

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Artificial Intelligence

www.elsevier.com/locate/artintThe metric distortion of multiwinner voting [☆]Ioannis Caragiannis ^a, Nisarg Shah ^b, Alexandros A. Voudouris ^{c,*}^a Department of Computer Science, Aarhus University, Denmark^b Department of Computer Science, University of Toronto, Canada^c School of Computer Science and Electronic Engineering, University of Essex, United Kingdom

ARTICLE INFO

Article history:

Received 1 February 2022

Received in revised form 23 September 2022

Accepted 29 September 2022

Available online 3 October 2022

Keywords:

Distortion

Multiwinner voting

Social choice

ABSTRACT

We extend the recently introduced framework of metric distortion to multiwinner voting. In this framework, n agents and m alternatives are located in an underlying metric space. The exact distances between agents and alternatives are unknown. Instead, each agent provides a ranking of the alternatives, ordered from the closest to the farthest. Typically, the goal is to select a single alternative that approximately minimizes the total distance from the agents, and the worst-case approximation ratio is termed distortion. In the case of multiwinner voting, the goal is to select a committee of k alternatives that (approximately) minimizes the total cost to all agents. We consider the scenario where the cost of an agent for a committee is her distance from the q -th closest alternative in the committee. We reveal a surprising trichotomy on the distortion of multiwinner voting rules in terms of k and q : The distortion is unbounded when $q \leq k/3$, asymptotically linear in the number of agents when $k/3 < q \leq k/2$, and constant when $q > k/2$.

© 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

The most canonical problem in voting theory is to aggregate ranked preferences of n individual agents over a set of m alternatives to reach a collective decision. Examples of such decisions include selecting a single alternative (single-winner voting), selecting k out of m alternatives for a fixed $k > 1$ (multiwinner voting), and selecting a set of costly alternatives subject to a budget constraint (participatory budgeting). In centuries of research on voting, perhaps the most prominent approach to designing voting rules has been the axiomatic approach, in which one fixes several qualitative axioms and seeks voting rules satisfying them. Unfortunately, this approach has often led to impossibility results [17,34].

Procaccia and Rosenschein [32] proposed the (utilitarian) *distortion* framework for analyzing, comparing, and designing single-winner voting rules. Under this framework, the ordinal preferences expressed by agents are viewed as proxies for their underlying cardinal utilities, and the goal of a voting rule is to optimize the worst-case approximation ratio (distortion) of the social welfare (the total utility of the agents). Under minimal assumptions, this framework offers a *quantitative* comparison of voting rules. It has been used successfully to analyze the distortion of well-known methods [8] and to identify voting rules with optimal distortion [6]. The framework has also been extended to multiwinner voting [9] and participatory budgeting [5], under the assumption that the utility of an agent for a set of alternatives is the maximum and the sum of her utilities for the alternatives in the set, respectively.

[☆] A preliminary version of the paper appeared in the 36th AAAI Conference on Artificial Intelligence (AAAI 2022).

* Corresponding author.

E-mail address: alexandros.voudouris@essex.ac.uk (A.A. Voudouris).

Anshelevich et al. [2] built on this idea to propose the *metric distortion* framework, in which agents and alternatives are embedded in an underlying metric space, and the *cost* of an agent for an alternative is the distance between them. An agent still ranks the alternatives, but now in a non-decreasing order of their distance from her. Instead of maximizing the social welfare, the goal is now to minimize the social cost (the total cost of the agents). Like in the utilitarian case, scholars have analyzed the metric distortion of prominent voting rules [19,24,29,35] and have identified rules with optimal metric distortion [18]. For a detailed overview, we refer the reader to the survey by Anshelevich et al. [3].

The goal of our work is to extend the metric distortion framework to multiwinner voting, where the objective is to select a subset of k alternatives (committee) for a given $k > 1$. To the best of our knowledge, the only prior works to address multiwinner voting with general metric costs are those of Goel et al. [20] and Chen et al. [12]. Both focus on a model in which an agent's cost for a committee is the sum of her distances to the alternatives in the committee. In a sense, this assumes that an agent cares equally about all the alternatives chosen in the committee.

However, in many applications, an agent may care only about a few alternatives in the committee, typically the ones she prefers more. For example, when parliament members are chosen in a political election, each voter may associate with just one (or a few) of the elected candidates as her representative(s). Similarly, if a city builds parks at multiple locations, each resident may only be able to access a few parks closest to her. Motivated by these applications, we consider the case where an agent's cost for a committee of k alternatives is her distance to the q -th closest alternative in the committee, for a given $q \leq k$. Note that for $q = 1$, this is the distance to the closest alternative in the committee, whereas for $q = k$, it is the distance to the farthest one. The application under consideration may inform a natural choice of q . For example, when electing parliament members who would subsequently vote on issues using majority voting, a voter may naturally wish to have a majority of the parliament members close to her, which corresponds to $q = \lfloor k/2 \rfloor + 1$. Similarly, supermajority voting can correspond to other values of $q > k/2$. We remark that our results are the most positive (constant distortion) when $q > k/2$. Our main research question is:

What is the optimal distortion for selecting a committee of k alternatives under this cost model? How does it depend on the relation between q and k ? Can the optimal distortion be achieved via computationally efficient voting rules?

Before proceeding further, note that another reason why it may not be desirable to model an agent's cost for a committee as the *sum* of her distances to the alternatives in the committee is that the optimal committee can suffer from the tyranny of the majority; that is, it may consist entirely of the alternatives liked by the majority and include none liked by a minority. This is somewhat reflected by the fact that repeatedly applying a single-winner voting rule is known to work well in this model [20]. Due to this, prior work on multiwinner voting, both in the distortion literature [9] and elsewhere [10,25,27,33], aims to ensure that there is *at least one* alternative in the committee that every agent likes well (which corresponds to $q = 1$ in our model). An interesting byproduct of our cost model is that an agent's submitted ranking of individual alternatives can be used to determine her ordinal preferences over committees; we use this fact to derive some of our results.

1.1. Our contributions

Recall that n is the number of agents and m is the number of alternatives. We reveal a surprising trichotomy on the best possible distortion for multiwinner voting.

When $q \leq k/3$, the distortion is unbounded. This holds even for randomized voting rules, and even when n and m are only linear in k . When $q \in (k/3, k/2]$, the best possible distortion is $\Theta(n)$. Here, the upper bound is obtained via a novel voting rule that is deterministic and (computationally) efficient, while the lower bound holds even for randomized voting rules and when m is linear in k or q . Finally, when $q > k/2$, the best possible distortion is 3 for deterministic rules, and between 2.1126 and $3 - 2/n$ for randomized rules. For this case, we show that the costs of agents for committees satisfy the triangle inequality. Hence, we can reduce multiwinner voting to single-winner voting by treating each committee as a separate alternative and applying any single-winner voting rule, allowing us to borrow known best possible distortion bounds from the single-winner case to the multiwinner one. This is where we use the fact that under our cost model, we can deduce an agent's ordinal preferences over committees from her provided ranking of individual alternatives.

However, this reduction is not computationally efficient as we need to apply the single-winner voting rule on an instance with $\binom{m}{k}$ alternatives. To that end, we also provide an efficient reduction. We show that there exists an agent such that the committee consisting of the k alternatives she prefers the most has social cost no worse than 3 times the optimal. Thus, we can apply any single-winner voting rule with distortion ρ on a reduced instance with only n committees (one committee per agent) as alternatives and obtain a multiwinner rule with distortion at most 3ρ . In particular, by applying the PLURALITYMATCHING rule of Gkatzelis et al. [18], which is known to achieve the best possible distortion of 3 in the single-winner case, we obtain distortion at most 9 in polynomial time. We also show that an efficient reduction of this type cannot be used to achieve distortion better than 5.207. An overview of our results is given in Table 1.

1.2. Related work

The works of Goel et al. [20] and Chen et al. [12] are the most closely related to ours as they consider metric distortion for multiwinner voting. As mentioned before, both focus on a setting where the cost of an agent for a committee is the

Table 1
An overview of our distortion bounds.

	Deterministic	Randomized
$q \leq k/3$	$+\infty$	$+\infty$
$q \in (k/3, k/2]$	$\Theta(n)$	$\Theta(n)$
$q > k/2$	3 (exp time, general n) 3 (poly time, constant n) ≤ 9 (poly time, general n)	[2.1126, 3)

sum of her distances to the alternatives in the committee.¹ Goel et al. [20] show that selecting a committee by repeatedly applying any single-winner voting rule with distortion δ yields a distortion of at most δ . Since the best possible single-winner distortion is known to be 3 [18], this implies that a distortion of 3 can be achieved for any k in this model. Chen et al. [12] directly present a voting rule achieving distortion 3 for the case of $k = m - 1$ and prove this to be best possible.

Jaworski and Skowron [22] consider a model where voters (agents) and candidates (alternatives) have preferences over a set of *binary issues*, which induce the preferences of the voters over the candidates; specifically, voters rank candidates based on the number of issues they agree on. The elected committee uses majority voting to decide on each issue, and the cost of a voter is the number of issues for which the decision differs from her preferred outcome. This can be viewed as a metric distortion model, but with the specific Hamming distance metric. Also, the cost of a voter for a committee depends on the locations of the candidates in the committee, and not just on their distances to her. Meir et al. [26] consider a similar model where voters are also candidates, and show that sortition – picking k of the voters uniformly at random – leads to low distortion.

The distortion of randomized rules has also received significant attention. For single-winner voting, the best possible distortion under the utilitarian model is known to be $\Theta(\sqrt{m})$ [6,13], while that under the metric model is between 2.1126 and $3 - 2/\min\{n, m\}$ and still remains a challenging open question [1,11,23]. Caragiannis et al. [9] provide bounds on the best possible distortion of randomized multiwinner voting rules under the utilitarian model, and our work does so under the metric model.

More broadly, there is a huge literature on multiwinner voting that focuses on desiderata other than distortion, such as proportional representation [4,30], committee diversity [7], monotonicity or consistency axioms [14], explainability [31], etc. We refer the interested reader to the book chapter by Faliszewski et al. [15] for an overview.

2. Preliminaries

For $p \in \mathbb{N}$, define $[p] = \{1, \dots, p\}$. An instance of our problem is given by the tuple $I = (N, A, d, k, q)$, where:

- N is a set of $n \geq 2$ agents.
- A is a set of $m \geq 2$ alternatives.
- d is a pseudometric over $N \cup A$ with $d(x, y)$ denoting the *distance* between $x, y \in N \cup A$. Being a pseudometric, d satisfies, for all $x, y, z \in N \cup A$, $d(x, x) = 0$, $d(x, y) = d(y, x)$, and the *triangle inequality* $d(x, y) \leq d(x, z) + d(z, y)$. Since our framework only uses distances between agents and alternatives (and not between two agents or between two alternatives), we use the following equivalent formulation of the triangle inequality [19]: $d(i, x) \leq d(i, y) + d(j, y) + d(j, x)$ for all agents $i, j \in N$ and alternatives $x, y \in A$.²
- k and q are positive integers such that $1 \leq q \leq k < m$.

Every agent $i \in N$ ranks the alternatives based on her distances from them, from smallest (most preferable) to largest (least preferable), breaking ties arbitrarily; that is, the pseudometric d induces the ordinal preferences of agent i given by a *ranking* \succ_i over the alternatives such that $x \succ_i y$ implies $d(i, x) \leq d(i, y)$. We refer to $\succ_d = (\succ_i)_{i \in N}$ as the *preference profile*.

For any $S \subseteq A$ with $|S| \geq q$, we denote by $\text{top}_{i,q}(S)$ the set of q most preferred alternatives of agent i in S . A *committee* $C \subseteq A$ is a set of alternatives of size exactly equal to k . We define the q -cost of agent i for C , denoted $c_{i,q}(C|d)$, to be the distance of i from her q -th closest alternative in C : $c_{i,q}(C|d) = \max_{x \in \text{top}_{i,q}(C)} d(i, x)$. The q -social cost of committee C , denoted $\text{SC}_q(C|d)$, is then defined as the total q -cost of the agents for C : $\text{SC}_q(C|d) = \sum_{i \in N} c_{i,q}(C|d)$.

A (randomized) *multiwinner voting rule* f takes as input a preference profile \succ_d and outputs a distribution over committees; we say that f is deterministic if the distribution it returns has singleton support. Given $k \in \mathbb{N}$ and $q \in [k]$, the (k, q) -distortion of f is the worst-case ratio, over all possible instances with these parameters, between the (expected) q -social cost of the committee chosen by f and the minimum possible q -social cost of any committee, i.e.,

$$\sup_{I=(N, A, d, k, q): |A|=m > k} \frac{\mathbb{E}[\text{SC}_q(f(\succ_d)|d)]}{\min_{C \subseteq A: |C|=k} \text{SC}_q(C|d)}.$$

¹ Chen et al. [12] focus only on the case of $k = m - 1$, but also consider another cost model.

² When proving lower bounds, we will often design a worst-case pseudometric d by embedding agents and alternatives in the 1D Euclidean space and taking the Euclidean distance between them.

Table 2

An example of the preference profile used in the proof of Theorem 1, when $k = 8$ and $q = 2$. Here, we have $L = \lfloor 8/2 \rfloor + 1 = 5$, $n = L = 5$, and $m = Lq = 10$. The alternatives are partitioned into 5 sets X_1, X_2, Y_1, Y_2, Y_3 of size 2 each. We also have that $X = \{X_1, X_2\}$ and $Y = \{Y_1, Y_2, Y_3\}$. As an example, u_2 ranks the alternatives of Y higher than those of X , i.e., $\{Y_1, Y_2, Y_3\} \succ \{X_1, X_2\}$. Within Y , the sets therein are ranked based on their index distance from 2, e.g., $Y_2 \succ Y_1 \succ Y_3$. Within X , the order is fixed so that $X_2 \succ X_1$. The two alternatives in every subset are ranked arbitrarily.

agent	ranking
v_1	$X_1 \succ X_2 \succ Y_1 \succ Y_2 \succ Y_3$
v_2	$X_2 \succ X_1 \succ Y_1 \succ Y_2 \succ Y_3$
u_1	$Y_1 \succ Y_2 \succ Y_3 \succ X_2 \succ X_1$
u_2	$Y_2 \succ Y_1 \succ Y_3 \succ X_2 \succ X_1$
u_3	$Y_3 \succ Y_2 \succ Y_1 \succ X_2 \succ X_1$

Our goal is to design multiwinner rules with as low (k, q) -distortion as possible.

To simplify notation, we drop q and d whenever they are clear from the context. In particular, we will use $c_i(C)$ instead of $c_{i,q}(C|d)$, $SC(C)$ instead of $SC_q(C|d)$, and $\text{top}_i(S)$ instead of $\text{top}_{i,q}(S)$. We also denote by $T_i = \text{top}_i(A)$ the set of q alternatives ranked highest by agent i .

3. Unbounded distortion with $q \leq k/3$

We begin with a strong impossibility result for the case where $q \leq k/3$. In particular, we show that every multiwinner voting rule has unbounded distortion, even if it is allowed to use randomization.

Theorem 1. *For $q \leq k/3$, the (k, q) -distortion of every (even randomized) multiwinner voting rule is unbounded.*

Proof. Fix k and q such that $q \leq k/3$, and a multiwinner voting rule f . Let $L = \lfloor k/q \rfloor + 1 \geq 4$. We consider an instance with $n = L$ agents, partitioned into two sets: $V = \{v_1, \dots, v_{\lfloor L/2 \rfloor}\}$ and $U = \{u_1, \dots, u_{\lfloor L/2 \rfloor}\}$. There are $m = Lq$ alternatives, partitioned into L sets $X_1, \dots, X_{\lfloor L/2 \rfloor}, Y_1, \dots, Y_{\lceil L/2 \rceil}$ of size q each.³ Let $X = \bigcup_{\ell=1}^{\lfloor L/2 \rfloor} X_\ell$ and $Y = \bigcup_{\ell=1}^{\lceil L/2 \rceil} Y_\ell$.

Consider any preference profile such that:

- Every agent in V ranks the alternatives in X higher than those in Y .
- Every agent in U ranks the alternatives in Y higher than those in X .
- For $\ell \in [\lfloor L/2 \rfloor]$, every agent ranks the alternatives in X_ℓ as well as those in Y_ℓ arbitrarily.
- For $\ell \in [\lfloor L/2 \rfloor]$, agent $v_\ell \in V$ ranks the alternatives in X_i higher than those in X_j whenever $|\ell - i| < |\ell - j|$, for $i, j \in [\lfloor L/2 \rfloor]$; otherwise, agent v_ℓ ranks the alternatives in X_i lower than those in X_j . Also, agent v_ℓ ranks the sets of Y in the order $Y_1, \dots, Y_{\lceil L/2 \rceil}$ from lowest to highest.
- For $\ell \in [\lceil L/2 \rceil]$, agent $u_\ell \in U$ ranks the alternatives in Y_i higher than those in Y_j whenever $|\ell - i| < |\ell - j|$, for $i, j \in [\lceil L/2 \rceil]$; otherwise, agent $u_\ell \in U$ ranks the alternatives in Y_i lower than those in Y_j . Also, agent u_ℓ ranks the sets of X in the order $X_{\lfloor L/2 \rfloor}, \dots, X_1$ from highest to lowest.

Table 2 presents an example preference profile for $k = 8$ and $q = 2$; Fig. 1 depicts the possible underlying metric spaces used in the two cases below for this example instance.

Because $m = Lq = (\lfloor k/q \rfloor + 1) \cdot q > k$, not all alternatives can be included in the committee. We distinguish between the following two cases.

Case 1: Some alternative in Y is not included in the committee with a positive probability.

Suppose that an alternative in Y_{ℓ^*} is not included in the committee with a positive probability, for some $\ell^* \in [\lceil L/2 \rceil]$. Consider the following one-dimensional Euclidean metric, which is consistent with the ordinal preferences of the agents defined above:

- All agents in V and all alternatives in X are located at 0.

³ We use an instance with $n \leq m$. Note, however, that the lower bound continues to hold even if one assumes $n \geq m$. We can simply create t copies of each agent, for an appropriately large $t \in \mathbb{N}$.

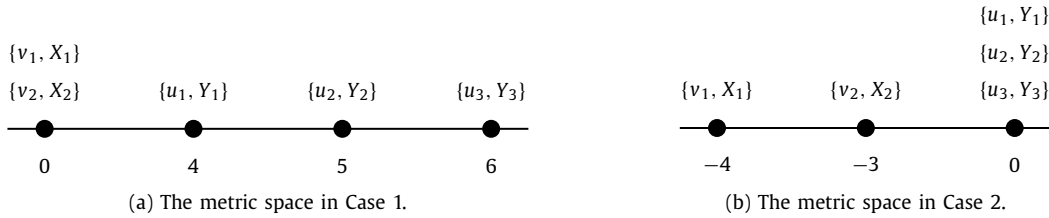


Fig. 1. The two metrics used in the proof of Theorem 1 when $k = 8$ and $q = 2$. Both are consistent with the ordinal preferences presented in Table 2. If the 6 alternatives of $Y = Y_1 \cup Y_2 \cup Y_3$ are not all included in the committee with positive probability, the expected q -cost of some agent of $U = \{u_1, u_2, u_3\}$ is positive in the first metric, thus leading to unbounded distortion, as the committee that includes the 6 alternatives of Y and 2 alternatives of $X = X_1 \cup X_2$ has q -social cost 0. If, on the other hand, all the alternatives of Y are included in the committee, not all 4 alternatives of X can be included in the committee, and thus some agent in $V = \{v_1, v_2\}$ has q -cost at least 1 in the second metric in any possible scenario; this again leads to unbounded distortion as the committee that includes the 4 alternatives of X and 4 alternatives of Y has q -social cost 0.

- For every $\ell \in [\lceil L/2 \rceil]$, agent $u_\ell \in U$ and the alternatives in Y_ℓ are located at $\lceil L/2 \rceil + \ell$.

Since some alternative in Y_{ℓ^*} is not included in the committee with a positive probability, the expected q -cost of agent u_{ℓ^*} is positive, and thus the expected q -social cost under f is also positive. However, it is possible to achieve q -social cost 0 by including in the committee all the alternatives of Y and an arbitrary subset of $k - \lceil L/2 \rceil q \geq q$ alternatives from X ; here, the inequality follows because $q \leq k/3$. So, the distortion of f is unbounded in this case.

Case 2: Every alternative in Y is included in the committee with probability 1.

Consider the following metric, which is consistent with the ordinal preferences of the agents defined above:

- For $\ell \in [\lceil L/2 \rceil]$, agent v_ℓ and the alternative in X_ℓ are located at $-L + \ell$.
- All agents of U and all alternatives in Y are located at 0.

Since f chooses a committee that includes all of Y with probability 1, it never includes all alternatives of X . Whenever an alternative in X_ℓ is excluded, note that agent v_ℓ has q -cost at least 1. Hence, the expected q -social cost is at least 1. However, it is possible to achieve q -social cost 0 by choosing the committee containing all the alternatives of X and an arbitrary subset of $k - \lceil L/2 \rceil q \geq q$ alternatives from Y . So, the distortion of f is unbounded in this case as well. \square

4. Linear distortion with $k/3 < q \leq k/2$

We now turn our attention to the case of $q \in (k/3, k/2]$. In this case, the best possible distortion turns out to be bounded, but linear in the number of agents, which could be very large.

We propose a novel deterministic multiwinner voting rule, called POLAROPPOSITES (see Algorithm 1), which runs in polynomial time and achieves a distortion of $O(n)$. Recall that, for a fixed value of q , $\text{top}_i(S)$ is the set of the q most favorite alternatives of agent i in S , $T_i = \text{top}_i(A)$, and also that we drop q and k from notation as they are clear from context. POLAROPPOSITES is relatively straightforward: we choose an agent i arbitrarily, and another agent j whose T_j has the highest cost⁴ for agent i ; then we output any committee that includes $T_i \cup T_j$. However, the analysis of the distortion upper bound of our rule is intricate.

Algorithm 1: PolarOpposites.

- 1 Choose an arbitrary agent $i \in N$;
 - 2 Choose an agent $j \in \arg \max_{\ell \in N \setminus \{i\}} c_i(T_\ell)$;
 - 3 Output an arbitrary committee $W \supseteq T_i \cup T_j$;
-

Before we proceed with bounding the distortion of our rule, we present a structural lemma, which holds for all possible values of k and q , and will be extremely useful in the proof of the bound.

Lemma 1. Consider any instance $I = (N, A, d, k, q)$ and let $O \in \arg \min_{C: |C|=k} SC(C)$ be an optimal committee. There exists a subset of agents S with $|S| \leq \lfloor k/q \rfloor$ such that for every agent $i \in N$ there exists an agent $j \in S$ with $\text{top}_i(O) \cap \text{top}_j(O) \neq \emptyset$ and $c_j(O) \leq c_i(O)$. In addition, for every agent $i \in N$ and committee $C \supseteq \bigcup_{j \in S} T_j$, we have $c_i(C) \leq 3 \cdot c_i(O)$.

⁴ Here, we use the fact that in our model, we can compare two sets of alternatives in terms of their cost to agent i using only agent i 's preference ranking over individual alternatives.

Proof. We construct the set S using Algorithm 2. Note that because we are only interested in arguing the existence of S , we can assume access to the underlying costs in order to construct S .

Algorithm 2: Construction of S in Lemma 1.

```

1 Sort the agents in  $N = [n]$  in non-decreasing order of their cost for  $O$  such that  $c_i(O) \leq c_j(O)$  for all  $i < j$ ;
2  $S \leftarrow \emptyset$ ;
3 for  $i = 1, \dots, n$  do
4   if  $\text{top}_i(O) \cap \text{top}_j(O) = \emptyset$  for all  $j \in S$  then
5      $S \leftarrow S \cup \{i\}$ ;
6   end
7 end

```

By construction, S is such that for every agent $i \in N$ either $i \in S$ (in which case the condition in the statement of the lemma holds for $j = i$), or there exists $j \in N \setminus \{i\}$ that is considered before i in Algorithm 2 (i.e., $c_j(O) \leq c_i(O)$) and $\text{top}_i(O) \cap \text{top}_j(O) \neq \emptyset$. Observe at any point during the algorithm, we maintain the invariant $\text{top}_i(O) \cap \text{top}_j(O) = \emptyset$ for all distinct $i, j \in S$. Since $|O| = k$ and $\text{top}_i(O)$ for each $i \in S$ consists of q unique alternatives from O , we always have $|S| \leq \lfloor k/q \rfloor$.

For the second claim, consider any committee $C \supseteq \bigcup_{i \in S} T_i$. Clearly, $c_i(C) \leq c_i(O)$ for every $i \in S$. By the property of S established above, for any agent $i \in N \setminus S$, there exist $j \in S$ and $x \in C$ such that $c_j(O) \leq c_i(O)$ and $x \in \text{top}_i(O) \cap \text{top}_j(O)$. Let $y \in C$ be the q -th favorite alternative of i in T_j . We make the following simple observations:

- Since $x \in \text{top}_i(O)$, $d(i, x) \leq c_i(O)$.
- By the choice of j , and since $x \in \text{top}_j(O)$ and $y \in T_j$, $d(j, x) \leq c_j(O) \leq c_i(O)$ and $d(j, y) \leq c_j(O) \leq c_i(O)$.

By the triangle inequality and the above observations, we have

$$c_i(C) \leq d(i, y) \leq d(i, x) + d(j, x) + d(j, y) \leq 3 \cdot c_i(O).$$

This completes the proof. \square

We are now ready to bound the distortion of the POLAROPPOSITES rule.

Theorem 2. For any $q \in (k/3, k/2]$, the (k, q) -distortion of POLAROPPOSITES is $O(n)$.

Proof. Let $I = (N, A, d, k, q)$ be an instance with $q \in (k/3, k/2]$. Let i and j be the agents chosen by POLAROPPOSITES on I , $W \supseteq T_i \cup T_j$ be the committee returned by it, and $O \in \arg \min_{C: |C|=k} \text{SC}(C)$ be an optimal committee for I . We will show that $c_\ell(W) \leq c_\ell(O) + 4 \cdot \text{SC}(O)$ for every agent $\ell \in N$. Then, by summing over all agents, we will obtain that $\text{SC}(W) \leq (4n + 1) \cdot \text{SC}(O)$, thus implying an upper bound of $4n + 1$ on the distortion of POLAROPPOSITES. We distinguish between the following two cases.

Case 1: $\text{top}_i(O) \cap \text{top}_j(O) \neq \emptyset$. Let $x \in \text{top}_i(O) \cap \text{top}_j(O)$ be an alternative that both i and j consider to be among their top q alternatives in the optimal committee O . For any agent $\ell \in N$, let

- $y_\ell = \arg \max_{z \in T_\ell} d(\ell, z)$ be the q -th overall favorite alternative of ℓ , and
- $y_{\ell j} = \arg \max_{z \in T_j} d(\ell, z)$ be the q -th favorite alternative of ℓ among those in T_j .

Since $T_j \subseteq W$, by the definition of $y_{\ell j}$, we have $c_\ell(W) \leq c_\ell(T_j) = d(\ell, y_{\ell j})$. Combining this with the triangle inequality, we obtain

$$\begin{aligned} c_\ell(W) &\leq d(\ell, y_{\ell j}) \\ &\leq d(\ell, y_\ell) + d(i, y_\ell) + d(i, x) + d(j, x) + d(j, y_{\ell j}). \end{aligned}$$

For any agent v , note that $c_v(T_v) \leq c_v(O)$, and for any agent u and alternative $y \in T_v$, we also have $d(u, y) \leq c_u(T_v)$. Combining these with the definitions of j, y_ℓ, x , and $y_{\ell j}$, we can now bound each of the terms in the expression above as follows:

- $d(\ell, y_\ell) = c_\ell(T_\ell) \leq c_\ell(O)$;
- $d(i, y_\ell) \leq c_i(T_\ell) \leq c_i(T_j)$;
- $d(i, x) \leq c_i(O)$;
- $d(j, x) \leq c_j(O)$;

- $d(j, y_{\ell j}) \leq c_j(T_j) \leq c_j(O)$.

Substituting these, we get

$$c_\ell(W) \leq c_\ell(O) + c_i(T_j) + c_i(O) + 2c_j(O).$$

We can bound the term $c_i(T_j)$ using the definition of alternative y_{ij} , the triangle inequality, and some of the above observations, as follows:

$$\begin{aligned} c_i(T_j) &= d(i, y_{ij}) \\ &\leq d(i, x) + d(j, x) + d(j, y_{ij}) \\ &\leq c_i(O) + c_j(O) + c_j(T_j) \\ &\leq c_i(O) + 2c_j(O). \end{aligned}$$

So, combining everything, and using the fact that $SC(O) \geq c_i(O) + c_j(O)$, we have that

$$\begin{aligned} c_\ell(W) &\leq c_\ell(O) + 2c_i(O) + 4c_j(O) \\ &\leq c_\ell(O) + 4 \cdot SC(O), \end{aligned}$$

as desired.

Case 2: $\text{top}_i(O) \cap \text{top}_j(O) = \emptyset$. Consider the set S that is guaranteed to exist by Lemma 1. Since $q \in (k/3, k/2]$, we have that $\lfloor k/q \rfloor = 2$, and hence $|S| \leq 2$.

If $|S| = 1$, then there exists a single agent $v \in S$ and $\text{top}_\ell(O) \cap \text{top}_v(O) \neq \emptyset$ for every agent $\ell \in N$. If $|S| = 2$, we claim that there is a function $g : N \rightarrow S$ such that for every agent $\ell \in N$, it holds that $\text{top}_\ell(O) \cap \text{top}_{g(\ell)}(O) \neq \emptyset$ and $S = \{g(i), g(j)\}$. Lemma 1 already guarantees a function g meeting the first condition. The second condition follows since $|O| = k < 3q$ and $\text{top}_i(O) \cap \text{top}_j(O) = \emptyset$ imply that there cannot exist $v \in S$ such that $\text{top}_i(O) \cap \text{top}_v(O) = \text{top}_j(O) \cap \text{top}_v(O) = \emptyset$; otherwise $\text{top}_i(O)$, $\text{top}_j(O)$, and $\text{top}_v(O)$ would be disjoint subsets of O of size q each.

Now, consider any agent $\ell \in N$, and suppose that $g(\ell) = g(j) = v \in S$; the case $g(\ell) = g(i)$ if $|S| = 2$ is similar. By the properties of S , there exist alternatives $x \in \text{top}_\ell(O) \cap \text{top}_v(O)$ and $z \in \text{top}_j(O) \cap \text{top}_v(O)$. As in case 1, let $y_{\ell j} \in T_j$ be the q -th favorite alternative of ℓ in T_j ; so, $c_\ell(T_j) = d(\ell, y_{\ell j})$. By the fact that $T_j \subseteq W$, and using the triangle inequality, we obtain

$$\begin{aligned} c_\ell(W) &\leq c_\ell(T_j) = d(\ell, y_{\ell j}) \\ &\leq d(\ell, x) + d(v, x) + d(v, z) + d(j, z) + d(j, y_{\ell j}). \end{aligned}$$

By the definitions of x , z and $y_{\ell j}$, we have $d(\ell, x) \leq c_\ell(O)$, $d(v, x) \leq c_v(O)$, $d(v, z) \leq c_v(O)$, $d(j, z) \leq c_j(O)$, and $d(j, y_{\ell j}) \leq c_j(T_j) \leq c_j(O)$. Combined with the fact that $SC(O) \geq c_v(O)$ and $SC(O) \geq c_j(O)$, we get

$$\begin{aligned} c_\ell(W) &\leq c_\ell(O) + 2c_v(O) + 2c_j(O) \\ &\leq c_\ell(O) + 4 \cdot SC(O), \end{aligned}$$

as desired. \square

We conclude this section by showing a matching lower bound of $\Omega(n)$ on the distortion of any (even randomized) multiwinner voting rule. Hence, when $q \in (k/3, k/2]$, POLAROPPOSITES is the asymptotically best possible rule in terms of distortion, even among randomized rules.

Theorem 3. For $q \in (k/3, k/2]$, the (k, q) -distortion of every (even randomized) multiwinner voting rule is $\Omega(n)$.

Proof. Fix k and $q \in (k/3, k/2]$, and let f be an arbitrary multiwinner voting rule. We consider instances with $n = 2\lambda + 1$ agents, partitioned into two sets V and U of size λ each and a singleton $\{w\}$. There are $m = 3q$ alternatives, partitioned into three sets X , Y , and Z consisting of q alternatives each.

Consider any preference profile subject to the following rules.

- Every agent ranks the alternatives that belong to the same set arbitrarily.
- Every agent in V has the ranking $X \succ Y \succ Z$.
- Every agent in U has the ranking $Y \succ X \succ Z$.
- Agent w has the ranking $Z \succ Y \succ X$.

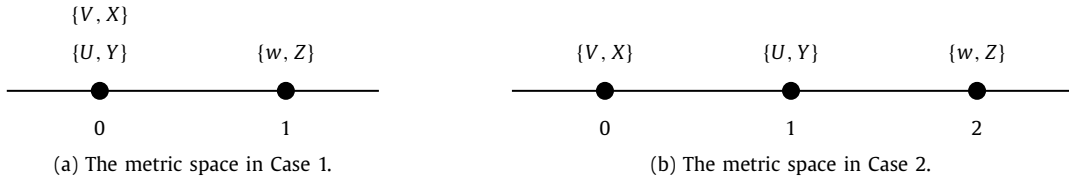


Fig. 2. The metric spaces used in the proof of Theorem 3. Each of the sets X, Y, Z has size q . Both metrics are consistent with the ordinal profile according to which the preference of the λ agents in V is $X > Y > Z$, the preference of all λ agents in U is $Y > X > Z$, and the preference of agent w is $Z > Y > X$. If the q alternatives of Z are not all included in the committee with positive probability (Case 1), then agent w has positive expected q -cost in the first metric space, leading to unbounded distortion as the committee that includes Z and q alternatives from $X \cup Y$ has q -social cost 0. If all alternatives of Z are included in the committee (Case 2), then the λ agents of V (in case not all alternatives of X are included in the committee) or the λ agents in U (in case not all alternatives of Y are included in the committee) have q -cost 1 in the second metric space, leading to a distortion of $\lambda = \Omega(n)$ as the committee that includes all the alternatives of $X \cup Y$ and some alternatives of Z has q -social cost 1. In either case, the distortion of the voting rule is $\Omega(n)$.

Since $m = 3q > k$, the committee returned by f cannot include every alternative with probability 1. We distinguish between the following two cases. Fig. 2 depicts the two metric spaces considered in these cases.

Case 1: With a positive probability, the committee chosen by f does not include all q alternatives of Z .

Consider the following metric, which is consistent (up to tie-breaking) with the preference profile defined above:

- The agents in $V \cup U$ and the alternatives in $X \cup Y$ are all located at 0.
- Agent w and the alternatives of Z are located at 1.

Since some alternative in Z is not included in the chosen committee with a positive probability, the expected q -cost of agent w , and thus the expected q -social cost under f , is positive. However, the committee that includes all of Z and any q alternatives from $X \cup Y$ has q -social cost 0. Therefore, the distortion of f is unbounded.

Case 2: With probability 1, the committee chosen by f includes every alternative of Z .

Consider the following metric, which is again consistent (up to tie-breaking) with the preference profile defined above:

- The agents in V and the alternatives in X are at 0.
- The agents in U and the alternatives in Y are at 1.
- Agent w and the alternatives in Z are at 2.

Recall that with probability 1, f chooses a committee that includes all q alternatives of Z . Any such committee cannot contain all alternatives in $X \cup Y$. If it does not include an alternative of X , the q -cost of every agent in V is at least 1, and if it does not include an alternative of Y , the q -cost of every agent in U is at least 1. Either way, we have that, with probability 1, f chooses a committee with q -social cost at least λ . Hence, the expected q -social cost under f is at least λ . However, the committee that includes all $2q$ alternatives of $X \cup Y$ and $k - 2q$ arbitrary alternatives of Z has q -social cost 1; only agent w has q -cost 1. So, the distortion of f is at least $\lambda = \Omega(n)$ in this case. \square

5. Constant distortion with $q > k/2$

We now turn our attention to the case $q > k/2$, where it is possible to achieve constant distortion. A crucial observation, which we exploit in this section, is that the q -costs of the agents form a new metric over the agents and all possible committees of alternatives.

Lemma 2. For any instance $I = (N, A, d, k, q)$ with $q > k/2$, the q -costs $c_i(C)$ of agents $i \in N$ for k -sized committees of alternatives $C \subseteq A$ form a pseudometric.

Proof. To prove the statement, we need to show that the q -costs satisfy the triangle inequality, i.e., $c_i(X) \leq c_i(Y) + c_j(Y) + c_j(X)$, for any two agents $i, j \in N$ and two k -sized committees X, Y .

For any $q \leq k$, there exists $x \in \text{top}_j(X)$ that is among the $k - q + 1$ least favorite alternatives (ranked in some position from q to k) of agent i in X ; thus, x is such that $c_i(X) \leq d(i, x)$ and $c_j(X) \geq d(j, x)$. Since $q > k/2$, there also exists $y \in \text{top}_i(Y) \cap \text{top}_j(Y)$, that is, y is such that $c_i(Y) \geq d(i, y)$ and $c_j(Y) \geq d(j, y)$. Combining these with the triangle inequality for the distances between agents and alternatives, we have that

$$\begin{aligned} c_i(X) &\leq d(i, x) \leq d(i, y) + d(j, y) + d(j, x) \\ &\leq c_i(Y) + c_j(Y) + c_j(X), \end{aligned}$$

as desired. \square

Since the q -costs form a pseudometric, we can transform the original profile in which the agents rank the alternatives to one in which the agents rank all possible k -sized committees. In particular, to decide whether agent i prefers a committee X over another committee Y , it suffices to compare her q -th favorite alternatives in X and Y ; we can break ties arbitrarily. Given the rankings of the agents over the committees, we can then employ any single-winner rule to decide the final committee. Specifically, using the PLURALITY-MATCHING rule of Gkatzelis et al. [18], we obtain a distortion bound of 3.⁵

Corollary 1. *For $q > k/2$, there exists a multiwinner voting rule with (k, q) -distortion at most 3.*

Unfortunately, the above approach requires us to apply a single-winner voting rule to a profile consisting of an exponential number of alternatives (the set of all k -sized committees). This is an inherent obstacle in our attempts to get constant distortion bounds by naively applying known deterministic single-winner voting rules.

Interestingly, it is still easy to compute the committee that includes the k most preferred alternatives of an agent; we refer to this as the *top- k* committee of the agent. Consequently, randomized dictatorship, the single-winner voting rule which selects an agent uniformly at random and returns her most favorite alternative, can be efficiently implemented in our setting; in particular, it chooses an agent uniformly at random and returns her top- k committee. Using its distortion analysis by Anshelevich and Postl [1], we obtain the following.

Corollary 2. *For any $q > k/2$, the (k, q) -distortion of randomized dictatorship, which can be implemented efficiently, is at most $3 - 2/n$.*

Proof. Let $I = (N, A, d, k, q)$ be an arbitrary instance with $q > k/2$, and denote by O an optimal k -sized committee. Let X_i be the top- k committee of agent $i \in N$, and let R be the random committee selected by randomized dictatorship. The expected q -social cost of R is

$$\begin{aligned} \mathbb{E}[\text{SC}(R)] &= \frac{1}{n} \sum_{i \in N} \text{SC}(X_i) = \frac{1}{n} \sum_{j \in N} \sum_{i \in N} c_i(X_j) = \frac{1}{n} \sum_{j \in N} \sum_{i \neq j} c_i(X_j) + \frac{1}{n} \sum_{j \in N} c_j(X_j) \\ &\leq \frac{1}{n} \sum_{j \in N} \sum_{i \neq j} \left(c_i(O) + c_j(O) + c_j(X_j) \right) + \frac{1}{n} \sum_{j \in N} c_j(X_j) \\ &= \left(2 - \frac{2}{n} \right) \sum_{j \in N} c_j(O) + \sum_{j \in N} c_j(X_j) \leq \left(3 - \frac{2}{n} \right) \cdot \text{SC}(O). \end{aligned}$$

The first inequality follows by applying the triangle inequality, and the second one follows since $c_j(X_j) \leq c_j(O)$ for every agent $j \in N$, and hence $\sum_{j \in N} c_j(X_j) \leq \text{SC}(O)$. \square

In the following, we restrict our attention to deterministic polynomial-time multiwinner voting rules. Our main result provides such a rule with distortion at most 9 by exploiting the following lemma.

Lemma 3. *Consider any instance $I = (N, A, d, k, q)$ with $q > k/2$, and let O be an optimal committee for I . There exists a committee C that is top- k for some agent, and such that $\text{SC}(C) \leq 3 \cdot \text{SC}(O)$.*

Proof. By Lemma 2, the q -costs of the agents form a metric space. Let j be the closest agent to the optimal committee O according to her q -cost, i.e., $c_j(O) \leq c_i(O)$ for every $i \in N$. Let C be the top- k committee of j , and thus $c_j(C) \leq c_j(O)$. Combining the above with the triangle inequality for the q -cost metric, for any agent $i \in N$, we have

$$\begin{aligned} c_i(C) &\leq c_i(O) + c_j(O) + c_j(C) \\ &\leq c_i(O) + 2c_j(O) \leq 3c_i(O). \end{aligned}$$

The lemma follows by summing over all agents. \square

Essentially, Lemma 3 suggests that the best top- k committee must 3-approximate the optimal committee in terms of social cost. So, by considering the n top- k committees (one per agent) as alternatives, we can deploy the single-winner rule of Gkatzelis et al. [18] to obtain in polynomial time a committee that is within a factor of 3 from the best top- k committee in terms of social cost, and thus has a (k, q) -distortion of at most 9.

⁵ Observe that the bound of 3 is best possible when $k < 2q \leq m$; to see this, consider an instance with $n/2$ agents that report $a_1 > \dots > a_q > b_1 > \dots > b_q$ (followed by any remaining alternatives) and $n/2$ agents that report $b_1 > \dots > b_q > a_1 > \dots > a_q$ (again followed by any remaining alternatives). However, it seems that we cannot define such an instance when $m < 2q$. It would thus be interesting to analyze the distortion as a function of q, k , and m in this regime.

Corollary 3. For any $q > k/2$, there is a polynomial-time multiwinner voting rule with (k, q) -distortion at most 9.

Actually, the approach used to obtain Corollary 3 can be used as a template that could potentially lead to even lower distortion bounds by deterministic polynomial-time voting rules. All we need is an algorithm that takes as input the ranking profile and decides – in polynomial time – on a set P of k -sized committees, which, for every distance metric consistent with the ranking profile, contains a committee that approximates the optimal social cost within a factor of ρ (e.g., Lemma 3 suggests that this is possible for $\rho = 3$). Then, applying the voting rule of Gkatzelis et al. [18] on the reduced ranking profile with only the committees in P as alternatives, we obtain a deterministic polynomial-time multiwinner voting rule with (k, q) -distortion at most 3ρ .

We present two results related to this template. The first one is positive and shows that a guarantee of $\rho = 1$ is possible when the number of agents is constant, making the optimal distortion bound of 3 achievable in polynomial time.

Theorem 4. For any $q > k/2$ and constant number of agents, there is a deterministic polynomial-time multiwinner voting rule with (k, q) -distortion at most 3.

Proof. We prove the theorem by defining a voting rule that follows our template with $\rho = 1$. In particular, given a ranking profile, our rule first identifies a set P of at most m^n committees such that one of minimum social cost is always included, no matter which distance function d , consistent with the preference profile, is used in the definition of the social cost. Then, by invoking the single-winner rule of Gkatzelis et al. [18] using the committees in P as alternatives, we get the desired multiwinner voting rule with (k, q) -distortion of at most 3.

To identify the committees in P , we enumerate over all possibilities for the alternatives in a k -sized committee that the agents can have as their q -th closest ones. In particular, for each of the m^n possible vectors of alternatives $\ell = \langle \ell_1, \ell_2, \dots, \ell_n \rangle$, we include in P a committee $C \subseteq A$ with $|C| = k$, $\bigcup_{i=1}^n \{\ell_i\} \subseteq C$, such that ℓ_i is the alternative in C that is the q -th closest to agent i , for $i \in [n]$, if such a committee exists. We refer to each committee having these properties as a (k, q) -completion of the vector of alternatives ℓ . Notice that including in P just one (k, q) -completion for each vector of alternatives is enough for our purposes since any two committees C, C' that are (k, q) -completions for the same vector ℓ of alternatives have the same social cost for a given distance function d : $SC_q(C|d) = SC_q(C'|d) = \sum_{i=1}^n d(i, \ell_i)$. Clearly, since n is a constant, P has polynomial size. To prove that its whole construction takes polynomial time, it suffices to show how a (k, q) -completion C for a given vector of alternatives ℓ can be identified (if it exists) in polynomial time. Essentially, we need to decide which alternatives in addition to the ones in ℓ form a k -sized committee and ensure that alternative ℓ_i is the q -th closest to agent i among those in C . To do so, we define the following classification of the alternatives into types from the set

$$\mathcal{T} = \{\langle t_1, t_2, \dots, t_n \rangle : t_i \in \{+1, 0, -1\}, i \in [n]\}.$$

An alternative $a \in A$ belongs to type $\mathbf{t} \in \mathcal{T}$ if and only if

- $a \succ_i \ell_i$ for each $i \in [n]$ with $t_i = +1$,
- $a = \ell_i$ for each $i \in [n]$ with $t_i = 0$, and
- $\ell_i \succ_i a$ for each $i \in [n]$ with $t_i = -1$.

Notice that the classification is such that replacing an alternative in a (k, q) -completion by another alternative from the same class results in another (k, q) -completion. Hence, to identify a (k, q) -completion C for a given vector of alternatives ℓ , we need to identify the number of alternatives from each type of \mathcal{T} that C should have.

For each type $\mathbf{t} \in \mathcal{T}$, denote by $H(\mathbf{t})$ the number of alternatives of type \mathbf{t} . Also, set $L(\mathbf{t}) = 1$ if \mathbf{t} is the type of some alternative in ℓ and $L(\mathbf{t}) = 0$, otherwise. The quantities $L(\mathbf{t})$ and $H(\mathbf{t})$ are lower and upper bounds on the number $h(\mathbf{t})$ of alternatives of type \mathbf{t} that can be included in a (k, q) -completion of ℓ . In particular, notice that each alternative in ℓ is the unique alternative in its type \mathbf{t} , i.e., $H(\mathbf{t}) = 1$. Setting $L(\mathbf{t}) = 1$ for this type guarantees that this alternative will be included in the (k, q) -completion. Now, the existence of a (k, q) -completion is equivalent to the feasibility of the following integer linear program:

$$\begin{aligned} \sum_{\mathbf{t} \in \mathcal{T}} h(\mathbf{t}) &= k \\ \sum_{\mathbf{t} \in \mathcal{T}: t_i \geq 0} h(\mathbf{t}) &= q, \forall i \in [n] \\ L(\mathbf{t}) &\leq h(\mathbf{t}) \leq H(\mathbf{t}), \forall \mathbf{t} \in \mathcal{T} \\ h(\mathbf{t}) &\in \mathbb{N}_{\geq 0}, \forall \mathbf{t} \in \mathcal{T} \end{aligned}$$

Notice that we can naively check the feasibility of the above ILP by enumerating all possible values for the variables $h(\mathbf{t})$. As there are 3^n types, there are 3^n such variables (i.e., constantly many since n is a constant), each taking at most $m + 1$

values. Even this naive solution takes only polynomial time. Finally, once we have a feasible solution for the above ILP, we can easily define the (k, q) -completion by just including $h(\mathbf{t})$ alternatives of type \mathbf{t} in it, for $\mathbf{t} \in \mathcal{T}$. \square

Our last result is negative and reveals a limitation of our template. It shows a lower bound of $1 + 2/e$ on ρ (unless $P = NP$) and, consequently, a lower bound of approximately 5.207 on the distortion that can be achieved via this template. In the appendix, we present a slightly weaker lower bound of 4.5 that is purely information-theoretic and does not rely on any complexity assumption.

Theorem 5. *Suppose there is an algorithm \mathcal{A} which when given as input a ranking profile with n agents and m alternatives, and integers k and q with $1 \leq k/2 < q \leq k < m$, runs in time polynomial in m and n and returns a set P of (polynomially many) k -sized committees of alternatives with the following property: For every distance function d , there is a committee $C \in P$ such that $SC_q(C|d) < (1 + 2/e - \varepsilon) \cdot SC_q(O|d)$, where O is a committee of minimum social cost according to d , and ε is a positive constant. Then, $P = NP$.*

Proof. We will prove the theorem using a reduction from the maximum K -cover problem. In maximum K -cover, we are given a universe U of elements, a collection \mathcal{S} of subsets of U , and an integer K . The objective is to find a subcollection of K sets from \mathcal{S} (or, a K -collection), whose union contains the maximum number of elements from U .

A well-known hardness result due to Feige [16] states that, for every constant $\varepsilon > 0$, it is NP-hard to distinguish between the next two cases for a maximum K -cover instance:

1. The instance has a K -collection that covers all elements of U .
2. Any K -collection covers at most a $(1 - 1/e + \varepsilon/2)$ -fraction of the elements of U .

Our reduction works as follows: Given an instance of maximum K -cover, we build a ranking profile with $n = |U|$ agents and $m = q - 1 + |\mathcal{S}|$ alternatives. Each agent corresponds to an element of U . There are $q - 1$ special alternatives and, additionally, a distinct set-alternative for every set $S \in \mathcal{S}$. For $i \in U$, agent i ranks the special alternatives first (in arbitrary order), followed by the set-alternatives that include element i , and then the remaining set-alternatives. The distance function d is such that agent i has distance 1 from the special alternatives and the set-alternatives which contain element i , and distance 3 from the remaining set-alternatives. The committee size is $k = q - 1 + K$.

Let us call *canonical* a k -sized committee that includes all special alternatives. Notice that a committee can be transformed to a canonical one by including the missing special alternatives and removing (arbitrarily) an equal number of set-alternatives. This transformation guarantees that the social cost (according to d) of the canonical committee is not worse than that of the original one.

We now claim that if algorithm \mathcal{A} satisfies the condition in the statement of Theorem 5, then we can distinguish between the two cases above in polynomial time, contradicting the hardness statement of Feige [16]. Indeed, consider the algorithm which, given a ranking profile produced by our reduction as input, runs algorithm \mathcal{A} first to compute a set P of (polynomially many) k -sized committees, then transforms these committees to canonical ones, and, among the K -collections corresponding to these canonical committees (each formed by the sets that correspond to the set-alternatives), returns the one that covers the maximum number of elements.

We now argue that by comparing the number of elements covered by the outcome of the algorithm to $(1 - 1/e + \varepsilon/2)n$, we can distinguish whether the original instance of maximum K -cover is of case 1 or case 2 in polynomial time. If the maximum K -cover instance is of case 2, then any K -collection returned by the algorithm clearly covers at most a $(1 - 1/e + \varepsilon/2)$ -fraction of the elements. If it is of case 1, let \widehat{O} be a K -collection that covers all elements of U . Then, the canonical committee O that contains the special alternatives and the set-alternatives corresponding to the sets of \widehat{O} has a minimum possible social cost of $SC_q(O|d) = n$. Indeed, since some of the sets of \widehat{O} contain element i , agent i has her q -th favorite alternative in O at distance 1; recall that she has the special alternatives in the top $q - 1$ positions of her preference, so the q -th closest alternatives is her top set-alternative in O . Now, by the approximation guarantee of algorithm \mathcal{A} , we know that there is a canonical committee among those obtained by the committees of P that has $SC_q(C|d) < (1 + 2/e - \varepsilon)n$. This means that the number of agents whose q -th closest alternative in C is at distance 1 must be more than $(1 - 1/e + \varepsilon/2)n$; the rest are at distance 3. Equivalently, by the definition of our reduction, the K -collection that corresponds to C covers more than $(1 - 1/e + \varepsilon/2)n$ elements of U . \square

6. Extensions and open problems

In this work, we extended the metric distortion framework to multiwinner voting. By modeling the cost of an agent for a k -sized committee as her distance from her q -th favorite alternative therein, we revealed a quite surprising trichotomy on the distortion of multiwinner rules in terms of q and k , providing asymptotically tight bounds. The main question that our work leaves open is to identify the best possible distortion that can be achieved by a computationally efficient deterministic rule in the case $q > k/2$. To achieve distortion at most 9 in this case, our voting rule applies the single-winner rule of Gkatzelis et al. [18] on the set of top- k committees as alternatives. This rule is just a member of a more general class of

rules that construct committees by exploiting single-winner voting. For example, following the work of Goel et al. [20], one could consider another rule of this class which constructs a committee iteratively by adding each time the alternative returned by the rule of Gkatzelis et al. [18] (or some other single-winner rule) when given as input the rankings of the agents for the alternatives not yet included in the committee. Is there any value of q large enough (such as $q = k$) for which this rule achieves constant distortion better than 9? It would be interesting to further explore the potential of such rules.

We exclusively focused on the social cost and did not consider other objectives, such as the maximum agent cost. It is not hard to observe that our methods provide bounds in terms of this objective as well, albeit not necessarily tight. When $q \leq k/3$, the distortion remains unbounded since, in the instances used in the proof of Theorem 1, the optimal committee guarantees cost 0 to all agents, whereas any voting rule yields a positive cost to some agent. When $q \in (k/3, k/2)$, by carefully inspecting the proof of the upper bound of POLAROPPOSITES in Theorem 2, we can show that it achieves constant distortion in terms of the maximum cost. Finally, when $q > k/2$, since the q -costs form a metric space, known results from single-winner voting extend to multiwinner voting. That is, there exists a deterministic multiwinner rule with distortion at most 3, which is based on the rule of Gkatzelis et al. [18]; this rule is known to achieve a *fairness ratio* of at most 3, which implies 3-approximation of the maximum cost as well. The efficient upper bound of 9 via our template also extends to the maximum cost, as the bound of 3 in Lemma 3 actually holds per agent. In the future, it would be interesting to go beyond objectives such as the social and maximum costs, and consider others that make sense in multiwinner voting.

Another interesting direction for future work is to consider multiwinner voting settings with different individual cost definitions. While in this paper we focused on the case where the cost of an agent for a committee is determined by the distance from her q -th favorite alternative in the committee, there are many different definitions that also make sense and capture different behaviors. For example, the individual cost can be defined as the total distance between an agent and her top- q alternatives in the committee. Both this and our q -cost definition are special cases of the broader family of ordered weighted average (OWA) costs, in which the cost to an agent is a weight w_q multiplied by her distance to her q -th favorite alternative in the committee, summed over all $q \in [k]$. Studying the distortion under this family, for different choices of $\mathbf{w} = (w_1, \dots, w_k)$, deserves future investigation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgements

We would like to thank Elliot Anshelevich for fruitful discussions in early stages of this work, and Piotr Skowron for suggesting an example demonstrating unbounded distortion for $q = 1$, which has been generalized to $q \leq k/3$ in Theorem 1. We also thank the anonymous reviewers for their comments and suggestions.

Appendix A. A weaker information-theoretic lower bound

Theorem 6. Let $\varepsilon > 0$ be a constant, $m \geq 2$ an even integer, and $q = k = m/2$. There exist a set A of m alternatives, a set N of m agent with a preference profile $(\succ_i)_{i \in N}$ over the alternatives of A , a distance function d that is consistent with the preference profile \succ , and a k -sized committee O , such that for any set P of fewer than $\exp(\varepsilon^2 m)$ k -sized committees of alternatives, $SC_q(C) \geq \left(\frac{3}{2} - \varepsilon\right) SC_q(O)$ for every committee $C \in P$.

Proof. Let $A = \{a_1, \dots, a_m\}$, $N = \{1, 2, \dots, m\}$, and consider the following preference profile $(\succ_i)_{i \in N}$: For $i = 1, 2, \dots, m/2$, both agents $2i - 1$ and $2i$ have alternatives a_{2i-1} and a_{2i} at the two bottom positions of their rankings, while the remaining alternatives are ranked arbitrarily in the remaining $m - 2$ positions. Agent $2i - 1$ has alternative a_{2i-1} last (i.e., $a_{2i} \succ_{2i-1} a_{2i-1}$) while agent $2i$ has alternative a_{2i} last (i.e., $a_{2i-1} \succ_{2i} a_{2i}$).

Our proof uses the probabilistic method (e.g., see [28]). We will define a random distance function d that is consistent with the profile above, as well as a random committee O of $k = m/2$ alternatives, so that by defining the social cost SC_q using d , we have $SC_q(O) = m$ and, with strictly positive probability, $SC_q(C) > \left(\frac{3}{2} - \varepsilon\right)m$ for each committee $C \in P$.

This probabilistic argument guarantees that there exists a distance function such that $SC_q(C) > \left(\frac{3}{2} - \varepsilon\right) SC_q(O)$, for every committee $C \in P$.

Initially, committee O is empty. For $i = 1, 2, \dots, m/2$, toss a fair coin, independently for different i 's:

- On heads, set $d(2i - 1, a_{2i-1}) = 3$, $d(2i - 1, a_j) = 1$ for $j \neq 2i - 1$, and $d(2i, a_j) = 1$ for $j = 1, \dots, m$. Include alternative a_{2i} in O .

- On tails, set $d(2i, a_{2i}) = 3$, $d(2i, a_j) = 1$ for $j \neq 2i$, and $d(2i - 1, a_j) = 1$ for $j = 1, \dots, m$. Include alternative a_{2i-1} in O .

Observe that for each agent, all alternatives are at distance 1, except (possibly) the alternative that is ranked last, which is at distance 3. Hence, d is indeed a distance function that is consistent with the profile. In addition, the alternative that is included in committee O at step i is the q -th closest (recall that $q = m/2$) to both agents $2i - 1$ and $2i$ and, furthermore, it is at distance 1 from both of them. Hence, $SC_q(O) = m$.

Now, consider a committee $C \in P$. For $i = 1, \dots, m/2$, let X_i be the random variable denoting the sum $c_q(2i - 1, C) + c_q(2i, C)$. Notice that

$$SC_q(C) = \sum_{i=1}^n c_q(i, C) = \sum_{i=1}^{m/2} X_i. \tag{1}$$

Now consider each random variable X_i separately. We distinguish between four cases:

- If C contains both alternatives a_{2i-1} and a_{2i} , then $c_q(2i - 1, C) = d(2i - 1, a_{2i-1})$ and $c_q(2i, a_{2i}) = d(2i, a_{2i})$. Notice that, by the definition of the random process, among these two quantities, one is equal to 1 and the other is equal to 3. Hence, $X_i = 4$.
- If C contains neither a_{2i-1} nor a_{2i} , then both agents $2i - 1$ and $2i$ are at distance 1 from all alternatives in C . Hence, $X_i = 2$.
- If C contains alternative a_{2i-1} but not alternative a_{2i} , notice that a_{2i-1} is the alternative of C that is the q -th closest to both agents $2i - 1$ and $2i$. Recall that $d(2i, a_{2i-1})$ is always equal to 1 while $d(2i - 1, a_{2i-1})$ is equiprobably equal to either 3 (and, hence, $X_i = 4$) or 1 (and, hence, $X_i = 2$), depending on whether the outcome of the i -th coin toss is heads or tails, respectively.
- The case when C contains alternative a_{2i} but not a_{2i-1} is completely symmetric. Then, $X_i = 2$ or $X_i = 4$ equiprobably, depending on whether the outcome of the i -th coin toss is heads or tails, respectively.

We conclude that $X_1, \dots, X_{m/2}$ are independent random variables, taking their values in $\{2, 4\}$, with

$$\mathbb{E}[X_i] = 2 + |C \cap \{a_{2i-1}, a_{2i}\}|. \tag{2}$$

Hence, we can use the well-known Hoeffding bound to estimate the probability $\Pr[SC_q(C) \leq (\frac{3}{2} - \varepsilon)m]$.

Lemma 4 ([21]). *Let Z_1, Z_2, \dots, Z_ℓ be independent random variables with $Z_i \in [a_i, b_i]$ and $Z = \sum_{i=1}^\ell Z_i$. Then, for every $t > 0$,*

$$\Pr[\mathbb{E}[Z] - Z \geq t] \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^\ell (b_i - a_i)^2}\right).$$

By (1), (2), and using linearity of expectation, we have

$$\mathbb{E}[SC_q(C)] = \sum_{i=1}^{m/2} \mathbb{E}[X_i] = \sum_{i=1}^{m/2} (2 + |C \cap \{a_{2i-1}, a_{2i}\}|) = 3m/2.$$

Using Lemma 4 with $\ell = m/2$, $Z = SC_q(C)$, $Z_i = X_i$, $b_i = 4$, and $a_i = 2$ for $i = 1, \dots, m/2$, and $t = \varepsilon m$, we get

$$\Pr\left[SC_q(C) \leq \left(\frac{3}{2} - \varepsilon\right)m\right] = \Pr[\mathbb{E}[SC_q(C)] - SC_q(C) \geq \varepsilon m] \leq \exp(-\varepsilon^2 m).$$

By applying the union bound, the probability that $SC_q(C) \leq (\frac{3}{2} - \varepsilon)m$ for some committee $C \in P$, is less than 1. We conclude that, with strictly positive probability, $SC_q(C) > (\frac{3}{2} - \varepsilon)m$ for every committee $C \in P$, as desired. \square

References

[1] Elliot Anshelevich, John Postl, Randomized social choice functions under metric preferences, *J. Artif. Intell. Res.* 58 (2017) 797–827.
 [2] Elliot Anshelevich, Onkar Bhardwaj, John Postl, Approximating optimal social choice under metric preferences, in: *Proceedings of the 29th AAAI Conference on Artificial Intelligence (AAAI)*, 2015, pp. 777–783.
 [3] Elliot Anshelevich, Aris Filos-Ratsikas, Nisarg Shah, Alexandros A. Voudouris, Distortion in social choice problems: the first 15 years and beyond, in: *Proceedings of the 30th International Joint Conference on Artificial Intelligence (IJCAI)*, 2021, pp. 4294–4301.
 [4] Haris Aziz, Markus Brill, Vincent Conitzer, Edith Elkind, Rupert Freeman, Toby Walsh, Justified representation in approval-based committee voting, *Soc. Choice Welf.* 48 (2) (2017) 461–485.
 [5] Gerdus Benadè, Swaprava Nath, Ariel D. Procaccia, Nisarg Shah, Preference elicitation for participatory budgeting, *Manag. Sci.* 67 (5) (2021) 2813–2827.

- [6] Craig Boutilier, Ioannis Caragiannis, Simi Haber, Tyler Lu, Ariel D. Procaccia, Or Sheffet, Optimal social choice functions: a utilitarian view, *Artif. Intell.* 227 (2015) 190–213.
- [7] Robert Bredereck, Piotr Faliszewski, Ayumi Igarashi, Martin Lackner, Piotr Skowron, Multiwinner elections with diversity constraints, in: *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI)*, 2018, pp. 933–940.
- [8] Ioannis Caragiannis, Ariel D. Procaccia, Voting almost maximizes social welfare despite limited communication, *Artif. Intell.* 175 (9–10) (2011) 1655–1671.
- [9] Ioannis Caragiannis, Swaprava Nath, Ariel D. Procaccia, Nisarg Shah, Subset selection via implicit utilitarian voting, *J. Artif. Intell. Res.* 58 (2017) 123–152.
- [10] John R. Chamberlin, Paul N. Courant, Representative deliberations and representative decisions: proportional representation and the Borda rule, *Am. Polit. Sci. Rev.* 77 (3) (1983) 718–733.
- [11] Moses Charikar, Prasanna Ramakrishnan, Metric distortion bounds for randomized social choice, in: *Proceedings of the 2022 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2022, pp. 2986–3004.
- [12] Xujin Chen, Minming Li, Chenhao Wang, Favorite-candidate voting for eliminating the least popular candidate in a metric space, in: *Proceedings of the 34th AAAI Conference on Artificial Intelligence (AAAI)*, 2020, pp. 1894–1901.
- [13] Soroush Ebadian, Anson Kahng, Dominik Peters, Nisarg Shah, Optimized distortion and proportional fairness in voting, in: *Proceedings of the 23rd ACM Conference on Economics and Computation (EC)*, 2022.
- [14] Edith Elkind, Piotr Faliszewski, Piotr Skowron, Arkadii Slinko, Properties of multiwinner voting rules, *Soc. Choice Welf.* 48 (3) (2017) 599–632.
- [15] Piotr Faliszewski, Piotr Skowron, Arkadii Slinko, Nimrod Talmon, Multiwinner voting: a new challenge for social choice theory, *Trends Comput. Soc. Choice* 74 (2017) (2017) 27–47.
- [16] Uriel Feige, A threshold of $\ln n$ for approximating set cover, *J. ACM* 45 (4) (1998) 643–652.
- [17] Allan Gibbard, Manipulation of voting schemes, *Econometrica* 41 (1973) 587–602.
- [18] Vasilis Gkatzelis, Daniel Halpern, Nisarg Shah, Resolving the optimal metric distortion conjecture, in: *Proceedings of the 61st IEEE Annual Symposium on Foundations of Computer Science (FOCS)*, 2020, pp. 1427–1438.
- [19] Ashish Goel, Anilesh K. Krishnaswamy, Kamesh Munagala, Metric distortion of social choice rules: lower bounds and fairness properties, in: *Proceedings of the 2017 ACM Conference on Economics and Computation (EC)*, 2017, pp. 287–304.
- [20] Ashish Goel, Reyna Hulett, Anilesh Kollagunta Krishnaswamy, Relating metric distortion and fairness of social choice rules, arXiv:1810.01092, 2018.
- [21] Wassily Hoeffding, Probability inequalities for sums of bounded random variables, *J. Am. Stat. Assoc.* 58 (301) (1963) 13–30.
- [22] Michal Jaworski, Piotr Skowron, Evaluating committees for representative democracies: the distortion and beyond, in: *Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI)*, 2020, pp. 196–202.
- [23] David Kempe, Communication, distortion, and randomness in metric voting, in: *Proceedings of the 34th AAAI Conference on Artificial Intelligence (AAAI)*, 2020, pp. 2087–2094.
- [24] David Kempe, An analysis framework for metric voting based on LP duality, in: *Proceedings of the 34th AAAI Conference on Artificial Intelligence (AAAI)*, 2020, pp. 2079–2086.
- [25] Tyler Lu, Craig Boutilier, Budgeted social choice: from consensus to personalized decision making, in: *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI)*, 2011, pp. 280–286.
- [26] Reshef Meir, Fedor Sandomirskiy, Moshe Tennenholtz, Representative committees of peers, *J. Artif. Intell. Res.* 71 (2021) 401–429.
- [27] Burt L. Monroe, Fully proportional representation, *Am. Polit. Sci. Rev.* 89 (4) (1995) 925–940.
- [28] Rajeew Motwani, Prabhakar Raghavan, *Randomized Algorithms*, Cambridge University Press, 1995.
- [29] Kamesh Munagala, Kangning Wang, Improved metric distortion for deterministic social choice rules, in: *Proceedings of the 2019 ACM Conference on Economics and Computation (EC)*, 2019, pp. 245–262.
- [30] Dominik Peters, Piotr Skowron, Proportionality and the limits of welfarism, in: *Proceedings of the 21st ACM Conference on Economics and Computation (EC)*, 2020, pp. 793–794.
- [31] Dominik Peters, Grzegorz Pierczynski, Nisarg Shah, Piotr Skowron, Market-based explanations of collective decisions, in: *Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI)*, 2021, pp. 5656–5663.
- [32] Ariel D. Procaccia, Jeffrey S. Rosenschein, The distortion of cardinal preferences in voting, in: *International Workshop on Cooperative Information Agents (CIA)*, 2006, pp. 317–331.
- [33] Ariel D. Procaccia, Jeffrey S. Rosenschein, Aviv Zohar, On the complexity of achieving proportional representation, *Soc. Choice Welf.* 30 (3) (2008) 353–362.
- [34] Mark A. Satterthwaite, Strategy-proofness and Arrow's conditions: existence and correspondence theorems for voting procedures and social welfare functions, *J. Econ. Theory* 10 (1975) 187–217.
- [35] Piotr Skowron, Edith Elkind, Social choice under metric preferences: scoring rules and STV, in: *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI)*, 2017, pp. 706–712.