m,k}(k' - 2).$$

Using inductive reasoning, we see that the above definition of a convex top-k-counting function is equivalent to requiring that for each k', k'', and d such that $k'' \leq k' \leq k$, $k'' - d \geq 0$, and $k' - d \geq 0$, it holds that:

$$g_{m,k}(k') - g_{m,k}(k' - d) \geq g_{m,k}(k'') - g_{m,k}(k'' - d).$$

Condition (ii) of Theorem 5 is of the same form, except that we fix k' to be k (i.e., we 'focus on $g_{m,k}(k)$'), set $d = k_2$, and set $k'' = k_1 + k_2$.

The notions of convexity and concavity are standard, but allow us to express many features of top-k-counting rules in a very intuitive way. For example, the following corollary is an immediate consequence of Theorem 5.

Corollary 6 *Let $f = (f_{m,k})_{2k \leq m}$ be a family of top-k-counting committee scoring functions with the corresponding family $(g_{m,k})_{2k \leq m}$ of counting functions. The following statements hold:*

1. *if $g_{m,k}$ are convex, then $\mathcal{R}_f$ satisfies the fixed-majority criterion, and*
2. *if $g_{m,k}$ are concave but not linear (that is, $\mathcal{R}_f$ is not Bloc) then $\mathcal{R}_f$ fails the fixed-majority criterion.*

The counting function for the Bloc rule is linear (and, thus, both convex and concave), and the counting function for the Perfectionist rule is convex, so these two rules satisfy the fixed-majority criterion. On the other hand, the counting function for α_k-CC is concave and, so, this rule fails the criterion (as we observed in Example 2). (It may be helpful to remark here that committee scoring rules are uniquely represented by their committee scoring functions, up to affine transformations; this result is provided in the technical report version of the work of Faliszewski et al. (2016).)

By Proposition 3, a family of concave counting functions $g_{m,k}$ corresponds to a nonincreasing OWA operator and a family of convex counting functions corresponds to a nondecreasing one. Skowron et al. (2016) provided evidence that rules based on nonincreasing OWA operators are computationally easier than those based on general OWA operators (while computing the exact winning committees tends to be computationally hard in both cases, there are, for example, polynomial-time constant-factor approximation algorithms whenever the operators are nonincreasing; unless P = NP, such algorithms do not exist for many rules based on the other OWA operators). In Sect. 4 we show that this seems to be the case for top-k-counting rules as well, but we also provide a striking example highlighting a certain dissimilarity.

3.4 Characterization of Bloc within committee scoring rules

We conclude this section by noting that Theorems 4 and 5, together with a result of Faliszewski et al. (2016), suffice to characterize Bloc within the class of committee scoring rules. To present this result, we need the following definition of Elkind et al. (2017):

Definition 7 A multiwinner rule $\mathcal{R}$ is noncrossing-monotone if the following holds: Whenever committee W of size k is winning in some election E, then W also is winning

in every election E' resulting from shifting some member c of W one position forward in some vote (provided that c does not pass any other member of W).

Faliszewski et al. (2016) have shown that a committee scoring rule is noncrossing monotone if and only if it is weakly separable, that is, if and only if its scoring functions $f = (f_{m,k})_{k \leq m}$ are of the form:

$$f_{m,k}(i_1, \ldots, i_k) = \gamma_{m,k}(i_1) + \gamma_{m,k}(i_2) + \cdots + \gamma_{m,k}(i_k), \tag{1}$$

where $\gamma = (\gamma_{m,k})_{k \leq m}$ is a family of single-winner scoring functions. Since the scoring functions of the Bloc rule are the only top-k-counting scoring functions of this form [this also follows by uniqueness of representation of committee scoring rules (Faliszewski et al. 2016)], by Theorems 4 and 5 we get the following corollary.

Corollary 7 *Bloc is the only committee scoring rule that is both fixed-majority consistent and noncrossing monotone.*

This corollary calls for two comments. First, the reader may complain that Theorems 4 and 5 assume that the number of candidates is at least twice as large as the committee size, but in Corollary 7 we do not make this assumption. Indeed, Theorems 4 and 5 suffice for Corollary 7 only for the case where $2k \leq m$. For the case where $2k > m$, one can show that the result still holds by using the fact that noncrossing monotonicity guarantees that our committee scoring rule have scoring functions of the form (1). Indeed, it suffices to consider an election with:

$$n + 1 \text{ votes of the form } c_1 \succ c_2 \succ \cdots \succ c_k \succ c_{k+1} \succ \cdots \succ c_m, \text{ and}$$
$$n \text{ votes of the form } c_{k+1} \succ c_1 \succ \cdots \succ c_{k-1} \succ c_{k+2} \succ \cdots \succ c_m \succ c_k.$$

If it were the case that $\gamma_{m,k}(1) > \gamma_{m,k}(k)$ then, for sufficiently large n, candidate c_{k+1} would have higher $\gamma_{m,k}$-score than c_k in the above election and, in consequence, the committee $\{c_1, \ldots, c_k\}$ would not be winning. Thus our rule would not be fixed-majority consistent. The same would hold if we had that $\gamma_{m,k}(1) = \cdots = \gamma_{m,k}(k)$, but $\gamma_{m,k}(k + 1) > \gamma_{m,k}(m)$: For sufficiently large n, c_{k+1} would have higher $\gamma_{m,k}$ score than c_k. Naturally, if $\gamma_{m,k}(1) = \cdots = \gamma_{m,k}(m)$, then all committees would always win and the rule would not be fixed-majority consistent either. Thus the only functions $\gamma_{m,k}$ remaining are such that $\gamma_{m,k}(1) = \cdots = \gamma_{m,k}(k) > \gamma_{m,k}(k + 1) = \cdots = \gamma_{m,k}(m)$. Such functions generate exactly the Bloc rule, which is fixed-majority consistent.

Faliszewski et al. (2016) characterize Bloc as the only committee scoring rules that is top-k-counting and weakly separable, which is the same result as ours, but phrased in terms of syntactic properties of scoring functions and not in terms of axiomatic properties.

4 Complexity of top-k-counting rules

In this section, we consider the computational complexity of winner determination for top-k-counting rules which are based on either convex or concave counting func-

tions. Throughout this section, we focus on committee scoring functions of the form $f_{m,k} \colon [m]_k \to \mathbb{N}$, that is, on functions that always return nonnegative integers as scores. This is a technical assumption, motivated by the fact that representing arbitrary real numbers on a computer can be problematic. To avoid confusion, we mention this assumption explicitly in each relevant theorem.

Remark 4 For a committee scoring rule $\mathcal{R}_f$, when we say that this rule is NP-hard to compute, we formally mean that, given an election $E = (C, V)$, a committee size k, and a nonnegative integer T, the problem of deciding if there exists a committee S of size k whose score is at least T is NP-hard. Indeed, if we were able to compute an $\mathcal{R}_f$ winning committee of size k in polynomial time, then we could solve this decision problem in polynomial time as well, by checking if the score of the winning committee is at least T (provided that f were polynomial-time computable). Conversely, if we knew that our decision problem were NP-hard, then we would also know that the ability to compute winning committees under $\mathcal{R}_f$ implies the ability to solve NP-hard problems.

We start by considering several examples. It is well-known that Bloc winners can be computed in polynomial time; this is so since one can compute the score of each candidate separately. It turns out that the same holds for the Perfectionist rule, albeit following different reasoning.

Proposition 8 *Both Bloc and Perfectionist winners are computable in polynomial time.*

Proof The case of Bloc is well-known (to form a winning committee of size k it suffices to pick k candidates with the highest k-Approval scores). To find a size-k winning committee under the Perfectionist rule, for each voter v we consider the set of his or her top-k candidates as a committee, and compute the score of that committee in the election. We output those committees—among the considered ones—that have the highest score. Correctness follows by noting that the committees considered by the algorithm are the only ones with nonzero scores. □

While the above result is very simple, it is also very interesting. For example, Perfectionist is the first example of a polynomial-time computable committee scoring rule that is not weakly separable [see the discussions of Elkind et al. (2017) and Faliszewski et al. (2016)]. Further, it stands in sharp contrast to the results of Skowron et al. (2016). By Proposition 3, Perfectionist is defined through the OWA operator $(0, \ldots, 0, 1)$, and Skowron et al. have shown that, in general, rules defined through this operator are NP-hard to compute and very difficult to approximate. Their result, however, relies on the fact that voters can approve any number of candidates, while in our case they must approve exactly k of them. This shows very clearly that even though top-k-counting rules are OWA-based, we cannot simply carry-over the computational hardness results of Skowron et al. (2016) or Aziz et al. (2015) to our framework.

We can generalize Proposition 8 to rules that are, in some sense, similar to Perfectionist. To this end, and to facilitate our later discussion regarding the com-

plexity of top-k-counting rules, we define the following property of counting functions.

Definition 8 Let $g_{m,k}$ be a counting function for a top-k-counting function $f_{m,k} \colon [m]_k \to \mathbb{N}$. We define the *singularity* of $g_{m,k}$, denoted by $\mathrm{sing}(g_{m,k})$, to be

$$\mathrm{sing}(g_{m,k}) = \operatorname*{arg\,min}_{2 \le i \le k} \big(g_{m,k}(i) - g_{m,k}(i-1) \ne g_{m,k}(i-1) - g_{m,k}(i-2)\big).$$

Loosely speaking, $\mathrm{sing}(g_{m,k})$ is the smallest integer in $\{2, \ldots, k\}$ for which the differential of $g_{m,k}$ changes. For Bloc (which is an exception) we define $\mathrm{sing}(g_{m,k})$ to be ∞, since the differential is a constant function. Naturally, for all other non-constant rules, the singularity is finite. For example, for Perfectionist we have $\mathrm{sing}(g_{m,k}) = k$.

We generalize the polynomial-time algorithm for Perfectionist to similar rules, for which the value $\mathrm{sing}(g_{m,k})$ is close to k.

Proposition 9 *Let $\mathcal{R}_f$ be a top-k-counting rule for a family $f = (f_{m,k})_{2k \le m}$ of polynomial-time computable committee scoring functions ($f_{m,k} \colon [m]_k \to \mathbb{N}$) with the corresponding family of counting functions $(g_{m,k})_{2k \le m}$. Let q be a constant, positive integer such that $k - \mathrm{sing}(g_{m,k}) \le q$ holds for all m and k. Then $\mathcal{R}_f$ has a polynomial-time computable winner determination problem.*

Proof Let the input consist of election $E = (C, V)$ and positive integer k, and let W be a winning committee in $\mathcal{R}(E, k)$. We assume that $q < \frac{k}{2}$ (if it were not the case, then $k \le 2q$ would be small and we could solve the problem using brute-force). We consider two cases: (1) there is at least one voter that has at least $\mathrm{sing}(g_{m,k})$ of his or her top k candidates in W; (2) every voter has less than $\mathrm{sing}(g_{m,k})$ of his or her top k candidates in W.

If case (1) holds, then we can compute W (or some other winning committee) by checking, for each voter v, all the committees that consist of at least $\mathrm{sing}(g_{m,k})$ candidates that v ranks among his or her top k positions. Since $k - \mathrm{sing}(g_{m,k}) \le q$, the number of committees that we have to check for each voter is:

$$\sum_{t=\mathrm{sing}(g_{m,k})}^{k} \binom{k}{t}\binom{m}{k-t} \le (q+1) \cdot \binom{k}{k - \mathrm{sing}(g_{m,k})}\binom{m}{k - \mathrm{sing}(g_{m,k})},$$

which is a polynomial in k and m. The above inequality requires some care: We have that $\mathrm{sing}(g_{m,k}) > \frac{k}{2}$ (because $k - \mathrm{sing}(g_{m,k}) \le q < \frac{k}{2}$) and, in effect, we have that for each $t \in \{\mathrm{sing}(g_{m,k}), \ldots, k\}$ it holds that $\binom{k}{t} = \binom{k}{k-t} \le \binom{k}{k-\mathrm{sing}(g_{m,k})}$ and $\binom{m}{k-t} \le \binom{m}{k-\mathrm{sing}(g_{m,k})}$.

If case (2) holds, then from the fact that $g_{m,k}(x) - g_{m,k}(x-1)$ is a constant for $x \le \mathrm{sing}(g_{m,k})$, we infer that $g_{m,k}(x)$ is effectively linear. Then, it suffices to compute the winning committee using the Bloc rule. While we do not know which of the two cases holds, we can compute the two committees, one as in case (1) and one as in case (2), and output the one with the higher score (or either of them, in case of a tie). $\qquad\square$

Example 3 Consider the following committee scoring function:

$$f'_{m,k}(i_1, \ldots, i_k) = f_{\text{Bloc}}(i_1, \ldots, i_k) + f_{\text{Perf}}(i_1, \ldots, i_k)$$
$$= \alpha_k(i_1) + \cdots + \alpha_k(i_{k-1}) + 2\alpha_k(i_k).$$

As a simple application of Proposition 9, we get that the committee scoring rule $\mathcal{R}_{f'}$ defined through f' is polynomial-time computable. This rule can be seen as a variant of Bloc, where a voter gives additional one bonus point to a committee if he or she approves of all its members. By Corollary 6, this rule is fixed-majority consistent.

It is also interesting to consider the rule which is defined through the following committee scoring function:

$$f''_{m,k}(i_1, \ldots, i_k) = f_{\text{SNTV}}(i_1, \ldots, i_k) + f_{\text{Perf}}(i_1, \ldots, i_k) = \alpha_1(i_1) + \alpha_k(i_k).$$

The corresponding rule is also polynomial-time computable (it suffices to compute an SNTV winning committee, and compare it with such committees whose all members stand on first k positions in some voter's preference ranking), but it is not a top-k-counting rule and, so, it fails the fixed-majority criterion.

Yet, as one might expect, not all top-k-counting rules are polynomial-time solvable and, indeed, most of them are not (under standard complexity-theoretic assumptions). For example, α_k-CC is NP-hard (this follows quite easily from Theorem 1 of Procaccia et al. (2008); we include a brief proof to substantiate the discussion and give the reader some intuition).

Proposition 10 *For α_k-CC it is* NP-*hard to decide whether or not there exists a committee with at least a given score (recall that k in α_k-CC is the committee size and, thus, is part of the input).*

Proof sketch The NP-hardness follows easily from a standard reduction from the EXACT COVER BY 3-SETS problem, abbreviated as X3C. In an instance of X3C we are given a family of m subsets, $S_1, \ldots, S_m$, each of cardinality 3, chosen from a given universal set $U = \{x_1, \ldots, x_{3n}\}$, and we ask if there are n subsets from the family whose union is U. Additionally, we assume that each element of U belongs to at most three subsets [it is well-known that this variant of X3C remains NP-complete (Garey and Johnson 1979)].

Given an instance of X3C, we create a candidate for each subset and a voter for each element of U. Voters rank the subsets to which they belong in their top positions, then they rank some n dummy candidates (different ones for each voter), and then all the remaining candidates (in some arbitrary, easy to compute, order). We ask for a committee of size $k = n$ (and we assume that $n \geq 3$; this is a technical assumption

as for $n = 1$ and $n = 2$ our construction is formally incorrect[9]). There is a winning committee with score $3n$ if and only if the answer for the input instance is "yes." □

We generalize the above NP-hardness result to the case of convex top-k-counting rules $\mathcal{R}_f$ for which there is some constant c such that for each k and m it holds that $k - \text{sing}(g_{m,k}) \geq k/c$ (that is, to the case of convex counting functions for which the differential changes 'early'). The proof of this result is fairly technical and is available in Appendix A.

Theorem 11 *Let $\mathcal{R}_f$ be a top-k-counting rule defined through a family f of top-k-counting functions $f_{m,k} : [m]_k \to \mathbb{N}$ ($f_{m,k} : [m]_k \to \mathbb{N}$) with the corresponding family of counting functions $(g_{m,k})_{k \leq m}$ that do not depend on m, $g_{m,k} = g_k$, and such that:*

1. *For each x, $0 \leq x \leq k$, $g_k(x)$ is computable in polynomial time with respect to k (that is, there is a polynomial time algorithm that given x and k outputs $g_k(x)$). Moreover, for each k, $g_k(k)$ is polynomially bounded in k.*
2. *There is a constant c such that, for each size of committee k greater than some fixed constant k_0, g_k is convex and $k - \text{sing}(g_k) \geq k/c$.*

Then, deciding if there is a committee with at least a given score is NP-*hard for $\mathcal{R}_f$.*

Let us now discuss the assumptions of the theorem, where they come from and why we believe they are natural (or necessary).

First, the assumption that the counting functions are computable in polynomial time is standard and clear. Indeed, it would not be particularly interesting to seek hardness results if already the counting functions were hard to compute.

Second, we believe that the assumption that the counting functions $g_{m,k}$ do not depend on m is reasonable. For example, it is quite intuitive that adding some candidates that all the voters rank last should not have any effect on the committee selected by a top-k-counting rule. (The assumption is also very helpful on the technical level. Our construction uses a number of dummy candidates that depends on the values of the counting function. If the values of the counting function depended on the number of candidates, we might end up with a very problematic, circular dependence.)

Third, the assumption that there is a constant c such that for any large enough committee size k we have $k - \text{sing}(g_k) \geq k/c$ says that the function "shows its convex behavior" early enough. As shown in Proposition 9, some assumption of this form is necessary (though there is still a gap, since the bounds from the theorem and from Proposition 9 do not match perfectly), and it is the core of the theorem.

Finally, perhaps the least intuitive assumption in this theorem is the requirement that for a given committee size k, the highest value of the counting function is polynomially bounded in k. The reason for having it is that, if the highest value were extremely large (say, exponentially large with respect to k) then, for sufficiently few voters (for example, polynomially many), the rule might degenerate to a polynomial-time

---

[9] This is so, because we require that each voter (i.e., each element) ranks each of the sets to which he or she belongs among top k positions. By assumption, each element belongs to at most 3 sets, so we need $k = n \geq 3$. Given an instance of X3C where $n = 1$ or $n = 2$, we can solve it using a brute-force algorithm (we have to either try each set or each pair of sets as a solution) and, depending on the outcome, either output a fixed *yes*-instance or a fixed *no*-instance of our problem.

computable one (for example, it might resemble the Perfectionist rule for this case). Exactly to avoid such problems, in our proof we use a number of voters that depends on $g_k(k)$. Our reduction would not run in polynomial time if $g_k(k)$ were superpolynomial.

A result similar to Theorem 11, but for concave rules, is possible as well [and, in essence, follows from the proofs of Skowron et al. (2016) and Aziz et al. (2015)]. Thus, in general, top-k-counting functions tend to be NP-hard to compute. What can we do if we need to use them anyway? There are several possibilities. Next we consider approximability and fixed-parameter tractability as possible approaches.

4.1 Approximability

First, for concave top-k-counting rules we can obtain a constant-factor approximation algorithm [we deduce it from the result of Skowron et al. (2016), which—in essence—boils down to optimizing a submodular function using the seminal results of Nemhauser et al. (1978)]. In particular, the next result applies to the α_k-PAV rule (that is, to the top-k-counting rule based on the OWA operators of the form $(1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{k})$; recall its discussion from Sect. 2).

Theorem 12 *Let $\mathcal{R}_f$ be a top-k-counting rule defined through a family f of (polynomial-time computable) top-k-counting functions $f_{m,k} \colon [m]_k \to \mathbb{N}$ with corresponding counting functions $g_{m,k}$ that are concave. Then there is a polynomial-time algorithm that, given an election E and a committee size k, computes a committee W of size k, whose score, under $\mathcal{R}_f$, is at least a $(1 - \frac{1}{e})$ fraction of the score of the winning committee(s) from $\mathcal{R}_f(E, k)$.*

Proof This follows from the fact that concave top-k-counting rules correspond to OWA-based rules that use nonincreasing OWA operators. For such rules, there is a $(1 - \frac{1}{e})$-approximation algorithm for computing the score of the winning committees and for computing a committee with such a score (Skowron et al. 2016, Theorem 4). □

Such a general result for convex counting functions seems impossible. Let us consider a convex counting function $g_{m,k}(x) = \max(x - 1, 0)$ that is nearly identical to the linear counting function used by Bloc. Let us refer to the top-k-counting rule defined by $(g_{m,k})_{k \leq m}$ as NearlyBloc. If we had a polynomial-time constant-factor approximation algorithm for NearlyBloc, we would have a constant-factor approximation algorithm for the DENSEST AT MOST K SUBGRAPH problem (abbreviated as DAMKS; see below). Taking into account the results of Khuller and Saha (2009), Raghavendra and Steurer (2010), and Alon et al. (2011), this seems very unlikely.

Given a graph G, we refer to its sets of vertices and edges as $V(G)$ and $E(G)$, respectively. The *density* of a graph G is defined as $\delta = \frac{|E(G)|}{|V(G)|}$.

Definition 9 In the DENSEST AT MOST K SUBGRAPH problem, DAMKS, we are given a graph G and we ask for a subgraph of G of the highest possible density with at most K vertices.

The proof of the next theorem is available in Appendix B.

Theorem 13 *There is no polynomial-time constant-factor approximation algorithm for the problem of computing the score of a winning committee under NearlyBloc, unless such an algorithm exists for the* DAMKS *problem.*

Nonetheless, for top-k-counting rules that are not too far from α_k-CC, we have a polynomial-time approximation scheme (PTAS), that is, an algorithm that can achieve any desired approximation ratio, as long as the number of candidates is not too large relative to the committee size. This result holds even for rules that are not concave (provided they satisfy the conditions of the theorem); the result follows by noting that our voters have non-finicky utilities (Skowron et al. 2016).

Theorem 14 *Let $\mathcal{R}_f$ be a top-k-counting committee scoring rule, where the family $f = (f_{m,k})_{k \leq m}$ ($f_{m,k} \colon [m]_k \to \mathbb{N}$) is defined through a family of counting functions $(g_{m,k})_{k \leq m}$ that are: (a) polynomial-time computable and (b) constant for arguments greater than some given value ℓ. If $m = o(k^2)$, then there is a PTAS for computing the score of a winning committee under $\mathcal{R}_f$.*

Proof We use the concept of *non-finicky utilities* provided by Skowron et al. (2016). Adapting their terminology, we say that a single-winner scoring function $\gamma_m \colon [m] \to \mathbb{N}$ (for elections with m candidates) is (ξ, δ)-non-finicky for $\xi, \delta \in [0, 1]$, if each of the highest $\lceil \delta m \rceil$ numbers in the sequence $\gamma_m(1), \ldots, \gamma_m(m)$ is greater or equal to $\xi \gamma_m(1)$. It is easy to see that α_k is $(1, \frac{k}{m})$-non-finicky.

Consider an input election $E = (C, V)$ with m candidates, and committee size k, such that $m = o(k^2)$. By Proposition 3, we know that $f_{m,k}$ is OWA-based, that it uses some OWA operator $\Lambda_{m,k}$ that has nonzero entries on the top ℓ positions only, and that it uses scoring function α_k (which is a $(1, \frac{k}{m})$-non-finicky). Thus, due to Skowron et al. (2016), there is a polynomial-time $\left(1 - \ell \exp\left(-\frac{k^2}{m\ell^2}\right)\right)$-approximation algorithm for computing the score of a winning committee under f. Using the assumption that $m = o(k^2)$, the approximation ratio of the algorithm is:

$$\alpha = 1 - \ell \exp\left(-\frac{k^2}{m\ell^2}\right)$$
$$= 1 - \ell \exp\left(-\frac{k^2}{o(k^2)\ell^2}\right)$$
$$= 1 - \ell \exp\left(-\frac{1}{o(1)}\right) = 1 - o(1).$$

This completes the proof. □

Theorem 14 is quite remarkable even for the case of α_k-CC (let alone that it applies to a somewhat more general set of rules). Indeed, generally, variants of the Chamberlin–Courant rule that use some sort of approval scoring function are hard to compute (Procaccia et al. 2008; Betzler et al. 2013) and the best possible approximation ratio for a polynomial-time algorithm, in the general case, is $1 - \frac{1}{e}$ [this result was observed by Skowron and Faliszewski (2017) and follows from results for the MaxCover problem (Feige 1998)]. This upper bound, however, relies on the fact that

there is no connection between the size of the input election, the committee size, and the number of candidates that each voter approves. We obtain a PTAS because we assume that for the committee size k each voter approves of k candidates, and that the number m of candidates is such that $m = o(k^2)$.

One may ask how likely it is that this last assumption holds. As a piece of anecdotal evidence, we mention that in the 2015 parliamentary elections in Poland, there were $k = 460$ seats in the parliament and $m \approx 8000$ candidates. In this case, $m/k^2 \approx 0.0378$, which suggests that our algorithm could be effective (provided that the voters could say which k candidates they approve of; likely, this would require some sort of simplified ballots, for example, allowing one to approve blocks of candidates).

4.2 Fixed-parameter tractability

If one were not interested in approximation algorithms but still wanted to use top-k-counting rules, then one might seek fixed-parameter tractable algorithms. In parameterized complexity we concentrate on some distinguished parameter of the problem, such as the number of candidates or the number of voters. We say that a parameterized problem is fixed-parameter tractable (is in FPT) if there is an algorithm that, given an instance of this problem of size n with parameter t, computes an answer for the problem in time $f(t)n^{O(1)}$, where f is some computable function (such an algorithm is also said to run in FPT time with respect to parameter t). For a detailed description of parameterized complexity, we point the readers to the books by Downey and Fellows (1999), Niedermeier (2006), and Cygan et al. (2015).

We start with a simple observation, namely that a winning committee can be computed for every top-k-counting rule in FPT time for the parameterization by the number of candidates.

Proposition 15 *Let $\mathcal{R}_f$ be a top-k-counting committee scoring rule, where the family $f = (f_{m,k})_{k \leq m}$ ($f_{m,k} \colon [m]_k \to \mathbb{N}$) is defined through a family of counting functions $(g_{m,k})_{k \leq m}$ (that are computable in* FPT *time with respect to m). There is an algorithm that, given a committee size k and an election E, computes a winning committee from $\mathcal{R}_f(E, k)$ in* FPT *time with respect to the number m of candidates.*

Proof The algorithm simply computes the score of every possible committee and outputs the one with the highest score. With m candidates and committee size k, the algorithm has to check $\binom{m}{k} = O(m^m)$ committees, and checking each committee requires FPT time only. □

For rules based on concave counting functions we can also provide a far less trivial FPT algorithm for the parameterization by the number of voters (the proof, which uses a somewhat technical trick on top of solving a mixed integer linear program is available in Appendix C). The algorithm applies, for example, to the α_k-PAV rule, which uses OWA operators of the form $(1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{k})$, so its counting functions are of the form $g_{m,k}^{\text{PAV}}(x) = \sum_{t=1}^{x} \frac{1}{t}$, and are concave. (See Sect. 2 for literature pointers regarding the PAV rule.)

Theorem 16 *Let $\mathcal{R}_f$ be a top-k-counting committee scoring rule, where the family $f = (f_{m,k})_{k \leq m}$ ($f_{m,k} \colon [m]_k \to \mathbb{N}$) is defined through a family of concave counting functions $(g_{m,k})_{k \leq m}$ (that are polynomial-time computable). There is an algorithm that, given a committee size k and an election E, computes a winning committee from $\mathcal{R}_f(E, k)$ in FPT time with respect to the number n of voters.*

To summarize, it appears that most (but certainly not all) top-k-counting rules are NP-hard to compute. For top-k-counting rules based on concave counting functions, there are good polynomial-time approximation algorithms and some exact FPT algorithms. On the other hand, for rules based on convex functions the situation is much more difficult. Aside from several algorithms that do not depend on concavity or convexity of the counting function (for instance the algorithms from Theorem 14 and Proposition 15), so far we only have evidence for computational hardness.

5 Related literature

The rules considered in this paper form a subfamily of the OWA-based rules of Skowron et al. (2016). A specific subclass of OWA-based rules—when voters express their preferences in the form of approval ballots—has been already mentioned in the early work of Thiele (1895). More recently, Aziz et al. (2017), Brill et al. (2017), and Sánchez-Fernández et al. (2017) analyzed selected axiomatic properties of the Thiele methods, and Aziz et al. (2015) studied their computational complexity. For a more general overview of approval-based multiwinner rules we refer the reader to the book by Kilgour (2010). It is also worth noting that there exist other OWA-based approaches to multiwinner voting [see, e.g., the work of Elkind and Ismaili (2015)], which, however, do not directly apply to our setting.

More generally, the class of OWA-based rules is a subclass of the class of committee scoring rules (Elkind et al. 2017). Committee scoring rules have been recently axiomatically characterized by Skowron et al. (2016), and Faliszewski et al. (2016) classified voting rules within this class in the form of a hierarchy. The studies of axiomatic properties of other committee scoring rules also include the work of Debord (1992), who characterized k-Borda voting rule. There is also a substantial literature describing axiomatic properties of other types of multiwinner rules—for an overview of this literature we refer the reader to the work of Elkind et al. (2017) and to the survey of Faliszewski et al. (2017).

Establishing the complexity of winner determination under various multiwinner rules is an active area of research. These studies were pioneered by Procaccia et al. (2008), who proved that computing winners under the Chamberlin–Courant committee scoring rule is NP-hard[10] and, in consequence, motivated many researchers to seek ways of circumventing this result. For example, Betzler et al. (2013) have shown that the rule is polynomial-time computable for the case of single-peaked preferences and Yu et al. (2013) have done the same for single-crossing ones [Elkind and Lackner (2015), Skowron et al. (2015b), Peters (2018), and Peters and Lackner (2017) provided further generalizations and improvements to these results]. Betzler et al. (2013) studied

---

[10] They also considered Monroe's rule, which is closely related but is not a committee scoring rule itself.

the problem from the perspective of parameterized complexity theory, whereas Lu and Boutilier (2011) analyzed the possibility of approximation and proved that a simple greedy procedure guarantees the approximation ratio of $1 - 1/e$ (the ratio relates the scores of the winning committee and the one provided by the algorithm). Later, Skowron et al. (2015a) improved this result by showing a polynomial-time approximation scheme. Oren and Lucier (2014) proved that if the voters arrive in a random order then the greedy algorithm can be easily adapted to the online setting, preserving the approximation ratio arbitrarily close to $1 - 1/e$; they also observed that for certain specific distributions of votes this approximation ratio can actually improve. Skowron and Faliszewski (2017) studied FPT approximation algorithms of the approval-based Chamberlin–Courant rule and Faliszewski et al. (2016) showed that often in practice the quality of approximation can be improved by employing certain clustering algorithms.

So far, analysis of the complexity of other committee scoring rules received far less attention, but this seems to be changing quickly. For example, it was shown that finding winners according to the proportional approval voting rule (the PAV rule), another approval-based committee scoring rule, is NP-hard (Aziz et al. 2015; Skowron et al. 2016), but there exist good approximation algorithms for the problem (Skowron et al. 2016; Skowron 2016; Byrka et al. 2017). The complexity of other selected subclasses of committee scoring rules has been studied by Skowron et al. (2016) and by Faliszewski et al. (2016). There also exists a literature studying the computational complexity of other multiwinner rules, which do not belong to the class of committee scoring rules, such as Minimax Approval Voting (MAV): finding winners under MAV is NP-hard (LeGrand 2004), yet there exists a PTAS for the problem (Byrka and Sornat 2014). Parameterized complexity and parameterized approximations of the rule were considered by Misra et al. (2015) and Cygan et al. (2017). The computational complexity of these and other important issues pertaining to MAV were considered by Baumeister et al. (2010, 2015, 2016), Baumeister and Dennisen (2015).

Our work regards the model of multiwinner elections where the voters rank the candidates and it is the voting rule's task to (implicitly) derive rankings of the committees (in a systematic way, according to the principles that underlie the given rule). Another approach, pioneered by Fishburn (1981a, b), is to require that the voters rank the committees explicitly. This approach is useful when there are dependencies between the candidates that are hard (or impossible) to capture within simple preference orders (e.g., when it is important that all members of an elected committee can work together), but can be used directly only in very limited settings (for example, there are 252 committees of five out of ten candidates; it would be unreasonable to ask voters to rank them all). In other cases, one has to rely on concise means of expressing voters' preferences, such as the formalism of CP-nets (Boutilier et al. 2004). Multiwinner elections of this type are often studied within the area of voting in combinatorial domains (Lang and Xia 2015).

6 Conclusions and further research

Aiming at finding a multiwinner analogue of the single-winner Plurality rule, we have shown that the answer is quite involved. While it is tempting to view SNTV as a natural analogue of Plurality, a closer look reveals that it fails the fixed-majority criterion (which Plurality satisfies in the single-winner setting). We have found that, among all committee scoring rules, only the top-k-counting rules—a class of rules we have defined in this paper—have a chance of satisfying the fixed-majority criterion, and we have characterized when this happens. Specifically, we have shown that the committee scoring rules which satisfy the fixed-majority criterion are exactly those top-k-counting rules whose counting functions satisfy a relaxed variant of convexity.

For example, the Bloc and Perfectionist rules both satisfy the fixed-majority criterion and, so, in some sense, they are among the multiwinner analogues of Plurality (for the Perfectionist rule this goes quite deep). On the other hand, a variant of the Chamberlin–Courant rule based on the k-Approval scoring function is top-k-counting, but fails the fixed-majority criterion.

We believe that it is very interesting to focus on top-k-counting rules based either on convex or on concave counting functions. These two classes of rules are different in some interesting ways: top-k-counting rules based on convex counting functions are fixed-majority consistent, but seem very hard to compute (with a few exceptions); this stands in contrast to top-k-counting rules based on concave counting functions, which fail the fixed-majority criterion (the borderline case of Bloc rule excluded), but are much easier to compute (typically still NP-hard, but with constant-factor polynomial-time approximation algorithms and FPT algorithms for the parameterization by the number of voters).

Our work leads to a number of open questions. In the axiomatic direction, it would be interesting to consider notions analogous to the fixed-majority criterion for the setting where voters do not provide preference orders but, instead, simply indicate which candidates they do or do not approve. On the computational front, it would be interesting to find more powerful algorithms for computing winning committees under various top-k-counting rules (e.g., for the α_k-PAV rule).

A Proof of Theorem 11

We prove NP-hardness of the problem by giving a reduction from the CLIQUE problem on regular graphs (a graph is regular if all its vertices have the same degree). In the CLIQUE problem we are given a graph G and an integer h, and we ask if there exists a set of h pairwise adjacent vertices in G (such a set of vertices is referred to as a *size-h clique*). The problem remains NP-complete when restricted to regular graphs (Garey and Johnson 1979).

Let G be the input regular graph, let h be the size of the clique sought for, and let δ be the common degree of G's vertices. If $h > \delta + 1$, then, of course, the graph does not contain a size-h clique and we output a fixed "no"-instance of our problem. Otherwise, we output an instance according to the following construction (intuitively, since each g_k is convex, the rule promotes situations where voters rank many members of the committee among their top k candidates; we exploit this fact).

We set the committee size k to be $(c + 2)h$ (recall that c is defined in the theorem statement). Since g_k does not depend on the number of candidates in the election, this fixes the counting function that we work with and we will denote it g. If $k \leq k_0$ (recall that k_0 is defined in the statement of the theorem), then we solve the input instance using brute force in polynomial time and output either a fixed "yes"-instance or a fixed "no"-instance, depending on the result.

We note that for each i, $1 \leq i \leq \mathrm{sing}(g)$, all the values $g(i) - g(i - 1)$ are equal and, without loss of generality, we can assume them to either all be 0s or all be 1s (if this were not the case, we could scale g appropriately). Similarly, since g is convex, we can assume that $g(\mathrm{sing}(g)) - g(\mathrm{sing}(g) - 1) > 1$. We note that $k - \mathrm{sing}(g) \geq k/c = (c + 2)h/c > h$ and, so, $\mathrm{sing}(g) < k - h$.

We form an election with the following candidates:

1. For each vertex v from the graph G, we create a candidate v.
2. We create a set $\{c_1, \ldots, c_{\mathrm{sing}(g)-2}\}$ of candidates, called the *edge-filler candidates*. These candidates will be in the top-k positions of all the voters, and hence will be chosen to every winning committee.
3. We create a set $\{b_1, \ldots, b_{k-h-(\mathrm{sing}(g)-2)}\}$ of candidates, called *general-filler candidates*. There will be sufficiently many voters who rank them in their top-k positions so that they will also be in every winning committee.
4. We also create a set of dummy candidates, such that each dummy candidate is ranked among the top-k positions of exactly one voter.

Let m be the total number of edges in G. For each edge e, we create a set of $2g(k)$ voters corresponding to this edge; each voter in this set has the following candidates in the top k positions of his or her preference order:

1. The two candidates corresponding to the endpoints of e.
2. All the edge-filler candidates.
3. Sufficiently many dummy candidates (such that they are ranked among top k positions only by this voter).

Further, we create $2g(k) \cdot (m + h) \cdot g(k)$ filler voters, who rank the following candidates in the top k positions:

1. All the edge-filler candidates.
2. All the general-filler candidates.
3. Sufficiently many dummy candidates (different dummy candidates for each filler voter).

(The role of the $2g(k)$ multiplicity factor regarding both the edge voters and the filler voters is to ensure that the best committee does not contain any of the dummy candidates; this will become clear later in the proof.)

We ask whether there is a committee W whose score is at least $T = T_1+T_2+T_3+T_4$, where:

$$
\begin{aligned}
T_1 &= 2g(k) \cdot (m + h) \cdot g(k) \cdot g(k - h), \\
T_2 &= 2g(k) \cdot m \cdot g(\text{sing}(g) - 2), \\
T_3 &= 2g(k) \cdot \delta h \cdot \big(g(\text{sing}(g) - 1) - g(\text{sing}(g) - 2)\big), \\
T_4 &= 2g(k) \cdot \binom{h}{2}\big(g(\text{sing}(g)) - g(\text{sing}(g) - 2) \\
&\quad - 2(g(\text{sing}(g) - 1) - g(\text{sing}(g) - 2))\big).
\end{aligned}
$$

Note that each T_i, $1 \le i \le 4$, is nonnegative (for T_4 this is due to convexity of g). The meaning of these values will become clear throughout the proof. This finishes the construction. Due to the assumptions regarding the counting function, the reduction is polynomial-time computable.

Let us now argue that the reduction is correct. First, we claim that if a committee W has a score of at least T, then it must contain all the edge-filler candidates and all the general-filler candidates. We note that altogether we have $k - h$ edge-filler and general-filler candidates. Consider some committee W' that contains $k - h - x$ candidates of these two types, where $x \ge 1$. This means that W' contains at most $h+x$ dummy candidates.

Let y be the number of filler voters that rank at least $k - h$ members of W' among their top k positions. Let us call these filler voters well-satisfied. For each of the well-satisfied filler voters, the members of W' ranked on top k positions are (a) the $k - h - x$ edge-filler and general-filler candidates from W', and (b) at least x unique dummy candidates. Thus it must hold that $xy \le h+x$ and, so, $y \le \frac{h}{x}+1$. If $x \ge 2$, then it must be that $y \le h$. If $x = 1$, then this inequality gives us that $y \le h + 1$. However, for y to be $h + 1$, W' would have to consist of $k - h - 1$ edge-filler and general-filler candidates and $h+1$ dummy candidates. Each of these dummy candidates would have to be ranked among top k positions by exactly one of the y well-satisfied filler voters. This would mean that for each edge voter, the only members of W' ranked by this voter among top k positions would be (some of) the edge-filler candidates. Consequently, all the edge voters would rank at most $k - h - 1$ members of W' among their top k positions. In either case (that is, irrespective if $x = 1$ or $x \ge 2$), we can upper-bound the score of committee W' by assuming that there are $2g(k) \cdot (m + h) \cdot g(k) - h$ voters that assign score $g(k - h - 1)$ to W' and $2g(k) \cdot m + h$ voters that assign score $g(k)$

to it. In effect, we have the following inequalities (also see the explanations below):

$$\begin{aligned}
\text{score}(W') &\leq \big(2g(k) \cdot (m+h) \cdot g(k) - h)\big) \cdot g(k-h-1) + (2g(k) \cdot m + h) \cdot g(k) \\
&= 2g(k) \cdot (m+h) \cdot g(k) \cdot g(k-h-1) - h \cdot g(k-h-1) \\
&\quad + (2g(k) \cdot m + h) \cdot g(k) \\
&< 2g(k) \cdot (m+h) \cdot g(k) \cdot \big(g(k-h)-1\big) - h \cdot g(k-h-1) \\
&\quad + (2g(k) \cdot m + h) \cdot g(k) \\
&= T_1 - 2g(k) \cdot (m+h) \cdot g(k) - h \cdot g(k-h-1) \\
&\quad + (2g(k) \cdot m + h) \cdot g(k) \\
&= T_1 - 2g(k) \cdot (m+h) \cdot g(k) - h \cdot g(k-h-1) \\
&\quad + 2g(k) \cdot m \cdot g(k) + h \cdot g(k) \\
&= T_1 - 2g(k) \cdot h \cdot g(k) - h \cdot g(k-h-1) + h \cdot g(k) \leq T_1 < T.
\end{aligned}$$

The second inequality holds because $g(k-h) > g(k-h-1)+1$ (which holds due to the fact that g is convex, $g(\text{sing}(g)) - g(\text{sing}(g)-1) > 1$, and $\text{sing}(g) < k-h$). Further inequalities hold due to simple calculations. Due to the above reasoning, we can assume that every committee with score at least T contains all the $k-h$ filler candidates.

Consider some committee that contains all the $k-h$ filler candidates. We claim that if this committee contains some dummy candidates then there is another committee with a higher score. Why is this so? Assume that the committee contains some z dummy candidates ($z \leq h$). If we simply removed these dummy candidates (obtaining a smaller committee) then we would lose at most $z \cdot g(k)$ points. Then, we could bring the committee back to its intended size by performing the following operations sufficiently many times: Either adding to the committee a single vertex candidate (already connected by an edge to one from the committee) or adding to the committee two vertex candidates connected by an edge. Each of these actions increases the score of the committee by at least $2g(k)\big(g(\text{sing}(g)) - g(\text{sing}(g)-1)\big) > 2g(k)$ (because for each edge there are $2g(k)$ corresponding edge voters). Thus, we would obtain a committee with a score higher than the one we have started with. (Note that, technically, there might be no sequence of operations that brings our committee back to size k, but this would only happen if the graph had too few edges to contain a clique of size h and we could recognize that this is the case in polynomial time.)

Let W be some winning committee that contains all the $k-h$ filler candidates, and some h vertex candidates (by the above paragraph, this committee cannot contain any dummy candidates), and let r be the number of edges that connect the vertices corresponding to the vertex candidates from W. Let us now calculate the score of W. The filler voters provide score T_1. The situation regarding the edge voters requires more care.

Each edge voter gets score at least $g(\text{sing}(g)-2)$ due to the edge-filler candidates. For each edge for which at least one endpoint is in W, we get additional $g(\text{sing}(g)-1) - g(\text{sing}(g)-2)$ points, and for each edge whose both endpoints are in W, we get yet additional $g(\text{sing}(g)) - g(\text{sing}(g)-1)$ points. Thus, the edge voters give W the

following score (see detailed explanations below):

$$\underbrace{\left(2g(k) \cdot m \cdot g(\text{sing}(g) - 2)\right)}_{=T_2}$$

$$+ \underbrace{\left(2g(k) \cdot \delta h \cdot \left(g(\text{sing}(g) - 1) - g(\text{sing}(g) - 2)\right)\right)}_{=T_3}$$

$$+ \underbrace{\left(2g(k) \cdot r \cdot \left(g(\text{sing}(g)) - g(\text{sing}(g) - 2) - 2(g(\text{sing}(g) - 1) - g(\text{sing}(g) - 2))\right)\right)}_{\leq T_4}.$$

The first main term corresponds to the points all the edge voters receive, the second is the correction for edge voters that correspond to edges that have at least one endpoint in W (note that if for some edge both its endpoints belong to W, then we add $g(\text{sing}(g) - 1) - g(\text{sing}(g) - 2)$ twice, once for each endpoint), and the final term corresponds to the correction for edges that have two endpoints in W. Let us now explain why this final correction is appropriate. Consider some edge voter for an edge whose both endpoints are in W. For this voter, we account $g(\text{sing}(g) - 2)$ points that each edge voter gets, we account $g(\text{sing}(g)-1)-g(\text{sing}(g)-2)$ points for each of the endpoints, and $g(\text{sing}(g)) - g(\text{sing}(g) - 1) - 2(g(\text{sing}(g) - 1) - g(\text{sing}(g) - 2))$ points of the final correction. Altogether, this sums up to:

$$g(\text{sing}(g) - 2) + 2(g(\text{sing}(g) - 1) - g(\text{sing}(g) - 2)) + g(\text{sing}(g))$$
$$- g(\text{sing}(g) - 1) - 2(g(\text{sing}(g) - 1) - g(\text{sing}(g) - 2)) = g(\text{sing}(g)).$$

This means that, indeed, we compute the score of edge voters for edges whose both endpoints are in W correctly. The same holds for all the other edge voters (and follows directly from the above analysis).

Finally, we note that the score of W that we obtain from the edge voters is maximized when r is maximized. The maximum value that r may have is $\binom{h}{2}$, which happens if and only if the vertex candidates in W correspond to a clique. Then the score that the edge voters provide equals $T_2 + T_3 + T_4$ and the total score of the committee is T.

We conclude, that there exists a committee with score at least T if and only if the input graph contains a size-h clique.

B Proof of Theorem 13

Let θ be a positive real, $0 < \theta < 1$. For the sake of contradiction, let us assume that there is a polynomial-time algorithm $\mathcal{A}$ that, given an election E and committee size k, outputs a committee W such that, under NearlyBloc the score of W is at least a θ fraction of the score of the winning committee. Using $\mathcal{A}$, we will derive a $\frac{\theta}{2}$-approximation algorithm for the DAMKS problem.

Let I be an instance of the DAMKS problem with a graph G and an integer K.

Our algorithm proceeds as follows. For each B, $1 \leq B \leq K$, we form an election $E_B = (C_B, V_B)$ where:

1. The set of candidates is $C_B = V(G) \cup \bigcup_{e \in E(G)} D_e$, where for each $e \in E(G)$, $D_e = \{d_{e,1}, \ldots d_{e,B-2}\}$ is the set of dummy candidates needed for our construction.
2. The collection V_B of voters is such that for each edge $e = \{u_1, u_2\} \in E(G)$ we have exactly one voter with preference order of the form $\{u_1, u_2\} \succ D_e \succ \cdots$.

For each election E_B, we run algorithm $\mathcal{A}$ to find a committee W_B of size B. Each such committee W_B generates an induced graph G_B with the vertex set $V(G) \cap W_B$. We let G_0 be the trivial subgraph of G consisting of two vertices and their connecting edge (if G had no edges, then we could output a trivial optimal solution at this point). We output the densest graph among $G_0, G_1, \ldots, G_K$.

Let us now argue that the above algorithm is a $\frac{\theta}{2}$-approximation algorithm for the DAMKS problem. Let OPT be an optimal solution for I, with the densest subgraph G' consisting of B vertices and X edges. By definition, G' has density $\delta = \frac{X}{B}$. For each B let us consider two cases:

Case 1: $X \leq \frac{B}{\theta}$. In this case, the density of the optimal graph is at most equal to $\frac{1}{\theta}$. However, a trivial solution with two vertices connected with an edge has density equal to $\frac{1}{2}$. Thus, in this case this trivial solution is $\frac{\theta}{2}$-approximate.

Case 2: $X > \frac{B}{\theta}$. In this case we know that there exists a size-B committee for election E_B with score at least X. Indeed, the committee that consists of the vertices from G' obtains one point for each edge from G' and has score X. Thus $\mathcal{A}$ for E_B and committee size B outputs a committee W' with score at least θX. Let $U' = W' \cap V(G)$ (that is, let U' be the part of this committee that consists of the vertex candidates) and let $D' = W' - U'$ (that is, let D' be the set of dummy candidates from W'). We observe that the graph induced by U' has at least $\theta X - |D'|$ edges. To see this, note that since each dummy candidate is ranked among top B positions by exactly one voter, removing a dummy candidate from the committee—in effect decreasing the committee size—decreases the total score by at most one. Thus the committee consisting only of candidates from U' has score at least $\theta X - |D'|$ and each of the points obtained by this committee comes from an edge between some members of U'.

The graph induced by U' has density δ' such that:

$$\delta' = \frac{\theta X - |D'|}{B - |D'|} = \frac{\theta X}{B} \cdot \frac{B(\theta X - |D'|)}{\theta X \cdot (B - |D'|)} = \theta \delta \cdot \frac{B(\theta X - |D'|)}{\theta X \cdot (B - |D'|)} \geq \theta \delta,$$

where the last inequality follows from the assumption that $B < \theta x$. Indeed, note that:

$$B(\theta X - |D'|) = \theta X B - B|D'| \geq \theta X B - \theta X|D'| = \theta X \cdot (B - |D'|).$$

By our assumptions, one of these conditions must hold. This means that the graph induced by U' is a θ-approximate solution for I.

Since in both cases we obtain at least $\frac{\theta}{2}$-approximate solutions, our algorithm is $\frac{\theta}{2}$-approximate. Since it is clear that it runs in polynomial time, the proof is complete.

C Proof of Theorem 16

Our algorithm is based on solving a mixed integer linear program (MILP) in FPT time with respect to the number of integral variables. The key trick is to use non-integral variables in such a way that in every optimal solution they have to take integral values [this technique was first used by Bredereck et al. (2015)].

Let k be the input committee size and $E = (C, V)$ be the input election, where $C = \{c_1, \ldots, c_m\}$ is the set of candidates, $V = (v_1, \ldots, v_n)$ is the collection of voters.

We enumerate all the nonempty subsets of V as $S_1, \ldots, S_{2^n-1}$. For each $i \in [2^n-1]$, let $\mathcal{T}(S_i)$ denote the largest set of candidates that satisfies the following condition: Every voter in S_i ranks each candidate from $\mathcal{T}(S_i)$ among the top k positions and no other voter ranks either of the candidates from $\mathcal{T}(S_i)$ among top k positions. Note that $\mathcal{T}(S_1), \ldots, \mathcal{T}(S_{2^n})$ is a partition of C. We illustrate this partition in the following example.

Example 4 Consider an election $E = (C, V)$ with $C = \{a, b, c, d, e, f\}$ and $V = (v_1, \ldots, v_6)$, where the voters have the following preference orders (we set the committee size $k = 3$ and, thus, we list only top k positions for each vote):

$$v_1 : c \succ d \succ f \succ \cdots, \quad v_2 : c \succ d \succ e \succ \cdots, \quad v_3 : a \succ b \succ c \succ \cdots,$$
$$v_4 : c \succ e \succ f \succ \cdots, \quad v_5 : d \succ e \succ f \succ \cdots, \quad v_6 : a \succ b \succ e \succ \cdots.$$

We have the following sets: $\mathcal{T}(\{v_3, v_6\}) = \{a, b\}$ since only voters v_3 and v_6 rank a and b on top three positions (and there are no other candidates they both rank among their top three positions). Then, we have: $\mathcal{T}(\{v_1, v_2, v_3, v_4\}) = \{c\}$, $\mathcal{T}(\{v_1, v_2, v_5\}) = \{d\}$, $\mathcal{T}(\{v_2, v_4, v_5, v_6\}) = \{e\}$, and $\mathcal{T}(\{v_1, v_4, v_5\}) = \{f\}$. For every other subset S_i of voters, we have $\mathcal{T}(S_i) = \emptyset$. For example, $\mathcal{T}(\{v_4, v_5\}) = \emptyset$ for the following reasons: The candidates that both v_4 and v_5 rank on top three positions are e and f. However, each of these candidates is ranked among top three positions also by some other voter(s).

Our algorithm forms a mixed integer linear program with the following variables. We have $2^n - 1$ integer variables, $z_1, \ldots z_{2^n-1}$, where, intuitively, each z_i describes how many candidates from the set $\mathcal{T}(S_i)$ we take into the winning committee. For each $i \in [n]$ we also have an integer variable x_i, which describes how many candidates from the top k positions of the preference order of voter v_i belongs to the winning committee. Finally, for each variable x_i, we have rational variables $x_{i,j}, 0 \leq x_{i,j} \leq 1$, such that (intuitively) each $x_{i,j}$ is 1 if x_i is at least j. We present our mixed integer linear program in Fig. 2. To solve this program, we invoke Lenstra's famous result in its variant for mixed integer programming (Lenstra 1983, Section 5).

Now it remains to argue that it indeed outputs a correct solution, that is, that the variables $z_1, \ldots, z_{2^n-1}$ describe a winning committee. If all the variables have the intended, intuitive values (as described in the preceding paragraph), then—with our

$$\text{maximize } \sum_{i=1}^{n} \sum_{j=1}^{k} x_{i,j} \cdot (g_{m,k}(j) - g_{m,k}(j-1))$$

subject to:

(a) $\displaystyle\sum_{i=1}^{2^n-1} z_i = k,$

(b) $\displaystyle x_i = \sum_{j:\, i \in S_j} z_j,$ $\qquad\qquad i \in [n]$

(c) $\displaystyle\sum_{j=1}^{k} x_{i,j} = x_i,$ $\qquad\qquad i \in [n]$

(d) $0 \le x_{i,j} \le 1,$ $\qquad\qquad i \in [n]; j \in [k]$

(e) $0 \le z_i \le |\mathcal{T}(S_i)|,$ $\qquad\qquad i \in [2^n - 1]$

Fig. 2 The Mixed Integer Linear Program used in the proof of Theorem 16

maximization goal in mind—one can verify that variables $z_1, \ldots, z_{2^n-1}$ describe a winning committee. Thus we show that, indeed, all the variables have their intended values.

Due to constraints (a) and (e), variables $z_1, \ldots, z_{2^n-1}$ certainly describe a possible committee of size k (from each set $\mathcal{T}(S_i)$ we take z_i arbitrary candidates). Constraints (b) ensure the correct values of variables $x_1, \ldots, x_n$. Finally, the maximization goal and constraints (c) ensure that each variable $x_{i,j}$ is 1 exactly if $x_i \ge j$ and is 0 otherwise. This is so, because $g_{m,k}$ is concave. Thus, if for some values j and j' with $j < j'$ it were the case that $x_{i,j} < 1$ and $x_{i,j'} > 0$ then increasing $x_{i,j}$ and decreasing $x_{i,j'}$ by the same amount [without breaking constraint (d)] would yield a higher value of the function to be maximized.

References

Alon N, Arora S, Manokaran R, Moshkovitz D, Weinstein O (2011) Inapproximabilty of densest k-subgraph from average case hardness. http://www.nada.kth.se/~rajsekar/papers/dks.pdf. Accessed 15 Apr 2018

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

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, pp 107–115

Barberà S, Bossert W, Pattanaik P (2004) Ranking sets of objects. In: Barberà S, Hammond P, Seidl C (eds) Handbook of utility theory. Springer, New York, pp 893–977

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

Baumeister D, Böhnlein T, Rey L, Schaudt O, Selker AK (2016) Minisum and minimax committee election rules for general preference types. In: Proceedings of the 22nd European Conference on Artificial Intelligence, pp 1656–1657

Baumeister D, Dennisen S (2015) Voter dissatisfaction in committee elections. In: Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems, pp 1707–1708

Baumeister D, Dennisen S, Rey L (2015) Winner determination and manipulation in minisum and minimax committee elections. In: Proceedings of the 4th International Conference on Algorithmic Decision Theory, pp 469–485

Baumeister D, Erdélyi G, Hemaspaandra E, Hemaspaandra L, Rothe J (2010) Computational aspects of approval voting. In: Laslier J, Sanver R (eds) Handbook of approval voting. Springer, New York, pp 199–251

Betzler N, Slinko A, Uhlmann J (2013) On the computation of fully proportional representation. J Artif Intell Res 47:475–519

Boutilier C, Brafman R, Domshlak C, Hoos H, Poole D (2004) CP-nets: a tool for representing and reasoning with conditional ceteris paribus preference statements. J Artif Intell Res 21:135–191

Bredereck R, Faliszewski P, Niedermeier R, Skowron P, Talmon N (2015) Elections with few candidates: Prices, weights, and covering problems. In: Proceedings of the 4th International Conference on Algorithmic Decision Theory, pp 414–431

Brill M, Laslier J, Skowron P (2017) Multiwinner approval rules as apportionment methods. In: Proceedings of the 31st AAAI Conference on Artificial Intelligence, pp 414–420

Byrka J, Skowron P, Sornat K (2017) Proportional approval voting, harmonic k-median, and negative association. Technical Report arXiv:1704.02183 [cs.DS]

Byrka J, Sornat K (2014) PTAS for minimax approval voting. In: Proceedings of the 10th Conference on Web and Internet Economics, pp 203–217

Chamberlin B, Courant P (1983) Representative deliberations and representative decisions: proportional representation and the Borda rule. Am Polit Sci Rev 77(3):718–733

Cygan M, Fomin F, Kowalik Ł, Lokshtanov D, Marx D, Pilipczuk M, Pilipczuk M, Saurabh S (2015) Parameterized algorithm. Springer, New York

Cygan M, Kowalik L, Socala A, Sornat K (2017) Approximation and parameterized complexity of minimax approval voting. In: Proceedings of the 31st AAAI Conference on Artificial Intelligence, pp 459–465

Debord B (1992) An axiomatic characterization of Borda's k-choice function. Soc Choice Welfare 9(4):337–343

Debord B (1993) Prudent k-choice functions: Properties and algorithms. Math Soc Sci 26:63–77

Downey R, Fellows M (1999) Parameterized complexity. Springer-Verlag, New York

Dummett M (1984) Voting procedures. Oxford University Press, Oxford

Elkind E, Faliszewski P, Skowron P, Slinko A (2017) Properties of multiwinner voting rules. Soc Choice Welfare 48:599–632

Elkind E, Ismaili A (2015) OWA-based extensions of the Chamberlin-Courant rule. In: Proceedings of the 4th International Conference on Algorithmic Decision Theory, pp 486–502

Elkind E, Lackner M (2015) Structure in dichotomous preferences. In: Proceedings of the 24th International Joint Conference on Artificial Intelligence, pp 2019–2025

Faliszewski P, Sawicki J, Schaefer R, Smolka M (2017) Multiwinner voting in genetic algorithms. IEEE Intell Syst 32(1):40–48

Faliszewski P, Skowron P, Slinko A, Talmon N (2016) Committee scoring rules: Axiomatic classification and hierarchy. In: Proceedings of the 25th International Joint Conference on Artificial Intelligence, pp 250–256. For the full version, see the arXiv report arXiv:1802.06483 [cs.GT]

Faliszewski P, Skowron P, Slinko A, Talmon N (2016) Multiwinner analogues of the plurality rule: Axiomatic and algorithmic perspectives. In: Proceedings of the 30th AAAI Conference on Artificial Intelligence, pp 482–488

Faliszewski P, Skowron P, Slinko A, Talmon N (2017) Multiwinner rules on paths from k-Borda to Chamberlin–Courant. In: Proceedings of the 26th International Joint Conference on Artificial Intelligence, pp 192–198

Faliszewski P, Skowron P, Slinko A, Talmon N (2017) Multiwinner voting: a new challenge for social choice theory. In: Endriss U (ed.), Trends in Computational Social Choice, ch. 2. MIT Press, Elsevier

Faliszewski P, Slinko A, Stahl K, Talmon N (2016) Achieving fully proportional representation by clustering voters. In: Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems, pp 296–304

Feige U (1998) A threshold of ln n for approximating set cover. J ACM 45(4):634–652

Fishburn P (1981a) An analysis of simple voting systems for electing committees. SIAM J Appl Math 41(3):499–502

Fishburn P (1981b) Majority committees. J Econ Theory 25(2):255–268

Garey M, Johnson D (1979) Computers and intractability: a guide to the theory of NP-completeness. W. H. Freeman and Company, New York

Khuller S, Saha B (2009) On finding dense subgraphs. In: Proceedings of the 36th International Colloquium on Automata, Languages, and Programming, pp 597–608

Kilgour M (2010) Approval balloting for multi-winner elections. In: Handbook on approval voting, Ch. 6. Springer, New York

Kleinberg J, Papadimitriou C, Raghavan P (2004) Segmentation problems. J ACM 51(2):263–280

Lackner M, Skowron P (April 2017) Consistent approval-based multi-winner rules. Technical Report arXiv:1704.02453 [cs.GT]

Lang J, Xia L (2015) Voting in combinatorial domains. In: Brandt F, Conitzer V, Endriss U, Lang J, Procaccia AD (eds) Handbook of computational social choice, ch. 9. Cambridge University Press, Cambridge

Laslier J (2012) And the loser is... plurality voting. In: Felsenthal D, Machover M (eds) Electoral systems: paradoxes, assumptions, and procedures. Springer, New York, pp 327–351

LeGrand R (2004) Analysis of the minimax procedure. Technical Report WUCSE-2004-67, Department of Computer Science and Engineering, Washington University

Lenstra H Jr (1983) Integer programming with a fixed number of variables. Math Oper Res 8(4):538–548

Lijphart A, Aitkin D (1994) Electoral systems and party systems: a study of twenty-seven democracies, 1945–1990. Oxford University Press, Oxford

Lu T, Boutilier C (2011) Budgeted social choice: From consensus to personalized decision making. In: Proceedings of the 22nd International Joint Conference on Artificial Intelligence, pp 280–286

Lu T, Boutilier C (2015) Value-directed compression of large-scale assignment problems. In: Proceedings of the 29th AAAI Conference on Artificial Intelligence, pp 1182–1190

Misra N, Nabeel A, Singh H (2015) On the parameterized complexity of minimax approval voting. In: Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems, pp 97–105

Monroe B (1995) Fully proportional representation. Am Polit Sci Rev 89(4):925–940

Nemhauser G, Wolsey L, Fisher M (1978) An analysis of approximations for maximizing submodular set functions. Math Program 14(1):265–294

Niedermeier R (2006) Invitation to fixed-parameter algorithms. Oxford University Press, Oxford

Oren J, Lucier B (2014) Online (budgeted) social choice. In: Proceedings of the 28th AAAI Conference on Artificial Intelligence, pp 1456–1462

Peters D (2018) Single-peakedness and total unimodularity: new polynomial-time algorithms for multi-winner elections. In: Proceedings of the 32nd AAAI Conference on Artificial Intelligence (**to appear**)

Peters D, Lackner M (2017) Preferences single-peaked on a circle. In: Proceedings of the 31st AAAI conference on artificial intelligence, pp 649–655

Procaccia A, Rosenschein J, Zohar A (2008) On the complexity of achieving proportional representation. Soc Choice Welfare 30(3):353–362

Raghavendra P, Steurer D (2010) Graph expansion and the unique games conjecture. In: Proceedings of the 42nd ACM Symposium on Theory of Computing, pp 755–764

Sánchez-Fernández L, Elkind E, Lackner M, Fernández N, Fisteus JA, Val P. Basanta, Skowron P (2017) Proportional justified representation. In: Proceedings of the 31st AAAI Conference on Artificial Intelligence, pp 670–676

Sawicki J, Smolka M, Łoś M, Schaefer R, Faliszewski P (2017) Two-phase strategy managing insensitivity in global optimization. In: Proceedings of the 20th International Conference on the Applications of Evolutionary Computation, pp 266–281

Skowron P (2016) FPT approximation schemes for maximizing submodular functions. In: Proceedings of the 12th Conference on Web and Internet Economics, pp 324–338

Skowron P, Faliszewski P (2017) Chamberlin-courant rule with approval ballots: approximating the max-cover problem with bounded frequencies in FPT time. J Artif Intell Res 60:687–716

Skowron P, Faliszewski P, Lang J (2016) Finding a collective set of items: from proportional multirepresentation to group recommendation. Artif Intell 241:191–216

Skowron P, Faliszewski P, Slinko A (2015a) Achieving fully proportional representation: approximability result. Artif Intell 222:67–103

Skowron P, Faliszewski P, Slinko A (April 2016) Axiomatic characterization of committee scoring rules. Technical Report arXiv:1604.01529 [cs.GT]

Skowron P, Yu L, Faliszewski P, Elkind E (2015b) The complexity of fully proportional representation for single-crossing electorates. Theor Comput Sci 569:43–57

Thiele T (1895) Om flerfoldsvalg. In: Oversigt over det Kongelige Danske Videnskabernes Selskabs Forhandlinger, pp 415–441

Yager R (1988) On ordered weighted averaging aggregation operators in multicriteria decisionmaking. IEEE Trans Syst Man Cybern 18(1):183–190

Yu L, Chan H, Elkind E (2013) Multiwinner elections under preferences that are single-peaked on a tree. In: Proceedings of the 23rd International Joint Conference on Artificial Intelligence, pp 425–431

Zanjirani F, Hekmatfar M (eds) (2009) Facility location: concepts, models, and case studies. Springer, New York