Correlating Preferences and Attributes: Nearly Single-Crossing Profiles

Foram Lakhani , **Dominik Peters** and **Edith Elkind**

University of Oxford, Oxford, UK

foramlakhani@gmail.com, dominik.peters@cs.ox.ac.uk, elkind@cs.ox.ac.uk

Abstract

We use social choice theory to develop correlation coefficients between ranked preferences and an ordinal attribute such as educational attainment or income level. For example, such correlations could be used to formalise statements such as "voters' preferences over parties are better explained by age than by income level". In the literature, preferences that are perfectly explained by a single-dimensional agent attribute are commonly taken to be single-crossing preferences. Thus, to quantify how well an attribute explains preferences, we can order the voters by the value of the attribute and compute how far the resulting ordered profile is from being single-crossing, for various commonly studied distance measures (Kendall tau distance, voter/alternative deletion, etc.). The goal of this paper is to evaluate the computational feasibility of this approach. To this end, we investigate the complexity of computing these distances, obtaining an essentially complete picture for the distances we consider.

1 Introduction

A key task of social science is to understand and explain people's preferences and attitudes. A psychologist might ask survey takers how much they agree with a statement on a 1–10 scale (from "disagree" to "agree"). The researcher can then compute correlations between this quantity and participants' *attributes* such as age, income, or educational attainment.

This way of explaining attitudes and preferences in terms of attributes has proved powerful—but it is intrinsically limited. In particular, this method cannot easily explain complex preferences that are not captured by a single numerical response. For example, suppose we want to ask participants to rank various forms of energy generation (such as coal, solar, nuclear) in terms of desirability in their country's energy mix. We might suspect people's answer to be mainly a function of their education level, or maybe of their left-right political orientation. How can we check this?

Statistics textbooks [e.g., de Vaus, 2002, p. 103] would recommend handling each energy source independently, and find the correlation between participants' education level (maybe expressed as number of school years) and the rank of the

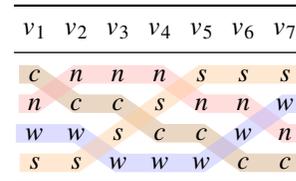| v_1	v_2	v_3	v_4	v_5	v_6	v_7
c	n	n	n	s	s	s
n	c	c	s	n	n	w
w	w	s	c	c	w	n
s	s	w	w	w	c	c

Figure 1: A single-crossing profile.

alternative in their ranking. A first complaint might be that this approach cardinalises inherently ordinal information, but there are notions of correlation for ordinal measures (e.g., Healey, 2008, Sec. 14.6; for example, take the Kendall tau distance between ordering people by education and ordering them by how they rank the alternative). A more severe problem of the approach is that it ignores the finer structure of the preferences, by considering each alternative on its own. Fundamentally, this approach only answers questions like "do more-educated people like solar more?" and not "are people's preferences over all energy sources well-explained by their education level?"

What would be a better way to answer questions of the latter type? We use the lens of social choice theory, a well-developed literature on handling preferences. While its main focus is on preference *aggregation*, parts of it can be repurposed for the goal of understanding preferences. This is particularly true for work concerning *structured preferences*, which have received much recent attention from computer scientists [Elkind *et al.*, 2016; 2017]. Our focus in this paper is a concept called *single-crossing preferences*, which goes back to work on income taxation [Mirrlees, 1971]. It turns out to be particularly relevant for our goal of "correlating" preferences and attributes.

Suppose we have obtained rankings from 7 participants over coal, nuclear, wind, and solar (c, n, w, s respectively). In Figure 1, we write these rankings as columns, and sort them left-to-right by the voter attribute. The coloured lines track the rank of each alternative and show an intriguing feature: any two of these 'trajectories' cross only once. For example, people on the left prefer wind to solar and people on the right prefer solar to wind; and a similar pattern holds for every pair of alternatives. This is the single-crossing property. In our view, this property implies that the rankings have been sorted "correctly", and it is easy to check that a single-crossing order is always unique (up to mirroring and reordering identical votes). Thus, we hold that an attribute that orders people as

shown "perfectly correlates" with the preferences.

In profiles with a large number of agents, there may not exist an ordering which makes the profile single-crossing, either due to noise or due to intrinsic complexity of the preferences. Still, some orderings might be better than others, in the sense of making the profile *almost* single-crossing. In this paper, we propose a way to measure the quality of different orderings in terms of notions of distance to being single-crossing.

To be concrete, suppose we are given a profile of rankings which are ordered in a fixed way (e.g., sorted by an exogenously given attribute). This profile is not single-crossing. How far away is it? A natural measure is provided by the *Kendall tau* distance: we ask how many times we need to swap two adjacent alternatives in one of the rankings until the profile is single-crossing ("global swaps"). Instead of counting the total number of swaps, we can also count the maximum number of swaps applied to any single ranking ("local swaps"). More coarse measures can also be sensible: for example, some of the rankings may be outliers, so we can count the number of rankings that need to be deleted until the rest is single-crossing ("voter deletion"). We consider some other distances as well; the best choice of distance will depend on the context.

Compared to the traditional method of computing correlations independently for each alternative, our proposals are more combinatorial and complex. In fact, it is not clear how to compute these distance measures efficiently. While we develop an efficient algorithm to evaluate the voter deletion distance, for most other metrics we prove that evaluation is NP-hard (see Table 1). This should not be a major impediment to using our notions in practice, and we show that evaluating the swap-based metrics becomes easy (in the FPT or XP sense) if the profile in question is very close to being single-crossing.

The literature on structured preferences has introduced and studied problems similar to ours. The most popular notion of structure is *single-peakedness*, which assumes that the alternative space is one-dimensional. The concept of nearly single-peaked preferences was suggested by Faliszewski *et al.* [2014], and then a number of authors have considered the complexity of identifying preferences that are nearly single-peaked [Erdélyi *et al.*, 2017; Bredereck *et al.*, 2016; Elkind *et al.*, 2012; Cornaz *et al.*, 2013], i.e., can be made single-peaked by deleting a few voters or alternatives, merging small groups of similar alternatives, performing a few swaps of adjacent alternatives in voters' rankings, or partitioning voters or candidates into a few groups so that the profile is single-peaked when restricted to one of these groups. Nearly single-crossing preferences have also been studied [Bredereck *et al.*, 2016; Jaeckle *et al.*, 2018]. However, those papers focused on a setting that is crucially different: they do not fix the ordering of the rankings, so the algorithms may rearrange them arbitrarily to obtain a single-crossing profile. That regime makes sense for voting, if we are aiming to use fast winner determination algorithms that require single-crossing input [e.g., Skowron *et al.*, 2015], but it is not relevant for our aim of quantifying the explanatory power of a given voter attribute.

For consistency with the literature on nearly structured preferences, in what follows we will use the terminology of computational social choice; for example, we refer to agents as voters. Due to space constraints, we omit some proofs.

2 Preliminaries

For an integer n, we write $[n] = \{1, \ldots, n\}$.

Preferences. Let A be a finite set of *alternatives*. A *linear order v* over A is a binary relation that is asymmetric, total, and transitive. We write $\mathcal{L}(A)$ for the set of linear orders over A. A *profile $P = (v_1, \ldots, v_n)$* is a list of linear orders; we write $|P| = n$ and refer to elements of P as *votes*, and to elements of $[n]$ as *voters*. We write $a >_i b$ when $(a, b) \in v_i$, i.e., when voter i prefers a to b. Let $\mathcal{P}(A)$ denote the set of all profiles over A. Given t profiles $P_1, \ldots, P_t \in \mathcal{P}(A)$, let $P_1 + \cdots + P_t$ be the profile obtained by concatenating $P_1, \ldots, P_t$ (in the given order); if $P_1 = \cdots = P_t = P$, we write tP for $P_1 + \cdots + P_t$. Let $X \subseteq A$ be a subset of alternatives. For a linear order v, we write $v|_X = v \cap (X \times X)$, and for a profile P, we write $P|_X = (v_1|_X, \ldots, v_n|_X)$; $v|_X$ and $P|_X$ are, respectively, the linear order and the profile obtained by restricting the alternative set to X. Given a subset $I \subseteq [n]$, we denote by P_I the profile obtained from P by deleting all voters in $[n] \setminus I$.

Single-crossing profiles. A profile $P \in \mathcal{P}(A)$ is *single-crossing (in the given order)* if for each pair $a, b \in A$ of alternatives such that $a >_1 b$ there is some $i \in [n]$ such that $a >_j b$ for all $j \in [i]$ and $b >_j a$ for all $j \in [n] \setminus [i]$. We say that a profile P is *single-crossing up to a permutation* if the votes in P can be permuted so that the resulting profile is single-crossing in the given order. Given a profile $P \in \mathcal{P}(A)$, $|P| = n$, we say that a pair of alternatives $\{a, b\}$ is *multi-crossing* if there exist $i, j, k \in [n]$, $i < j < k$, such that $a >_i b$, $b >_j a$, $a >_k b$. Note that a profile is single-crossing in the given order if and only if no pair of alternatives is multi-crossing.

Graph-theoretic background. An *undirected graph* is a pair $G = (A, L)$, where A is a set of *vertices* and L is a collection of size-two subsets of A, which are referred to as *edges*. A vertex set $X \subseteq A$ is an *independent set* if for every $a, b \in X$ we have $\{a, b\} \notin L$. A mapping $\chi : A \rightarrow [k]$ is a *k-colouring* of G if for each $i \in [k]$ the set $\{a \in A \mid \chi(a) = i\}$ is an independent set. A graph $G = (A, L)$ is *s-regular* if for each $a \in A$ we have $|\{e \in L \mid a \in e\}| = s$. We denote the set of all graphs with vertex set $\{a_1, \ldots, a_m\}$ by $\mathcal{G}_m$.

Parameterised complexity. A popular approach to deal with NP-complete problems is to consider their complexity as a function of certain problem parameters [see, e.g., Downey and Fellows, 2013], such as the number of alternatives or voters. Problems that admit algorithms whose running time is of the form $f(p)\text{poly}(n)$, where p is the value of the parameter, f is a computable function, and n is the instance size, are called fixed parameter-tractable (FPT) with respect to p; these are the 'easiest' problems in the parameterised sense. A problem belongs to the complexity class XP with respect to p if it can be solved in polynomial time for each fixed value of p (it admits an algorithm with running time $n^{f(p)}$ for a computable function f). Clearly, FPT $\subseteq$ XP.

3 Pairwise Swaps

We begin with our most fine-grained distance concepts, namely, the ones that are based on swaps. Swapping a pair of adjacent alternatives is, arguably, a minimal change to a preference

Computational problem	Recognition	Small n and parameterised
ALTERNATIVE DELETION	NPc (Th.6)	NPc for $n \geqslant 4$ (Th.7), in P for $n = 3$ (Cor.9), FPT wrt k (Pr.5)
k-ALTERNATIVE PARTITION	NPc (Th.6)	NPc for $n \geqslant 4$ (Th.8), in P for $n = 3$ (Cor.9) or $k = 2$ (Pr.5)
GLOBAL SWAPS	NPc (Th.1)	FPT wrt k (Th.3)
LOCAL SWAPS	NPc (Th.2)	XP wrt k (Cor.4)
VOTER DELETION	$O(n^3 m^2)$ (Pr.12)	n/a
CONTIGUOUS VOTER PARTITION	$O(nm^2)$ (Pr.12)	n/a
INTERLEAVING VOTER PARTITION	?	XP wrt k (Th.13)
VOTER REORDERING	$O(n(\log n + m^2))$ (Pr.14)	n/a

Table 1: Overview of complexity results. Here n is the number of voters, m is the number of alternatives, and k is the 'budget', such as, e.g., the maximum number of alternatives that can be deleted. 'NPc' stands for 'NP-complete'.

profile, and so it makes sense to count the number of such changes as a distance. This yields the *Kendall tau distance*, which counts the number of pairs of alternatives on which two linear orders disagree. This distance model is also well-motivated when voters' preferences are noisy, and this is the reason why the profile is not single-crossing. Formally, given two votes v_i, v_j, we define $d_K(v_i, v_j) = |\Delta(v_i, v_j)|$, where $\Delta(v_i, v_j) = \{\{a, b\} \mid a \succ_i b, \ b \succ_j a\}$. This distance can be extended to n-voter profiles: if $P = (v_1, \ldots, v_n)$ and $Q = (u_1, \ldots, u_n)$, we set $d_K(P, Q) = \sum_{i \in [n]} d_K(v_i, u_i)$. The *egalitarian*, or ℓ_0, variant of this distance uses the maximum number of swaps per voter: $d_{\text{egal}}(P, Q) = \max_{i \in N} d(v_i, u_i)$.

These definitions suggest the following decision problems.

Definition 1. *An instance of* GLOBAL SWAPS *(respectively,* LOCAL SWAPS*) problem is given by a profile P and an integer k. It is a yes-instance if and only if there exists a profile Q that is single-crossing in the given order and satisfies $d_K(P, Q) \leqslant k$ (respectively, $d_{\text{egal}}(P, Q) \leqslant k$).*

We will show that both of these problems are NP-complete. Our proof is by reduction from the problem of computing, respectively, the Kemeny distance and the egalitarian Kemeny distance. Thus, we reduce from one voting problem to another, and this allows the proof to be short and simple.

A profile $P = (v_1, \ldots, v_n)$ is *unanimous* if $v_i = v_j$ for all $i, j \in [n]$. An instance of KEMENY SCORE (respectively, EGALITARIAN KEMENY SCORE) is given by a profile P and an integer s; it is a yes-instance if there exists a unanimous profile Q with $d_K(P, Q) \leqslant s$ (respectively, with $d_{\text{egal}}(P, Q) \leqslant s$). Both of these problems are NP-complete [Bartholdi III *et al.*, 1989; Biedl *et al.*, 2009].

Theorem 1. GLOBAL SWAPS *is NP-complete.*

Proof. This problem is clearly in NP. To prove NP-hardness, we reduce from KEMENY SCORE. We map an instance $\langle P, s \rangle$ of KEMENY SCORE with $P \in \mathcal{P}(A)$, $|A| = m$, to an instance $\langle tP, ts \rangle$ of GLOBAL SWAPS, where $t = (s + 1)\binom{m}{2} + 1$.

If $\langle P, s \rangle$ is a yes-instance of KEMENY SCORE, there is a unanimous profile Q with $d_K(P, Q) \leqslant s$. So tQ is unanimous and hence single-crossing in the given order, and $d_K(tP, tQ) \leqslant ts$. Hence $\langle tP, ts \rangle$ is a yes-instance of GLOBAL SWAPS.

Conversely, suppose $\langle tP, ts \rangle$ is a yes-instance of GLOBAL SWAPS, and R is a profile that is single-crossing in the given order and satisfies $d(tP, R) \leqslant ts$. Split R as $R_1 + \cdots + R_t$, $|R_j| = n$ for each $j \in [t]$. We say that the i-th vote in R

is a *crossing point* for alternatives $a, b \in A$ if the i-th voter in R prefers a to b, while the $(i + 1)$-st voter prefers b to a. Since R is single-crossing in the given order, there is at most one crossing point for each pair $\{a, b\} \subseteq A$, and hence at most $\binom{m}{2}$ crossing points altogether. Let $J = \{j \in [t] \mid R_j$ contains no crossing points$\}$. Then $|J| \geqslant t - \binom{m}{2} \geqslant 1$. Now, if $d_K(P, R_j) > s$ for all $j \in J$, we can lower-bound $d_K(tP, R)$ as

$$(s + 1)|J| \geqslant (s + 1)\left(t - \binom{m}{2}\right) = st + t - (s + 1)\binom{m}{2} > st,$$

a contradiction. Hence, $d_K(P, R_j) \leqslant s$ for some $j \in J$. Since R_j contains no crossing points, it is unanimous, which means that we started with a yes-instance of KEMENY SCORE. □

A similar reduction from EGALITARIAN KEMENY SCORE shows that LOCAL SWAPS is also hard.

Theorem 2. LOCAL SWAPS *is NP-complete.*

On the positive side, GLOBAL SWAPS turns out to be fixed-parameter tractable with respect to the number of swaps.

Theorem 3. GLOBAL SWAPS *is in FPT with respect to k.*

Proof. We use the bounded search tree technique. Specifically, we describe an algorithm that starts with a budget of k swaps and aims to transform P into a profile that is single-crossing in the given order. It makes some non-deterministic choices, but the number of such choices at each decision node can be bounded by a function that depends on k only; the algorithm rejects a given non-deterministic branch if it has already performed k swaps along this branch, but has not obtained a single-crossing profile yet. This algorithm can be simulated by a deterministic algorithm that explores all branches of non-determinism in order of depth-first search.

Throughout the algorithm, we label some triples (i, a, b), where $i \in [n]$ and $a, b \in A$ as *fixed*; intuitively, once a triple has been fixed, we cannot change the preferences of voter i over the pair $\{a, b\}$. Initially, the set of fixed triples is empty.

Given a voter i and two alternatives $a, b \in A$ such that $a \succ_i b$, we denote by $i[a, b]$ the set of alternatives that i ranks between a and b (including a and b). We modify voters' preferences while executing the algorithm; the set $i[a, b]$ is defined with respect to the *current* preferences of voter i.

The algorithm maintains a stack of pairs of alternatives $\{a, b\}$. At the beginning of the algorithm, we push all pairs of alternatives that are multi-crossing in P onto this stack. At

each stage of the algorithm we pop one pair of alternatives off the stack and change some voters' preferences so that this pair is no longer multi-crossing; the swap budget is updated to account for these changes, and we reject if we run out of budget or if it is clear that the current pair cannot be 'fixed' within the remaining budget. As a result of our changes, some new pairs of alternatives may become multi-crossing, and they are then pushed on the stack. The algorithm returns 'yes' if for some sequence of non-deterministic choices it can reach a state where at the end of a stage the stack is empty.

It remains to describe what we do once we have popped some pair $\{a, b\}$ off the stack. Let r be the remaining swap budget at this point, i.e., we assume that the algorithm has already performed $k - r$ swaps.

We say that $i \in \{0\} \cup [n]$ is a *crossing point* for $\{a, b\}$ if (1) $i = 0$, (2) $i = n$, or (3) $0 < i < n$ and voters i and $i + 1$ disagree on $\{a, b\}$, i.e., $a >_i b, b >_{i+1} a$. Let $(i_1, \ldots, i_t), i_1 < \cdots < i_t$, be the ordered list of crossing points for $\{a, b\}$. For each $\ell \in [t - 1]$, all voters between $i_\ell + 1$ and $i_{\ell+1}$ agree on $\{a, b\}$, i.e., the list of voters can be split into $t - 1$ intervals, where all voters in the same interval agree on $\{a, b\}$. Note that a single swap reduces the number of intervals by at most 2 (e.g., if two intervals where all voters prefer a to b are separated by a single voter who prefers b to a, swapping b and a in the preferences of that voter merges these three intervals into one). Thus, if $t - 1 - 2r > 2$, we can safely reject, as a and b will necessarily remain multi-crossing after we perform r or fewer swaps. Thus, from now on we assume that $t \leqslant 2r + 2 \leqslant 2k + 2$.

Now, we have to 'fix' the pair $\{a, b\}$ so that there is at most one crossing point. To this end, we non-deterministically guess the first voter's preference over the pair $\{a, b\}$ and the new crossing point $i \in [n]$ ($i = n$ corresponds to the case where all voters agree on $\{a, b\}$); these two guesses completely determine all voters' preferences over $\{a, b\}$. Importantly, i need not belong to $\{i_1, \ldots, i_t\}$, but we can limit our guesses for i to lie within an r-neighborhood of some existing crossing point for $\{a, b\}$, i.e., we will only consider values of i such that $|i - i_\ell| \leqslant r$ for some $\ell \in [t]$. Indeed, if $i_p < i < i_{p+1}$ and $i - i_p > r, i_{p+1} - i > r$, then, to make i the unique crossing point, we would have to swap a and b in the preferences of all voters $i_p + 1, \ldots, i$ or in the preferences of all voters $i + 1, \ldots, i_{p+1}$, i.,e., perform more than r swaps. This means that we only have to consider $(2r + 1)t \leqslant (2k + 1)(2k + 2)$ positions for the new crossing point; this quantity does not depend on n or m.

Let us say that a voter i is *bad* if her current preferences over $\{a, b\}$ are different from what is implied by our guess for the crossing point and the preferences of the first voter; denote the set of all bad voters by B. Next, we would like to swap a and b in the preferences of the bad voters. There is a difficulty, however: if for some voter $i \in B$ alternatives a and b are not adjacent in i's current vote, i.e., if $|i[a, b]| > 2$, changing i's preferences over $\{a, b\}$ results in preferences that are not transitive. If we want i's preferences to be a (transitive) linear order, each alternative in $i[a, b] \setminus \{a, b\}$ would have to be swapped with a or b. This means, in particular, that we can reject if $|i[a, b]| > r + 2$ for some bad voter i, so from now on we assume that $|i[a, b]| \leqslant r + 2$ for each $i \in B$.

Now, we process the voters in B one by one. For each $i \in B$

we non-deterministically guess an ordering of all alternatives in $i[a, b]$ in i's preferences; for a given voter, there are at most $|i[a, b]|! \leqslant (r + 2)! \leqslant (k + 2)!$ options to consider. We reject our guess if (1) it does not swap a and b, or (2) the Kendall tau distance between the new vote for i and her current vote exceeds r, or (3) our guess is inconsistent with the current set of fixed triples: i.e., $c >_i d$, (i, c, d) is a fixed triple, and we have guessed a permutation of $i[a, b]$ where d is placed above c. Otherwise, we reorder the alternatives in $i[a, b]$ in the ranking of voter i according to our guess, and add all triples of the form $(i, c, d), c, d \in i[a, b]$, to the set of fixed triples.

Once we have processed all voters in B, the pair $\{a, b\}$ is no longer multi-crossing; however, some other pairs may have become multi-crossing. Thus, once all voters in B have been processed, we scan the resulting profile and add all the newly multi-crossing pairs to the stack. This completes the description of a single stage of our algorithm.

It is immediate that our algorithm only accepts if it has identified a profile that is single-crossing in the given order and can be obtained from P by performing at most k swaps. Conversely, we only reject if each non-deterministic branch is eventually rejected, and this only happens if every choice we consider does not result in a single-crossing profile after k or fewer swaps. As we have argued that our algorithm explores all options that may possibly lead to a positive answer, this establishes correctness. For the analysis of running time, we observe that each non-deterministic branch performs at most k swaps and executes some fairly simple operations between the swaps. Further, the number of non-deterministic choices at each point is bounded by a function of k. This implies the desired bound on the running time. □

For Local Swaps, we do not know if this problem is in FPT with respect to k. However, Jaeckle et al. [2018] show that the analogue of this problem for preferences that are single-crossing up to a permutation is in XP with respect to k. This proof can be modified to apply to our setting.

Corollary 4. Local Swaps *is in XP with respect to k.*

We note that (Egalitarian) Kemeny score is known to be NP-hard even for $n = 4$, but our reductions do not preserve the number of voters. We conjecture that Global Swaps and Local Swaps remain hard for small values of n.

4 Alternative Deletion and Partition

In this section, we consider two notions of distance to single-crossing preferences that are based on modifying the set of alternatives. The first measure asks how many alternatives need to be deleted to make the given profile single-crossing; the second measure asks for the smallest value k such that the alternatives can be split into k groups so that restricting the voters' preferences to each group results in a profile that is single-crossing in the given order. Both of these measures have been considered in the context of profiles that are nearly single-peaked or single-crossing up to a permutation [Erdélyi et al., 2017; Bredereck et al., 2016; Jaeckle et al., 2018].

Definition 2. *An instance of* Alternative Deletion *is given by a profile $P \in \mathcal{P}(A)$ and an integer $k \geqslant 1$. It is a yes-instance*

if and only if there is a subset $X \subseteq A$ with $|X| \geq |A| - k$ such that $P|_X$ is single-crossing in the given order.

Definition 3. *An instance of k-ALTERNATIVE PARTITION is given by a profile $P \in \mathcal{P}(A)$. It is a yes-instance if and only if A can be partitioned into k sets $A_1, \ldots, A_k$ so that for each $j \in [k]$ the profile $P|_{A_j}$ is single-crossing in the given order.*

In a companion paper [Cohen *et al.*, 2019] we introduce a mapping between preference profiles and undirected graphs that is useful for analysing the problems we consider here. Specifically, given a profile $P \in \mathcal{P}(A)$, we define its *crossing graph* as an undirected graph $G = (A, L)$ such that $\{a, b\} \in L$ if and only if $\{a, b\}$ is a multi-crossing pair in P. Note that P can be made single-crossing by deleting a subset of alternatives $X \subseteq A$ if and only if X is a vertex cover of G; similarly, a partition of A into k sets such that the restriction of A to each set is single-crossing corresponds to a k-coloring of G. Since VERTEX COVER is in FPT with respect to the target cover size k, and one can decide in polynomial time whether a graph is 2-colourable, we obtain the following easiness results.

Proposition 5. ALTERNATIVE DELETION *is in FPT with respect to k, and 2-ALTERNATIVE PARTITION is in P.*

Cohen *et al.* [2019] prove, conversely, that any undirected graph on m vertices is a crossing graph of a preference profile with $O(m)$ voters. The following hardness result is a corollary.

Theorem 6 (Cohen *et al.*, 2019). ALTERNATIVE DELETION *is NP-complete. k-ALTERNATIVE PARTITION is NP-complete for $k \geq 3$.*

A limitation of the technique introduced by Cohen *et al.* [2019] is that the construction behind Theorem 6 requires an unbounded number of voters. We strengthen Theorem 6 by showing that these problems remain hard even for a small constant number of voters, namely even for $n = 4$ voters.

Theorem 7. *For every $n \geq 4$, ALTERNATIVE DELETION remains NP-complete when restricted to instances with n voters.*

Proof. We reduce from the INDEPENDENT SET problem. This problem remains NP-hard for 3-regular graphs [Garey *et al.*, 1976]; thus, in our reduction we assume that the input graph G is 3-regular. We establish hardness for $n = 4$; the result can be extended to any $n > 4$ by creating additional $n - 4$ copies of the last vote. Consider an instance $\langle G, k \rangle$ of INDEPENDENT SET, where $G = (A, L)$, $A = \{a_1, \ldots, a_m\}$, $L = \{e_1, \ldots, e_s\}$ and G is 3-regular. For every edge $e = \{a_i, a_j\}$ we create two alternatives $c(e, i)$ and $c(e, j)$ and set $C^e = \{c(e, i), c(e, j)\}$; also, for each $a_i \in A$ we create two alternatives d_i and d_i'. For each $a_i \in A$, let $C_i = \{c(e, i) \mid a_i \in e\}$ and $D_i = \{d_i, d_i'\}$; note that $|C_i| = 3$, since G is 3-regular.

We will now describe the profile P. For readability, we first explain how voters rank various groups of alternatives, and then describe their preferences within these groups.

The first two voters' rankings are of the form
$$C^{e_1} > \cdots > C^{e_s} > D_1 > \cdots > D_m,$$
the third voter's ranking is of the form
$$D_1 >_3 C_1 >_3 \cdots >_3 D_m >_3 C_m,$$
and the fourth voter's ranking is of the form
$$C_1 >_4 D_1 >_4 \cdots >_4 C_m >_4 D_m;$$

the order of alternatives within each group is described next.

For each $a_i \in A$ all voters rank the alternatives in D_i as $d_i > d_i'$. For each $e = \{a_i, a_j\} \in L$ with $i < j$, set $c(e, i) >_1 c(e, j)$, $c(e, j) >_2 c(e, i)$. Now, fix an $i \in [m]$ and suppose that $C_i = \{c(e_x, i), c(e_y, i), c(e_z, i)\}$, where $x < y < z$. The third and the fourth voter rank the alternatives in C_i as $c(e_x, i) > c(e_y, i) > c(e_z, i)$.

In the resulting profile for each $e = \{a_i, a_j\} \in L$ the alternatives $c(e, i)$ and $c(e, j)$ are multi-crossing; also, for each $a_i \in A$ every pair $\{c, d\}$ with $c \in C_i, d \in D_i$ is multi-crossing. No other pair of alternatives is multi-crossing.

Suppose $X \subseteq A$ is an independent set of size s in G. Let

$$Y = \bigcup_{a_i \in X} C_i \cup \bigcup_{a_i \notin X} D_i.$$

Then $|Y| = 3|X| + 2(m - |X|) = 2m + s$, and $P|_Y$ is single-crossing in the given order. Conversely, suppose that there is a set of alternatives Z, $|Z| \geq 2m + s$, such that $P|_Z$ is single-crossing in the given order. Let $X = \{a_i \in A \mid C_i \subseteq Z\}$. Then X is an independent set in G: if $e = \{a_i, a_j\} \subseteq X$ for some $e \in L$ then $C^e \subseteq Z$, a contradiction with $P|_Z$ being single-crossing in the given order. We will now show that $|X| \geq s$. Note that for each $i \in [m]$ we have $Z \cap C_i = \varnothing$ or $Z \cap D_i = \varnothing$ and hence $|Z \cap (C_i \cup D_i)| \leq 3$ and, moreover, $|Z \cap (C_i \cup D_i)| = 3$ if and only if $C_i \subseteq Z$, i.e., $a_i \in X$. We have $2m + s \leq |Z| \leq 3|X| + 2(m - |X|)$ and hence $|X| \geq s$. Thus, $\langle G, s \rangle$ is a yes-instance of INDEPENDENT SET if and only if $\langle P, 2m + s \rangle$ is a yes-instance of ALTERNATIVE DELETION. □

The proof for alternative partition is similar.

Theorem 8. *For each $n \geq 4$ and each $k \geq 3$, k-ALTERNATIVE PARTITION remains NP-complete when restricted to instances with n voters.*

Now, clearly, ALTERNATIVE PARTITION and ALTERNATIVE DELETION are easy for $n = 2$, as every 2-voter profile is single-crossing. It turns out that the remaining case $n = 3$ is easy, too, for an interesting graph-theoretic reason: Cohen *et al.* [2019] prove that crossing graphs of 3-voter profiles are perfect (for a definition of perfect graphs see, e.g., Diestel, 2012). The problems of finding a vertex cover or a k-colouring are in P for perfect graphs, and so we obtain the following corollary, which shows that the hardness results in Theorems 7 and 8 are tight with respect to the number of voters.

Corollary 9 (Cohen *et al.*, 2019). ALTERNATIVE DELETION *and k-ALTERNATIVE PARTITION are polynomial-time solvable when restricted to instances with three voters.*

5 Voter Deletion, Partition and Reordering

Finally, we look at distance measures that alter the set of voters.

Definition 4. *An instance of VOTER DELETION is given by a profile $P \in \mathcal{P}(A)$, $|A| = m$, $|P| = n$, and an integer $k \geq 1$. It is a yes-instance if and only if there is a subset $I \subseteq [n]$ with $|I| \geq n - k$ such that P_I is single-crossing in the given order.*

Bredereck *et al.* [2016] give a dynamic program for the voter deletion problem aiming for single-crossing up to a permutation. The same idea works in our setting.

Proposition 10. VOTER DELETION *can be solved in time* $O(n^3m^2)$.

We can also partition voters into several single-crossing subprofiles. We define two variants of this problem: in the first variant, each part of the partition must form a contiguous block, and in the second variant parts may interleave.

Definition 5. *An instance of* CONTIGUOUS VOTER PARTITION *is given by a profile* $P \in \mathcal{P}(A)$, $|A| = m$, $|P| = n$, *and an integer* $k \geqslant 1$. *It is a yes-instance if and only if there exist* k *profiles* $P_1, \ldots, P_k$ *such that* P_j *is single-crossing in the given order for each* $j \in [k]$ *and* $P = P_1 + \cdots + P_k$.

Definition 6. *An instance of* INTERLEAVING VOTER PARTITION *is given by a profile* $P \in \mathcal{P}(A)$, $|A| = m$, $|P| = n$, *and an integer* $k \geqslant 1$. *It is a yes-instance if and only if the set* $[n]$ *can be partitioned into* k *sets* $I_1, \ldots, I_k$ *so that for each* $j \in [k]$ *the profile* P_j *obtained from* P *by removing the voters in* $\{v_i \mid i \in [n] \setminus I_j\}$ *is single-crossing in the given order.*

These two problems seem different from a complexity perspective: CONTIGUOUS VOTER PARTITION can be solved by a simple greedy algorithm, but we do not know if INTERLEAVING VOTER PARTITION is in P. However, the latter problem is in the class XP with respect to k. Our analysis makes use of the following simple lemma [see, e.g., Jaeckle *et al.*, 2018, Sec. 2].

Lemma 11. *Let* $P = (v_1, \ldots, v_n) \in \mathcal{P}(A)$ *be a single-crossing profile. Let* v_{n+1} *be a vote in* $\mathcal{L}(A)$. *Then* $P + (v_{n+1})$ *is single-crossing in the given order if and only if* (v_1, v_n, v_{n+1}) *is.*

Proposition 12. CONTIGUOUS VOTER PARTITION *can be solved in time* $O(nm^2)$.

Proof. Let $P = (v_1, \ldots, v_n)$. We construct the profiles $P_1, \ldots, P_k$ greedily. Initially we set $j = 1$, $P_1 = \varnothing$. We process the votes in P from left to right. When considering the vote v_i, we check whether $P_j + (v_i)$ is single-crossing in the given order (this can be done in time $O(m^2)$ by Lemma 11). If yes, we set $P_j := P_j + (v_i)$ and otherwise we set $j := j + 1$, $P_j = (v_i)$. We stop once all votes have been processed, and output the number of subprofiles in our partition. We omit the proof of correctness. $\square$

Theorem 13. INTERLEAVING VOTER PARTITION *can be solved in time* $O(n^{2k+3}m^2)$.

Proof. We guess k voters $f_1, \ldots, f_k$; we then check if there is an interleaving partition such that the first voter in part j is f_j, for $j \in [k]$. Note that there are at most n^k guesses to consider. We then proceed by dynamic programming. For each $i \in [n]$ and $\ell_1, \ldots, \ell_k \in [i] \cup \{\bot\}$, let $A[i; \ell_1, \ldots, \ell_k] = 1$ if there is a partition of $[i]$ into k pairwise disjoint sets $I_1, \ldots, I_k$ such that for each $j \in [k]$ it holds that P_{I_j} is single-crossing in the given order, $\ell_j = \bot$ if and only if $f_j > i$, and, moreover, if $f_j \leqslant i$ then the first voter in I_j is f_j and the last voter in I_j is ℓ_j; otherwise, let $A[i; \ell_1, \ldots, \ell_k] = 0$.

For $i = 1$, we have $A[i; \ell_1, \ldots, \ell_k] = 1$ if and only if there is a $j \in [k]$ with $f_j = \ell_j = 1$, $\ell_t = \bot$ for $t \in [k] \setminus \{j\}$. For $i > 1$ we have $A[i; \ell_1, \ldots, \ell_k] = 1$ if and only if $i = \ell_j$ for some $j \in [k]$ and either (1) $A[i-1; \ell_1, \ldots, \ell_{j-1}, \bot, \ell_{j+1}, \ldots, \ell_k] = 1$ and $f_j = i$ or (2) $A[i-1; \ell_1, \ldots, \ell_{j-1}, t, \ell_{j+1}, \ldots, \ell_k] = 1$ for some $t \neq \bot$ and $(v_{f_j}, v_t, v_{\ell_j})$ is single-crossing in the given order.

Here, case (1) corresponds to a partition of $[i]$ in which i is a singleton and case (2) corresponds to a partition where i is appended to an existing part. The correctness of this expression follows from Lemma 11. For each initial guess, our dynamic program has $O(n^{k+1})$ variables, and each variable can be computed in time $O(n^2m^2)$. This implies our bound on the running time. $\square$

Suppose we are given a profile that can be made single-crossing by rearranging the voters according to a permutation π. Then we can measure how good the input order is by the number of voter swaps performed by π.

Definition 7. *An instance of* VOTER REORDERING *is given by a profile* P *and an integer* k. *It is a yes-instance if we can perform at most* k *successive swaps of adjacent voters to make* P *single-crossing in the given order.*

This problem turns out to be polynomial-time solvable.

Proposition 14. VOTER REORDERING *can be solved in time* $O(nm^2 + n \log n)$.

6 Discussion and Conclusions

We have suggested a way of correlating voters' attributes with preferences, which in formal terms is the problem of deciding whether an input profile is close to being single-crossing in the given order. Let us close the paper by discussing technical connections to previous work on almost structured preferences. As with other papers on this topic, we have focused on computational issues, and have determined the problems' complexity for most notions of closeness that have been considered in the literature. Due to space constraints, we did not analyse closeness measures that are based on collapsing similar alternatives [Elkind *et al.*, 2012; Cornaz *et al.*, 2013]; it can be shown that the respective decision problems admit simple polynomial-time algorithms.

From the complexity perspective, our results for the fixed voter order essentially mirror the known results for deciding if a profile is nearly single-crossing up to a permutation: e.g., alternative deletion/partition is hard and voter deletion is easy in both settings, and our hardness results for global and local swaps can easily be extended to the setting where voters can be permuted. There are, however, a few cases where we obtained a positive algorithmic result for our setting, but it is not clear how to extend it to the setting where permutations are allowed; this is the case, e.g., for our FPT algorithm for GLOBAL SWAPS or our XP algorithm for INTERLEAVING VOTER PARTITION. On the other hand, there are no examples of problems that become computationally harder when the voter order is fixed.

In contrast, in the context of single-peaked preferences, knowing the axis reduces the complexity considerably, as shown by Erdélyi *et al.* [2017]. Thus, for nearly single-peaked preferences the main computational challenge is to determine the right axis, whereas the complexity of identifying nearly single-crossing preferences when the order of the voters is not known stems both from having to determine the order of voters and from figuring out the optimal modifications to the votes.

References

[Bartholdi III *et al.*, 1989] John Bartholdi III, Craig A. Tovey, and Michael A. Trick. Voting schemes for which it can be difficult to tell who won the election. *Social Choice and Welfare*, 6(2):157–165, 1989.

[Biedl *et al.*, 2009] Therese C. Biedl, Franz-Josef Brandenburg, and Xiaotie Deng. On the complexity of crossings in permutations. *Discrete Mathematics*, 309(7):1813–1823, 2009.

[Bredereck *et al.*, 2016] Robert Bredereck, Jiehua Chen, and Gerhard J. Woeginger. Are there any nicely structured preference profiles nearby? *Mathematical Social Sciences*, 79:61–73, 2016.

[Cohen *et al.*, 2019] Nathann Cohen, Edith Elkind, and Foram Lakhani. Single-crossing implementation, 2019. arXiv:1906.09671.

[Cornaz *et al.*, 2013] Denis Cornaz, Lucie Galand, and Olivier Spanjaard. Kemeny elections with bounded single-peaked or single-crossing width. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI)*, pages 76–82, 2013.

[de Vaus, 2002] David de Vaus. *Surveys in Social Research*. Taylor & Francis, 2002. 5th edition.

[Diestel, 2012] Reinhard Diestel. *Graph Theory*. Springer, 2012.

[Downey and Fellows, 2013] Rodney G. Downey and Michael R. Fellows. *Fundamentals of Parameterized Complexity*. Texts in Computer Science. Springer, 2013.

[Elkind *et al.*, 2012] Edith Elkind, Piotr Faliszewski, and Arkadii Slinko. Clone structures in voters' preferences. In *Proceedings of the 13th ACM Conference on Electronic Commerce (EC)*, pages 496–513. ACM, 2012.

[Elkind *et al.*, 2016] Edith Elkind, Martin Lackner, and Dominik Peters. Preference restrictions in computational social choice: Recent progress. In *Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 4062–4065, 2016.

[Elkind *et al.*, 2017] Edith Elkind, Martin Lackner, and Dominik Peters. Structured preferences. In U. Endriss, editor, *Trends in Computational Social Choice*, chapter 10, pages 187–207. AI Access, 2017.

[Erdélyi *et al.*, 2017] Gábor Erdélyi, Martin Lackner, and Andreas Pfandler. Computational aspects of nearly single-peaked electorates. *Journal of Artificial Intelligence Research*, 58:297–337, 2017.

[Faliszewski *et al.*, 2014] Piotr Faliszewski, Edith Hemaspaandra, and Lane A. Hemaspaandra. The complexity of manipulative attacks in nearly single-peaked electorates. *Artificial Intelligence*, 207:69–99, 2014.

[Garey *et al.*, 1976] Michael R. Garey, David S. Johnson, and Larry J. Stockmeyer. Some simplified NP-complete graph problems. *Theoretical Computer Science*, 1(3):237–267, 1976.

[Healey, 2008] Joseph F. Healey. *Statistics: A Tool for Social Research*. Cengage Learning, 2008. 8th edition.

[Jaeckle *et al.*, 2018] Florian Jaeckle, Dominik Peters, and Edith Elkind. On recognising nearly single-crossing preferences. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI)*, pages 1079–1086, 2018.

[Mirrlees, 1971] James A Mirrlees. An exploration in the theory of optimum income taxation. *The Review of Economic Studies*, 38(2):175–208, 1971.

[Skowron *et al.*, 2015] Piotr Skowron, Lan Yu, Piotr Faliszewski, and Edith Elkind. The complexity of fully proportional representation for single-crossing electorates. *Theoretical Computer Science*, 569:43–57, 2015.

Omitted Proofs

Proof of Theorem 2

Proof. The problem is clearly in NP. To prove NP-hardness, we reduce from Egalitarian Kemeny score. We map an instance $\langle P, s\rangle$ of Egalitarian Kemeny score with $P \in \mathcal{P}(A)$, $|A| = m$, to an instance $\langle tP, ts\rangle$ of Local Swaps, where $t = \binom{m}{2} + 1$.

If $\langle P, s\rangle$ is a yes-instance of Egalitarian Kemeny score, then there is a unanimous profile Q with $d_{\mathrm{egal}}(P, Q) \leqslant s$. Then tQ is unanimous and hence single-crossing in the given order, and $d_{\mathrm{egal}}(tP, tQ) \leqslant ts$. Hence $\langle tP, ts\rangle$ is a yes-instance of Local Swaps.

Conversely, suppose $\langle tP, t\rangle$ is a yes-instance of Local Swaps, and R is a profile that is single-crossing in the given order and satisfies $d_{\mathrm{egal}}(tP, R) \leqslant s$. As in the proof of Theorem 1, we partition R as $R_1 + \cdots + R_t$, and observe that there exists a $j \in [t]$ such that R_j contains no crossing points and is therefore unanimous. As we have $d_{\mathrm{egal}}(P, R_j) \leqslant d_{\mathrm{egal}}(tP, R) \leqslant s$, it follows that we have started with a yes-instance of Egalitarian Kemeny score. $\square$

Proof of Corollary 4

Proof. Recall that for linear orders v_i and v_j, we write $\Delta(v_i, v_j) = \{\{a, b\} \mid a >_i b,\ b >_j a\}$. We will use the following fact, which is easy to see: a profile $P = (v_1, \ldots, v_n)$ is single-crossing in the given order if and only if $\Delta(v_1, v_i) \subseteq \Delta(v_1, v_{i+1})$ for each $i = 1, \ldots, n-1$.

For a linear order v and an integer k, define the k-ball $B_k(v) = \{v' \in \mathcal{L} : d_K(v, v') \leqslant k\}$ as the set of linear orders obtainable from v in at most k swaps. Note that $|B_k(v)| = O(m^k)$.

Given a profile $P = (v_1, \ldots, v_n)$, our XP algorithm for local swaps works as follows. We first guess a vote $v_1' \in B_k(v_1)$ which will form the first vote in the profile after applying local swaps. Next, write $V_i = \{(i, v_i') : v_i' \in B_k(v_i)\}$ for $i = 2, \ldots, n$ and $V_1 = \{(1, v_1')\}$. We define a digraph $D = (V, A)$ with vertex set $V_1 \cup \cdots V_n$. For each $i \in [n-1]$, we add arcs between V_i and V_{i+1} as follows: there is an arc from (i, v_i') to $(i+1, v_{i+1}')$ whenever $\Delta(v_1', v_i') \subseteq \Delta(v_1', v_{i+1}')$. Clearly, D admits a topological order, and hence is acyclic. Thus, by a standard dynamic program, we can find a longest path in D in time polynomial in the size of the digraph. Our algorithm returns 'yes' if and only if the longest path found contains n vertices.

To show that the algorithm is correct, we claim that paths in D containing n vertices are in one-to-one correspondence with single-crossing profiles Q such that $d_{\mathrm{egal}}(P, Q) \leqslant k$ and whose first vote is v_1'. To see this, note that a path of length n has the form

$$(1, v_1') \to (2, v_2') \to \cdots \to (n, v_n').$$

This corresponds to a profile $Q = (v_1', v_2', \ldots, v_n')$ whose first vote is v_1'. By construction of the V_i's, we have $d_{\mathrm{egal}}(P, Q) \leqslant k$. Further, by construction of the arcs of the digraph D, we have $\Delta(v_1', v_i') \subseteq \Delta(v_1', v_{i+1}')$ for each $i = 1, \ldots, n-1$, and hence Q is single-crossing in the given order. $\square$

Proof of Theorem 8

Proof. We reduce from k-Colouring. Consider an instance $G = (A, L)$ of k-Colouring with $A = \{a_1, \ldots, a_m\}$, $L = (e_1, \ldots, e_s)$. Similarly to the proof of Theorem 7, for every edge $e = \{a_i, a_j\}$ we create two alternatives $c(e, i)$ and $c(e, j)$ and set $C^e = \{c(e, i), c(e, j)\}$; also, for each $a_i \in A$ we create k alternatives $d_i^1, \ldots, d_i^{k-1}$ and let $C_i = \{c(e, i) \mid a_i \in e\}$, $D_i = \{d_i^j \mid j \in [k-1]\}$.

We will now describe the profile P. As in the proof of Theorem 7, we first explain how voters rank various groups of alternatives, and then describe their rankings within these groups. At the level of blocks, this profile is identical to the one constructed in the proof of Theorem 7.

The first two votes in P are of the form

$$C^{e_1} > \cdots > C^{e_s} > D_1 > \cdots > D_m,$$

The third voter's ranking is of the form

$$D_1 >_3 C_1 >_3 \cdots >_3 D_m > C_m,$$

and the fourth voter's ranking is of the form

$$C_1 >_4 D_1 >_4 \ldots C_m >_4 D_m;$$

the order of alternatives within each block is described in the next paragraph.

For each $e = \{a_i, a_j\} \in L$ with $i < j$, set $c(e, i) >_1 c(e, j)$, $c(e, j) >_2 c(e, i)$. For each $i \in [m]$, voters 1, 3, and 4 rank the alternatives in D_i as $d_i^1 > \cdots > d_i^k$ and the voter 2 ranks the alternatives in D_i as $d_i^k > \cdots > d_i^1$. For each $i \in [m]$, voters 3 and 4 rank the alternatives in C_i so that $c(e_x, i)$ is ranked above $c(e_y, i)$ whenever $x < y$.

By construction, all pairs of alternatives within each set D_i are multi-crossing; also, each alternative in C_i is multi-crossing with each alternative in D_i. Similarly, for each edge e the alternatives in C^e are multi-crossing. No other pairs of alternatives are multi-crossing.

Suppose that G admits a k-colouring $\chi : A \to [k]$. We create a partition $A_1, \ldots, A_k$ as follows. For each $a_i \in A$ such that $\chi(i) = j$ we assign all alternatives in C_i to A_j. We then distribute the alternatives in D_i to other parts of the partition so that there is exactly one alternative from D_i in each of the parts $A_1, \ldots, A_{j-1}, A_{j+1}, \ldots, A_k$. This ensures that for every multi-crossing pair $\{a, b\}$ the alternatives a and b belong to different parts of the partition. Hence, $P|_{A_j}$ is single-crossing in the given order for each $j \in [k]$.

Conversely, suppose that A can be partitioned into k parts $A_1, \ldots, A_k$ so that for each $j \in [k]$ the profile $P|_{A_j}$ is single-crossing in the given order. The colouring $\chi : A \to [k]$ is defined as follows. First, we claim that for each $i \in [m]$ there exists a $j \in [k]$ such that $C_i \subseteq A_j$. Indeed, if $C_i \cap A_x \neq \varnothing$, $C_i \cap A_y \neq \varnothing$ for some $i \in [m]$ and $x \neq y$ then, as $|D_i| = k - 1$, at least one of the parts A_x, A_y must contain an alternative in D_i; as a result, the restriction of P to this part is not single-crossing in the given order, a contradiction. Hence, we can set $\chi(a_i) = j$ if $C_i \subseteq A_j$. Note that it cannot be the case that $\{a, b\} \in L$, but $\chi(a) = \chi(b) = j$, i.e., χ is a colouring of G, since otherwise $P|_{A_j}$ would not be single-crossing in the given order. This completes the proof. $\square$

Proof of Proposition 10

Proof. For each $i \in [n]$ and each $j = i, i+1, \ldots, n$, let

$$s(i,j) = \max\{|J| \mid \min J = i, \max J = j,$$
$$P_J \text{ is single-crossing in the given order}\}.$$

These quantities can be calculates as follows. For each $i \in [n]$ we have $s(i,i) = 1$ and

$$s(i,j) = \max\left\{s(i,\ell)+1 \mid i \leqslant \ell \leqslant j \text{ and } \Delta(v_i, v_\ell) \subseteq \Delta(v_i, v_j)\right\}.$$

We output $\max_{i \in [n], i \leqslant j \leqslant n} s(i,j)$. We have $O(n^2)$ quantities to compute, and each of them can be computed in time $O(nm^2)$. □

Full proof of Proposition 12

Proof. Let $P = (v_1, \ldots, v_n)$. We construct the profiles $P_1, \ldots, P_k$ greedily. Initially we set $j = 1$. We process the votes in P from left to right. When considering the vote v_i, we check whether $P_j + (v_i)$ is single-crossing in the given order (this can be done in time $O(m^2)$ by Lemma 11). If yes, we set $P_j := P_j + (v_i)$ and otherwise we set $j := j + 1$, $P_j = (v_i)$. We stop once all votes have been processed, and output the number of subprofiles in our partition.

To see that this results in an optimal solution, suppose for the sake of contradiction that our algorithm outputs a partition $(P_1, \ldots, P_k)$, but it is possible to partition P as $P = P'_1 + \cdots + P'_\ell$ for some $\ell < k$, where each P'_j is single-crossing in the given order. Among all such partitions, pick one that agrees with $(P_1, \ldots, P_k)$ on the longest possible prefix of $[n]$, and denote it by $(Q_1, \ldots, Q_\ell)$. Let $j = \min\{s \mid P_s \neq Q_s\}$. As $P_1 = Q_1, \ldots, P_{j-1} = Q_{j-1}$, the subprofiles P_j and Q_j have the same starting point, say, i; as both are contiguous subprofiles of P, they must end at different points. Let t be the last voter in P_j, and let t' be the last voter in Q_j. It cannot be the case that $t' > t$, since $(v_i, \ldots, v_{t+1})$ is not single-crossing in the given order by construction of P_j. Thus, $t' < t$. But then we can move $v_{t'+1}$ from Q_{j+1} to Q_j: $Q_j + \{v_{t'+1}\}$ remains single-crossing in the given order, and the resulting partition has at most ℓ parts and agrees with $(P_1, \ldots, P_k)$ on a longer prefix of $[n]$ than $(Q_1, \ldots, Q_\ell)$ does. This is a contradiction with our choice of $(Q_1, \ldots, Q_\ell)$. □

Proof of Proposition 14

Proof. We first check if P is single-crossing up to a permutation; if not, we reject P immediately. Now, suppose this condition is satisfied. It is known that if P is single-crossing up to a permutation, then there exists a unique ordering of voters in $[n]$, up to permuting voters with identical preferences and reversing the entire voter list, which makes P single-crossing in the given order [**?**]; further, some such ordering can be computed in time $O(nm^2)$ [**?**]. Let Q be the profile that corresponds to some such ordering, and let Q' be the profile obtained by reversing the order of voters in Q. We view P, Q and Q' as ordered lists of votes and compute the Kendall tau distance between P and Q (note that this differs from $d_K(P, Q)$, as defined earlier in the paper) and between P and Q'; the Kendall tau distance between two lists of length n can be computed in time $O(n \log n)$ [**?**]. We output 'yes' if one of these distances is at most k. □