How Similar Are Two Elections?

Piotr Faliszewski,[1] **Piotr Skowron,**[2] **Arkadii Slinko,**[3] **Stanisław Szufa,**[4] **Nimrod Talmon**[5]

[1]AGH University, Kraków, Poland, [2]University of Warsaw, Poland, [3]University of Auckland, New Zealand
[4]Jagiellonian University, Kraków, Poland, [5]Ben-Gurion University, Be'er Sheva, Israel
faliszew@agh.edu.pl, p.skowron@mimuw.edu.pl, a.slinko@auckland.ac.nz,
stanislaw.szufa@doctoral.uj.edu.pl, talmonn@bgu.ac.il

Abstract

We introduce the ELECTION ISOMORPHISM problem and a family of its approximate variants, which we refer to as d-ISOMORPHISM DISTANCE (d-ID) problems (where d is a metric between preference orders). We show that ELECTION ISOMORPHISM is polynomial-time solvable, and that the d-ISOMORPHISM DISTANCE problems generalize various classic rank-aggregation methods (e.g., those of Kemeny and Litvak). We establish the complexity of our problems (including their inapproximability) and provide initial experiments regarding the ability to solve them in practice.

1 Introduction

We introduce the ELECTION ISOMORPHISM problem and a family of its approximate variants, which measure the degree of similarity between two elections by using distances over preference orders. We study the complexity of problems that emerge from this idea and seek efficient algorithms for them.

We consider the ordinal model of elections, where each voter submits a preference order that ranks all the candidates from the most to the least desirable one. In the ELECTION ISOMORPHISM problem we are given two elections, E and E', both with the same numbers of candidates and the same numbers of voters, and we ask if it is possible to transform one into the other by renaming candidates and reordering voters. While this problem is similar in spirit to the famous GRAPH ISOMORPHISM problem (whose complexity status remains elusive; see the report of Babai et al. (2015) and further discussion on Babai's home page (2017) for recent progress on the problem), the structure of elections with ordinal ballots is such that it is very easy to provide a polynomial-time algorithm for ELECTION ISOMORPHISM. On the other hand, for approval-based elections ELECTION ISOMORPHISM is at least as hard as GRAPH ISOMORPHISM (a graph can be encoded as an approval election in a simple way; we consider ordinal elections only).

We are also interested in approximate variants of the ELECTION ISOMORPHISM problem, which turn out to define distances over elections. In the distance rationalizability framework, the distances between elections are obtained by extending distances over preference orders to elections in a

votewise manner (Elkind, Faliszewski, and Slinko 2012), assuming that both elections include the same candidates and the same voters (even if the voters have changed their preference orders in one of the elections). In contrast to that line of work, we extend the distance between preference orders to whole elections in a way that respects both anonymity and neutrality. Namely, we ask if, via appropriate renaming of candidates and reordering voters, it is possible to bring a given election within some small votewise distance of another given one. We will refer to these problems as the d-ISOMORPHISM DISTANCE problems, where d is the distance (over preference orders) that we use.

As a consequence, when d is the swap distance, d-ISOMORPHISM DISTANCE generalizes the problem of finding a Kemeny ranking (roughly speaking, finding a Kemeny ranking is equivalent to finding the smallest swap-based isomorphism distance between the given election and a "constant" one, where all the votes are identical). We note that approximate GRAPH ISOMORPHISM problems are also studied in the literature (Arvind et al. 2012; Grohe, Rattan, and Woeginger 2018); in spirit, they are very similar to our problems, but they differ on the technical level.

Motivation. Formalizing the concept of an isomorphism between elections allows us to classify domains of preference orders. In particular, for a given number of candidates, all maximal single-peaked domains are isomorphic, whereas this is not always the case for maximal single-crossing domains (see Section 3). Furthermore, the d-ISOMORPHISM DISTANCE problems allow for a principled way of performing classification, clustering, and other common tasks related to distances, over the space of elections. We give three more specific domains of application:

Choosing Elections to Evaluate Algorithms On. There is a growing body of work in computational social choice that provides experimental analyses of elections (see, e.g., the works of Narodytska and Walsh (2014), Aziz et al. (2015), Caragiannis et al. (2017), and BrederECK et al. (2017)), often using the PrefLib database (Mattei and Walsh 2013). Designing convincing experiments, however, requires care. For example, we should not simply use all the elections available in PrefLib, because the results would be biased toward the type of elections that happens to be most popular there. Currently researchers often

rely on the annotations available in PrefLib, which cluster elections regarding their origin. Thus, we may wish to verify if elections with the same origin are, indeed, similar to each other or not. Alternatively, instead of choosing elections based on their origins, we may choose a set of elections that are sufficiently different from each other. In either case, our distances may be helpful.

Analyzing Real-Life Elections. Given an election (e.g., from PrefLib), it is interesting to ask if it is similar to those that come from some distribution (such as, e.g., the impartial culture model, where each preference order is equally likely; or one of the Euclidean models, where candidates and voters are mapped to points in some Euclidean space, and the voters rank the candidates with respect to the distance from their points (Enelow and Hinich 1984; 1990)). To this end, we may generate a number of elections according to a given distribution and compute their isomorphism distances from the given one.

Comparing Election Distributions. Given several samples from two distributions over elections, we may ask how similar these distributions are to each other. For example, we may wonder how different are two models of elections with Euclidean preferences where in one model voters and candidates are distributed uniformly on a disc and in another they are distributed according to a Gaussian distribution.[1] While comparing real-valued distributions is a classic topic within statistics, doing the same for elections seems much harder, and we are not aware of any good solution. Yet, intuitively, the ability to measure the isomorphism distance between samples of elections would be helpful in designing an appropriate method (indeed, e.g., the classic χ^2 test relies on comparing frequencies of *similar* items in both distributions).

We stress that we do not address the above problems directly, but, rather, we claim that the d-ISOMORPHISM DISTANCE problem is an important tool for tackling them. Thus, our focus is on establishing the complexity of this problem and on designing practically relevant heuristics.

Our Contribution. Our main contributions are as follows:

1. We show that the ELECTION ISOMORPHISM problem is in P, and, using similar ideas, we show that the same is true for the ISOMORPHISM DISTANCE problem for the case of the discrete distance (but we do not expect this variant of the problem to be very relevant in practice).

2. We show that the ISOMORPHISM DISTANCE problem is NP-complete both for the swap distance and for the Spearman distance. In the former case, we inherit hardness from the Kemeny rule, but in the latter we use a new proof. We also provide parametrized complexity results for both problems.

3. We describe an ILP formulation of the d_{Spear}-ISOMORPHISM DISTANCE problem. We evaluate experimentally on what sizes of elections the state-of-the-art

---

[1]Such a comparison would be useful, e.g., in the work of Elkind et al. (2017), where the authors evaluated multiwinner voting rules on four Euclidean models of elections.

ILP solvers are capable of solving our ILP. We also propose a heuristic algorithm and evaluate its performance.

We plan to use our tools to address some of the issues described above, such as analyzing elections within PrefLib.

2 Preliminaries

For an integer n, we write $[n]$ to denote the set $\{1, \ldots, n\}$. By S_n, we mean the set of all permutations over $[n]$.

Preference Orders. Let C be a set of candidates. We refer to a linear order over C as a *preference order* (ranking the candidates from the most to the least appealing one); we also sometimes call it a *vote*. For a vote v, we write $v\colon a \succ b$ to indicate that candidate a is preferred to b under v; we write $\mathrm{pos}_v(c)$ to denote the position of candidate c in v (the top ranked candidate has position 1, the next one 2, and so on).

Domains of Preference Orders. We write $\mathcal{L}(C)$ to denote the set of all preference orders over C. Every subset $\mathcal{D}$ of $\mathcal{L}(C)$ is called a domain (of preference orders over C) and, in particular, $\mathcal{L}(C)$ itself is the general domain. Later we will consider the single-peaked and single-crossing domains.

Elections. An election $E = (C, V)$ consists of a set of candidates $C = \{c_1, \ldots, c_m\}$ and a collection of voters $V = (v_1, \ldots, v_n)$, where each voter v_i has a preference order, also denoted as v_i (the exact meaning will always be clear from the context and this convention will simplify our discussions). The preference orders always come from some domain $\mathcal{D}$ (the general domain unless stated otherwise).

Distances. Formally, for a set X a function $d\colon X \times X \to \mathbb{R}$ is a metric if for each $x, y, z \in X$ it holds that (i) $d(x, y) \geq 0$, (ii) $d(x, y) = 0$ if and only if $x = y$, (iii) $d(x, y) = d(y, x)$, and (iv) $d(x, z) \leq d(x, y) + d(y, z)$. A pseudometric relaxes condition (ii) to the requirement that $d(x, x) = 0$ for each $x \in X$ (in particular, for a pseudometric d it is possible that $d(x, y) = 0$ when $x \neq y$).

We focus on the following three distances between preference orders (below, let C be a set of candidates and let u and v be two preference orders from $\mathcal{L}(C)$):

Discrete Distance. The discrete distance between u and v, $d_{\mathrm{disc}}(u, v)$, is 0 when u and v coincide and is 1 otherwise.

Swap Distance. The swap distance between u and v (also known as Kendall's Tau distance in statistics), denoted $d_{\mathrm{swap}}(u, v)$, is the smallest number of swaps of consecutive candidates that need to be performed within u to transform it into v.

Spearman Distance. The Spearman's distance (known also as the Spearman's footrule or the displacement distance) measures the total displacement of candidates in u relative to their positions in v. Formally, it is defined as:

$$d_{\mathrm{Spear}}(u, v) = \sum_{c \in C} |\mathrm{pos}_v(c) - \mathrm{pos}_u(c)|.$$

We only consider distances over preference orders that are defined for all sets of candidates (as is the case for d_{disc},

d_{swap}, and d_{Spear}). For an in-depth discussion regarding distances between elections, we point to the literature on distance rationalizability of voting rules (Nitzan 1981; Meskanen and Nurmi 2008; Elkind, Faliszewski, and Slinko 2015) and, in particular, to the survey of Elkind and Slinko (2016).

Bijections Between Candidate Sets. Consider two sets of candidates, C and D, of the same cardinality. Let σ be a bijection from C to D. We extend σ to act on preference orders v in $\mathcal{L}(C)$ in the natural way: $\sigma(v) \in \mathcal{L}(D)$ is the preference order such that for each $c, c' \in C$ it holds that $v\colon c \succ c' \iff \sigma(v)\colon \sigma(c) \succ \sigma(c')$. For an election $E = (C, V)$, where $V = (v_1, \dots, v_n)$, a candidate set D, and a bijection σ from C to D, by $\sigma(E)$ we mean election with candidate set D and voter collection $(\sigma(v_1), \dots, \sigma(v_n))$.

3 Election Isomorphism

In this section we define the notion of election isomorphism, illustrate its usefulness, and show that testing if two elections are isomorphic is a polynomial-time computable task. We start with a formal definition of election isomorphism.

Definition 1. *We say that elections $E = (C, V)$ and $E' = (C', V')$, where $|C| = |C'|$, $V = (v_1, \dots, v_n)$, and $V' = (v'_1, \dots, v'_n)$, are isomorphic if there is a bijection $\sigma\colon C \to C'$ and a permutation $\nu \in S_n$ such that $\sigma(v_i) = v'_{\nu(i)}$ for all $i \in [n]$.*

Example 1. *Consider elections $E = (C, V)$ and $E' = (C', V')$, such that $C = \{a, b, c\}$, $C' = \{x, y, z\}$, $V = (v_1, v_2, v_3)$, $V' = (v'_1, v'_2, v'_3)$, with preference orders:*

$$v_1\colon a \succ b \succ c, \qquad v_2\colon b \succ a \succ c, \qquad v_3\colon c \succ a \succ b,$$
$$v'_1\colon y \succ x \succ z, \qquad v'_2\colon x \succ y \succ z, \qquad v'_3\colon z \succ x \succ y.$$

E and E' are isomorphic, by mapping candidates a to x, b to y, and c to z, and voters v_1 to v'_2, v_2 to v'_1, and v_3 to v'_3.

The idea of election isomorphism has already appeared in the literature, though without using this name and usually as a tool to achieve some specific goal. For example, Eğecioğlu and Giritligil (2013) refer to two isomorphic elections as members of the same *anonymous and neutral equivalence class (ANEC)* and study the problem of sampling representatives of ANECs uniformly at random. Hashemi and Endriss (2014) use the election isomorphism idea in their analysis of preference diversity indices.

In the ELECTION ISOMORPHISM problem we are given two elections and we ask if they are isomorphic. Surprisingly, the problem has an easy polynomial-time algorithm.

Proposition 1. ELECTION ISOMORPHISM *is in* P.

Proof. Let $E = (C, V)$ and $E' = (C', V')$ be two input elections where $C = \{c_1, \dots, c_m\}$, $C' = \{c'_1, \dots, c'_m\}$, $V = (v_1, \dots, v_n)$ and $V = (v'_1, \dots, v'_n)$. Without loss of generality, let us assume that v_1's preference order is $v_1\colon c_1 \succ c_2 \succ \dots \succ c_m$. For each v'_j there is a bijection σ_j from C to C' such that for the preference order of v'_j we have $\text{pos}_{v'_j}(\sigma_j(c_i)) = i$. For each σ_j, we build a bipartite graph where $v_1, \dots, v_n$ are the vertices on the left, $v'_1, \dots, v'_n$ are the vertices on the right, and there is an edge

between v_i and v_ℓ if $\sigma_j(v_i) = v_\ell$; we accept if this graph has a perfect matching for some σ_j and we reject otherwise.

The algorithm runs in polynomial time because there are n σ_j's to try, and computing perfect matchings is a polynomial-time computable task. Correctness follows from the fact that we need to map v_1 to some vote in E' and we try all possibilities. $\square$

As an extended example of usefulness of the isomorphism idea, we consider the single-peaked (Black 1958) and single-crossing (Mirrlees 1971; Roberts 1977) domains. They received extensive attention within (computational) social choice; we point the readers to the survey of Elkind et al. (2016) for more details.

Definition 2. *Let $\mathcal{D} \subseteq \mathcal{L}(C)$ be a domain.*

1. *$\mathcal{D}$ is single-crossing if its preference orders can be ordered as $v_1, \dots, v_n$ so that, for each pair of candidates $a, b \in C$, as we consider $v_1, \dots, v_n$ in this order, the relative ranking of a and b changes at most once.*
2. *$\mathcal{D}$ is single-peaked if there exists a linear order $>$ over C (referred to as the societal axis) such that for each v and each $j \in [\|C\|]$, the top j candidates from v form an interval according to the order $>$.*

A single-peaked (single-crossing) domain is maximal if it is not contained in any other single-peaked (single-crossing) domain. Each maximal single-peaked domain $\mathcal{D} \subseteq \mathcal{L}(C)$ contains $2^{|C|-1}$ preference orders (Monjardet (2009) attributes this fact to a 1962 work of Kreweras).

Since we can view a domain as an election that includes a single copy of every preference order from the domain, our notion of isomorphism directly translates to the case of domains, and we can formalize a fundamental difference between single-peakedness and single-crossingness.

Proposition 2. *Each two maximal single-peaked domains over candidate sets of the same size are isomorphic. There are two maximal single-crossing domains over the same set of candidates that are not isomorphic.*

Sketch of the proof. For the first part of the proposition, it suffices to note that if $\mathcal{D}$ and $\mathcal{D}'$ are two maximal single-peaked domains (over candidate sets $\{x_1, \dots, x_m\}$ and $\{y_1, \dots, y_m\}$, respectively), with axes $>_1$ and $>_2$, such that:

$$x_1 >_1 \dots >_1 x_m \qquad \text{and} \qquad y_1 >_2 \dots >_2 y_m,$$

then a bijection that maps each x_i to y_i witnesses that the two domains are isomorphic.

For the second part of the proposition, consider the following two single-crossing domains of four candidates (each preference order is shown as a column, with the first candidate on top and the last one on the bottom):

a	b	b	b	b	d	d		d	d	d	a	a	a	a
b	a	c	c	d	b	c		c	c	a	d	c	c	b
c	c	a	d	c	c	b		b	a	c	c	d	b	c
d	d	d	a	a	a	a		a	b	b	b	b	d	d

Both are maximal single-crossing domains and both are maximal Condorcet domains (Puppe and Slinko 2017). They are not isomorphic as the first one has three different candidates on top, and the second one has two. $\square$

Given this result, it is very natural to ask, e.g., how many non-isomorphic maximal single-crossing domains exist for a particular number of alternatives m. We recommend this issue for future research.

4   Isomorphism Distance

In this section we use the isomorphism idea to build distances between elections that respect both voter anonymity (so the order of the voters in an election is irrelevant) and candidate neutrality (so the names of the candidates are nothing more than temporary identifiers). Below we give our main definition (for two sets A, B of the same cardinality, by $\Pi(A, B)$ we mean the set of bijections from A to B).

Definition 3. *Let d be a distance between preference orders. Let $E = (C, V)$ and $E' = (C', V')$ be two elections, where $|C| = |C'|$, $V = (v_1, \ldots, v_n)$ and $V' = (v'_1, \ldots, v'_n)$. We define the d-isomorphism distance between E and E' to be:*

$$d\text{-ID}(E, E') = \min_{\nu \in S_n} \min_{\sigma \in \Pi(C, C')} \sum_{i=1}^{n} d(\sigma(v_i), v'_{\nu(i)})$$

We sometimes refer to the bijection σ as the candidate matching and to the permutation ν as the voter matching, and sometimes instead of ν, we use bijection $\tau \in \Pi(V, V')$ (depending on what is more convenient). The name, d-isomorphism distance, is justified by the fact that if $d\text{-ID}(E, E') = 0$ for some two elections (and d is a metric over preference orders), then these elections are isomorphic.

Remark 1. *Note that in the above definition we view elections as both anonymous and neutral. This is why we apply the minimum operator over all permutations of the voters and over all bijections between the candidates.*

Remark 2. *Some of the scenarios from the introduction require the ability of computing distances between elections with different numbers of candidates and voters, but our definition does not allow this directly. There are many ways of computing such distances, using our isomorphism distances as tools, but we leave the study of this issue for the future.*

Basic Properties

Formally, the d-isomorphism distances are pseudometrics over the space of elections with given numbers of candidates and voters. Indeed, it is immediate to see that for every election $E = (C, V)$ we have that $d\text{-ID}(E, E) = 0$, and for each two elections E and E' we have $d\text{-ID}(E, E') = d\text{-ID}(E', E)$. The triangle inequality requires some care.

We say that a distance d over preference orders is *permutation-invariant* if for each two candidate sets C and D of the same cardinality, each two preference orders $u, v \in \mathcal{L}(C)$, and each bijection σ from C to D, it holds that $d(u, v) = d(\sigma(u), \sigma(v))$.

Proposition 3. *For each permutation-invariant distance d over preference orders, d-ID satisfies the triangle inequality.*

The d-isomorphism distances inherit some properties from the underlying distances. For example, the Diaconis–Graham inequality says that for two preference orders u and

v we have $d_{\text{swap}}(u, v) \leq d_{\text{Spear}}(u, v) \leq 2 \cdot d_{\text{swap}}(u, v)$ (Diaconis and Graham 1977), and one can verify that the same holds for the isomorphism distances. Specifically, for each two elections E and E' (with the same numbers of candidates and voters) we have $d_{\text{swap}}\text{-ID}(E, E') \leq d_{\text{Spear}}\text{-ID}(E, E') \leq 2 \cdot d_{\text{swap}}\text{-ID}(E, E')$.

Complexity of Computing Isomorphism Distances

We now turn to the complexity of computing ismorphism distances. Formally, our problem is defined as follows.

Definition 4. *Let d be a distance over preference orders. In the d-ISOMORPHISM DISTANCE problem (the d-ID problem) we are given two elections, $E = (C, V)$ and $E' = (C', V')$ such that $|C| = |C'|$ and $|V| = |V'|$, and an integer k. We ask if $d\text{-ID}(E, E') \leq k$.*

We are also interested in two variants of this problem, the d-ID WITH CANDIDATE MATCHING problem, where the bijection σ between the candidate sets is given (and fixed), and the d-ID WITH VOTER MATCHING problem, where the voter permutation ν is given (and fixed). The former problem is in P for polynomial-time computable distances, but, as we will see later, this is not always true for the latter.

Proposition 4. *For a polynomial-time computable d, the problem d-ID WITH CANDIDATE MATCHING is in P.*

Proof sketch. Let E and E' be our input elections and let σ be the input matching between candidates from E and E'. It suffices to compute a distance between every pair of votes (one from $\sigma(E)$ and another from E'), build a corresponding bipartite graph (where vertices on the left are the voters from $\sigma(E)$, the vertices on the right are the voters from E', and all possible edges exist, weighted by the distances between the votes they connect), and find the smallest-weight matching (the weight of the matching gives the value of the distance, and the matching itself gives the permutation ν). $\square$

Using an argument very similar to that in the proof of Proposition 1, we show that d_{disc}-ID is in P.

Proposition 5. *The d_{disc}-ID problem is in P.*

The elections for which the d_{disc}-ID distance is small are, in fact, nearly identical (up to renaming of the candidates and reordering the voters). In consequence, we do not expect such elections to frequently appear in the applications that we mentioned in the introduction (for example, for two elections with n voters and a relatively large number of candidates, generated according to the impartial culture model, we would expect their d_{disc}-ID distance to typically be $n - 1$). Thus, we need more fine-grained distances, such as d_{swap}-ID and d_{Spear}-ID. Unfortunately, they are NP-hard to compute and, indeed, for d_{swap}-ID we inherit this result from the Kemeny rule.

Proposition 6. *The d_{swap}-ID problem is NP-complete, even for elections with four voters.*

Proof. Membership in NP- is easy to see. We give a reduction from the KEMENY SCORE problem. In the KEMENY SCORE problem we are given an election $E = (C, V)$ and an integer k, and we ask if there exists a preference order

p over C such that $\sum_{v \in V} d_{\text{swap}}(v, p) \leq k$. The problem is NP-complete (Bartholdi, Tovey, and Trick 1989) and remains NP-complete even for the case of four voters (Dwork et al. 2001). We reduce it to the d_{swap}-ID problem in a straightforward way: Given election $E = (C, V)$ and k, our reduction outputs election E, a newly constructed election $E' = (C', V')$, and an integer k, where $C' = \{c'_1, \ldots, c'_{|C|}\}$ and every voter in V' has identical preference order $v' : c'_1 \succ \cdots \succ c'_{|C|}$.

The reduction runs in polynomial time. The correctness follows directly from the definitions of the problems. $\quad\square$

Since the above reduction works even for elections with four voters, having a matching between the voters cannot make the problem simpler (this also follows from the fact that in our reduction one election consists of identical votes).

Corollary 1. d_{swap}-ID WITH VOTER MATCHING *is* NP-*complete.*

The situation for d_{Spear}-ID is somewhat different. In this case Litvak's rule (Litvak 1983), defined analogously to the Kemeny rule, but for the Spearman distance, is polynomial-time computable (Dwork et al. 2001) and we can lift this result to the case of d_{swap}-ID WITH VOTER MATCHING. Without the voter matching, d_{Spear}-ID is NP-complete.

Proposition 7. d_{Spear}-ID w/ VOTER MATCHING *is in* P.

Theorem 8. *The* d_{Spear}-ID *problem is* NP-*complete.*

Hardness of Approximation

Unfortunately, we cannot hope for good polynomial-time approximation algorithms for our problems, unless GRAPH ISOMORPHISM is in P, which seems unlikely. (In the discussion below, when we speak of d_{swap}-ID and d_{Spear}-ID we mean the optimization variants of these problems where instead of deciding if the distance between two given elections is at most a given value, we ask to compute the distance between these elections.)

Theorem 9. *For each* $\alpha < 1$, *there is no polynomial-time* $|C|^{\alpha}$-*approximation algorithm neither for* d_{Spear}-ID *nor for* d_{swap}-ID, *unless the* GRAPH ISOMORPHISM *problem is in* P.

Proof. Let us focus on the Spearman distance—the same construction works for the swap distance. Towards a contradiction, let us assume that there exists a $|C|^{\alpha}$-approximation algorithm $\mathcal{A}$ for d_{Spear}-ID. We will show that $\mathcal{A}$ can be used to solve GRAPH ISOMORPHISM.

Let I be an instance of the GRAPH ISOMORPHISM problem; I consists of two undirected graphs $G = (Vtx, Edg)$ and $G' = (Vtx', Edg')$. We use a non-standard definition of GRAPH ISOMORPHISM and assume that in I we ask whether there exist *two* bijections,[2] $\sigma \colon Vtx \to Vtx'$ and $\tau \colon Edg \to Edg'$, such that for each $edg = \{vtx_1, vtx_2\} \in Edg$ it holds that $\tau(edg) = \{\sigma(vtx_1), \sigma(vtx_2)\}$. Without loss of

generality, let us assume that $|Vtx| = |Vtx'| = n$ and $|Edg| = |Edg'| = m$, and that there are no isolated vertices in any of the two graphs.

From I we construct an instance I_{ID} of d_{Spear}-ID that consists of two elections, $E = (C, V)$ and $E' = (C', V')$, as follows. The voters in V and V' correspond to the edges from Edg and Edg', respectively. Further, for each vertex $vtx \in Vtx$ we add one candidate to C (denoted by the same symbol), and for each $vtx' \in Vtx'$ we add a candidate to C'. Additionally, we add $L = \lceil (2nm^2)^{\frac{1}{1-\alpha}} \rceil$ dummy candidates $D = \{d_1, \ldots, d_L\}$ to C and L dummy candidates $D' = \{d'_1, \ldots, d'_L\}$ to C'. In consequence, our elections have $n + L$ candidates and m voters each. Finally, we describe the preferences of the voters. Consider an edge $edg = \{vtx_1, vtx_2\}$; the voter corresponding to edg ranks vtx_1 and vtx_2 on top in some arbitrary order, next all the dummy candidates from D in the increasing order of their indices ($d_1 \succ d_2 \succ \cdots \succ d_L$) and then all other candidates from $Vtx \setminus \{vtx_1, vtx_2\}$ in some fixed arbitrary order. We construct the rankings of the voters from V' analogously.

Now, observe that if there exists an isomorphism (σ, τ) between the two graphs in the original instance, then the isomorphism distance between E and E' is at most nm^2. Indeed, it is apparent that this distance is witnessed by the bijection between voters σ and by the function renaming the candidates τ combined with the mapping τ_D such that $\tau_D(d_i) = d'_i$ for each $d_i \in D$. Consequently, our approximation algorithm $\mathcal{A}$ would find a distance that is lower than:

$$nm^2(L + m)^{\alpha} \leq nm^2 \left(1 + \frac{m}{L}\right)^{\alpha} L^{\alpha} < 2nm^2 L^{\alpha-1} L$$
$$\leq 2nm^2(2nm^2)^{\frac{1}{1-\alpha} \cdot (\alpha-1)} L = L.$$

Next, we assess the isomorphism distance in case when the answer to the original instance I is "no". Let σ and τ be two bijections between the sets of candidates and voters, respectively. First, let us analyze what happens when $\sigma(d_i) \notin D'$ for some $d_i \in D$. Since d_i is always ranked on the same position, and each vertex candidate appears at least once in the first two positions, and at least once in the last $m - 2$ positions, we infer that the isomorphism distance between E and E' is at least equal to L. Next, we move to the case when $\sigma(d_i) \in D'$ for each $d_i \in D$. Since σ and τ truncated to the vertex-candidates do not witness an isomorhism, it must be the case that there exists a voter v and a candidate c such that c is ranked in the top two positions by v, yet $\sigma(v)$ ranks $\tau(c)$ in the last $m - 2$ position, which witnesses that the isomorphism distance is at least equal to L.

Thus, $\mathcal{A}$ finds that the isomorphism distance between E and E' is lower or equal to L if and only if the answer to the original nstance I is "yes". This completes the proof. $\quad\square$

The next results follows by a similar reduction.

Theorem 10. *For each* $\alpha < 1$ *there is no polynomial-time* $|V|^{\alpha}$-*approximation algorithm neither for* d_{Spear}-ID *nor for* d_{swap}-ID, *unless* GRAPH ISOMORPHISM *is in* P.

While an m^2-approximation algorithm (which boils down to using an arbitrary matching between the candidates) exists, the existence of an $O(m)$-approximation or an $O(n)$-approximation for either of our distances remains open.

---

[2]Typically, one asks for a bijection between the vertices only, and it is assumed that the other bijection (between the edges) is induced by the first one. In our case it will be easier to work with two separate bijections.

d	d-ID	WITH VOTER MATCHING	WITH CANDIDATE MATCHING	parameter		
				m	n	k
d_{disc}	P	P	P	–	–	–
d_{Spear}	NP-complete	P	P	FPT	FPT	FPT
d_{swap}	NP-complete	NP-complete	P	FPT	para-NP-hard	FPT

Table 1: The complexity of computing isomorphism distances, also parametrized by the number of candidates m, the number of voters n, and the distance k.

Fixed Parameter Tractability

As d_{swap}-ID and d_{Spear}-ID are both NP-complete and hard to approximate, one further hope for their theoretical tractability lays within parametrized complexity theory. Next we show that such hope is, at least partially, not in vain.

For the parameterization by the number of candidates, it suffices to use simple brute-force algorithms (we guess the matching between the candidates and invoke Proposition 4).

Observation 1. *Both d_{swap}-ID and d_{Spear}-ID are in* FPT *for parametrization by the number m of candidates.*

For the parametrization by the number of voters, d_{swap}-ID is para-NP-hard (i.e., is NP-hard even for some constant number of voters) due to Proposition 6. For d_{Spear}-ID, Proposition 7 implies an FPT algorithm.

Observation 2. *d_{Spear}-ID is in* FPT *for the parametrization by the number of voters.*

Next, we consider the distance value k as the parameter, which in the FPT jargon would be referred to as the *natural parameter*. It turns out that both for the swap distance and the Spearman distance, we have FPT algorithms.

Proposition 11. *d_{Spear}-ID is in* FPT *for the parametrization by the distance value k.*

Proof. Let $E = (C, V)$ and $E' = (C', V')$ be two elections with $|C| = |C'|$ and $|V_1| = |V_2| = n$. If $k \geq n$ then the result follows from Observation 2, so we assume that $k < n$.

If $\sigma \colon C_1 \to C_2$ and $\tau \colon V_1 \to V_2$ witness the fact that d_{Spear}-ID$(E_1, E_2) \leq k$, then there is at least one voter v from V_1 such that $d_{\mathrm{Spear}}(\sigma(v), \tau(v)) = 0$. Thus we guess a voter v from V_1 and a voter v' from V_2 and compute the permutation σ which makes votes $\sigma(v)$ and v' the same. We then invoke Proposition 4 to test if this σ indeed leads to distance at most k between the elections. $\square$

Theorem 12. *d_{swap}-ID is in* FPT *for the parametrization by the distance value k.*

Proof. Let $E = (C, V)$ and $E' = (C', V')$ be our input elections, where $|C| = |C'|$, $V = (v_1, \ldots, v_n)$, and $V' = (v'_1, \ldots, v'_n)$. Our goal is to check if d_{swap}-ID(E, E') is at most k. We treat the case where $k < n$ as in Proposition 11 and, thus, we assume that $k \geq n$.

We proceed by guessing a matching $\nu \in S_n$ between the voters that witnesses that d_{swap}-ID$(E, E') \leq k$. With respect to this matching, we say that a candidate $c \in C$ is *happy* if there is a candidate $c' \in C'$ such that for every vote $v_i \in V$, we have that c is ranked on the same position in v_i

as c' is in $v_{\nu(i)}$ (i.e., $\mathrm{pos}_{v_i}(c) = \mathrm{pos}_{v_{\nu(i)}}(c')$). If a candidate is not happy, then we say that he or she is *sad*. Note that if the distance between E and E' is at most k, then there are no more than $2k$ sad candidates (indeed, each swap of two neighbouring candidates, in any of the votes from V, can make at most two sad candidates happy, and all the candidates must be happy after at most k swaps).

Given ν, our algorithm performs at most k iterations, where in each iteration we execute the following steps:

1. We compute the set of candidates that are currently sad. We accept if there are no sad candidates and we reject if there is more than $2k$ of them.

2. We guess a voter v_i from V, a sad candidate d, a candidate c such that $|\mathrm{pos}_{v_i}(c) - \mathrm{pos}_{v_i}(d)| \leq k$, and whether we swap c with the candidate right before or right after him or her. Then we perform this swap of c within v_i.

We accept if there is a sequence of guesses such that after at most k iterations there are no sad candidates, and we reject otherwise. Correctness follows from the fact that there is always a solution that starts by swapping a candidate that is within distance k swaps of a sad one (we omit the somewhat involved argument as to why this holds due to space restriction). Fixed-parameter tractability follows because we perform k iterations and in each we have at most $n \cdot 2k \cdot (2k + 1) \cdot 2$ possible guesses (corresponding to a voter, a sad candidate, distance from this sad candidate, and the direction of the swap), which is at most $8k^3 + 4k^2$. Altogether, the superpolynomial factor of the running time is $O^*(k!(8k^3 + 4k^2)^k) = O^*(k^{5k})$, where the $k!$ factor comes from guessing the matching between the voters. $\square$

5 Computing d-ID Distances in Practice

In spite of the intractability results of the previous section, it is important to be able to compute isomorphism distances in practice. Due to restricted space, we focus on the Spearman distance (on the one hand, it is simpler to deal with and, on the other hand, we know that its values are 2-approximations of the swap distance ones, so the nature of these two distances is similar). First, we provide an integer linear program (ILP) for d_{Spear}-ID.

Proposition 13. *There is an ILP for d_{Spear}-ID.*

Proof. Let $E = (C, V)$ and $E' = (C', V')$ be the elections we wish to compute the distance for, with $C = \{c_1, \ldots, c_m\}$, $C' = \{c'_1, \ldots, c'_m\}$, $V = \{v_1, \ldots, v_n\}$, and $V' = \{v'_1, \ldots, v'_n\}$. For each $k, \ell \in [n]$, we define a binary variable $N_{k,\ell}$ with the intention that value 1 indicates

that voter v_k is matched to voter v'_ℓ. Similarly, for each $i, j \in [m]$, we define a binary variable $M_{i,j}$ with the intention that value 1 means that candidate c_i is matched to candidate c'_j. For each $k, \ell \in [n]$ and each $i, j \in [m]$, we define a binary variable $P_{k,\ell,i,j}$ with the intention that $P_{k,\ell,i,j} = N_{k,\ell} \cdot M_{i,j}$. We introduce the following constraints:

$$\sum_{\ell=1}^{n} N_{k,\ell} = 1, \forall k \in [n]; \quad \sum_{k=1}^{n} N_{k,\ell} = 1, \forall \ell \in [n] \quad (1)$$

$$\sum_{j=1}^{m} M_{i,j} = 1, \forall i \in [m]; \quad \sum_{i=1}^{m} M_{i,j} = 1, \forall j \in [m] \quad (2)$$

$$\sum_{\ell \in [n], j \in [m]} P_{k,\ell,i,j} = 1, \quad \forall i \in [m], k \in [n] \quad (3)$$

$$\sum_{k \in [n], i \in [m]} P_{k,\ell,i,j} = 1, \quad \forall j \in [m], \ell \in [n] \quad (4)$$

$$P_{k,\ell,i,j} \le N_{k,\ell}, \qquad \forall i, j \in [m], k, \ell \in [n] \quad (5)$$

$$P_{k,\ell,i,j} \le M_{i,j}, \qquad \forall i, j \in [m], k, \ell \in [n] \quad (6)$$

Constraints (1) and (2) ensure that variables $N_{k,\ell}$ and $M_{i,j}$ describe matchings between voters and candidates, respectively. Constraints (3)–(6) implement the semantics of the $P_{k,\ell,i,j}$ variables (the former two ensure that for a given vote/candidate pair, there is exactly one vote/candidate pair in the other election to which they are matched; the latter two ensure connection between the $P_{k,\ell,i,j}$ variables and the $N_{k,\ell}$ and $M_{i,j}$ variables). The optimization goal is to minimize $\sum_{k,\ell \in [n], i,j \in [m]} P_{k,\ell,i,j} \cdot |\mathrm{pos}_{v_k}(c_i) - \mathrm{pos}_{v'_\ell}(c'_j)|$ (which, for values $P_{k,\ell,i,j}$ that satisfy the constraints of the program, defines the Spearman distance for the given matchings). Values $|\mathrm{pos}_{v_k}(c_i) - \mathrm{pos}_{v'_\ell}(c'_j)|$ are precomputed. $\quad\square$

While the ILPs described above find optimal solutions, they can be quite slow to solve for large instances (see experimental results below). Thus we also provide a heuristic algorithm. The main observation is that, for a given candidate matching, finding an optimal voter maching and the corresponding distance can be done efficiently, as described in Proposition 4. The heuristic is a local search on the candidate matching $\hat{\sigma}$, which we try to improve in each step:

1. Let $E = (C, V)$ and $E' = (C', V')$ be input elections. We initialize $\hat{\sigma}$ by ordering the candidates in each election according to their Borda scores, so that the Borda winner of one election is matched to the Borda winner of the other one, and so on (Borda score of a candidate c is $\sum_{v \in V} |C| - \mathrm{pos}_v(c)$).

2. We perform S iterations as follows (S is a parameter of the algorithm). We compute the distance d between the elections for the candidate matching $\hat{\sigma}$ (using Proposition 4). We find a candidate $c \in C$ that contributes most to the distance (i.e., the sum of values $|\mathrm{pos}_v(c) - \mathrm{pos}_{v'}(\hat{\sigma}(c))|$ for all the pairs of matched voters v, v' is highest for this c). We choose $d \in C$ uniformly at random and create matching $\hat{\sigma}'$ by swapping the candidates assigned to c and d. We compute the distance d' corresponding to $\hat{\sigma}'$ and set $\hat{\sigma} := \hat{\sigma}'$ if $d' < d$. (If in five consecutive iterations the matching remains unchanged then instead of considering the candidate that individually generates the highest distance, we simply swap the assignment of two candidates chosen uniformly at random.)

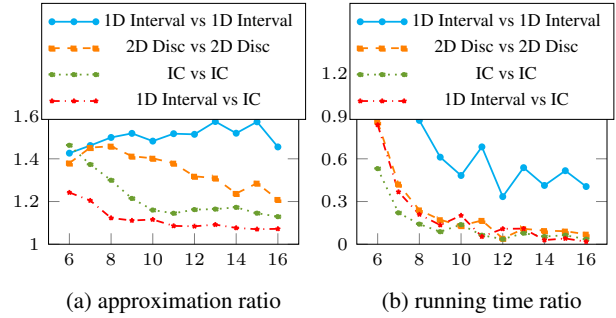(a) approximation ratio (b) running time ratio

Figure 1: Results of the local search heuristic: (a) average approximation ratio and (b) the ratio between its running time and that of CPLEX for our ILP, as functions of the number of voters (x-axis) and how the two elections are generated.

Note that we can halt the heuristic after each step, returning the best matching $\hat{\sigma}$ seen so far.

Experimental Evaluation. We report on experiments regarding the ILPs and the heuristic described above. Specifically, for a number of elections, generated according to several election distributions described below, we run both the ILP and the heuristic (we use $S = 50$ iterations) and compare them based on the Spearman isomorphism distances they find and their running time.

We consider the following distributions over elections. In the Impartial Culture model (IC), we choose the preference order of each voter uniformly at random among all possible ones. In the Euclidean models, each candidate and voter is a point in some Euclidean space and each voter v forms his or her preference order by sorting the candidates with respect to the increasing distance of their points from v's point. We use the 1D Interval model, where we choose candidates' and voters' points uniformly at random from the $[0, 1]$ interval, and the 2D Disc model, where we choose the points uniformly at random from a 2D disc (centered at point $(0, 0)$ and with radius 1). We consider elections with 6 candidates and between 6 and 16 voters (larger elections are infeasible to solve using our ILP formulation; as anecdotal evidence, for elections with 9 voters and 9 candidates it takes about 8 seconds to solve a single instance on our machine, but with 10 candidates and 10 voters it already takes 40 seconds).

We give the results of our experiments in Figure 1. In each case we consider values computed for elections generated according to given distributions. In the first plot, we report on the average ratio ρ between the distance computed by our heuristic and the optimal distance computed using ILP (we disregard cases of isomorphic elections, for which the optimal distance is 0). In the second plot, we give the ratio between the running times of the heuristic and the CPLEX solver used for the ILP. Each point in Figure 1 is averaged over 100 elections. We see that the average approximation ratio of the heuristic tends to decrease as the number of voters increases, and in our experiments it is at most 1.5. The running time of the heuristic is increasing much more slowly than that of CPLEX.

6 Conclusions

We have introduced the ELECTION ISOMORPHISM problem and the notion of isomorphism distances. We have shown that the former is computationally feasible, whereas for the latter there are strong intractability results. We present our complexity results in Table 1. Yet, we have found several FPT algorithms and a heuristic algorithm that gives some hope to be practical. Directions for future work include more involved tests of the heuristic and applying our tools to real-life problems (as suggested in the introduction). The latter requires introducing methodology for dealing with elections with different number of voters and candidates.

References

Arvind, V.; Köbler, J.; Kuhnert, S.; and Vasudev, Y. 2012. Approximate graph isomorphism. In *Proceedings of MFCS-2012*, 100–111.

Aziz, H.; Gaspers, S.; Mackenzie, S.; Mattei, N.; Narodytska, N.; and Walsh, T. 2015. Equilibria under the probabilistic serial rule. In *Proceedings of IJCAI-2015*, 1105–1112.

Babai, L.; Dawar, A.; Schweitzer, P.; and Torán, J. 2015. The graph isomorphism problem (dagstuhl seminar 15511). *Dagstuhl Reports* 5(12):1–17.

Bartholdi, III, J.; Tovey, C.; and Trick, M. 1989. Voting schemes for which it can be difficult to tell who won the election. *Social Choice and Welfare* 6(2):157–165.

Black, D. 1958. *The Theory of Committees and Elections*. Cambridge University Press.

Bredereck, R.; Faliszewski, P.; Kaczmarczyk, A.; Niedermeier, R.; Skowron, P.; and Talmon, N. 2017. Robustness among multiwinner voting rules. In *Proceedings of SAGT-2017*, 80–92.

Caragiannis, I.; Nath, S.; Procaccia, A.; and Shah, N. 2017. Subset selection via implicit utilitarian voting. *Journal of Artificial Intelligence Research* 58:123–152.

Diaconis, P., and Graham, R. 1977. Spearman's footrule as a measure of disarray. *Journal of the Royal Statistical Society. Series B* 39(2):262–268.

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

Elkind, E., and Slinko, A. 2016. Rationalizations of voting rules. In Brandt, F.; Conitzer, V.; Endriss, U.; Lang, J.; and Procaccia, A. D., eds., *Handbook of Computational Social Choice*. Cambridge University Press. chapter 8, 169–196.

Elkind, E.; Faliszewski, P.; Laslier, J.; Skowron, P.; Slinko, A.; and Talmon, N. 2017. What do multiwinner voting rules do? An experiment over the two-dimensional euclidean domain. In *Proceedings of AAAI-2017*, 494–501.

Elkind, E.; Faliszewski, P.; and Slinko, A. 2012. Rationalizations of condorcet-consistent rules via distances of hamming type. *Social Choice and Welfare* 39(4):891–905.

Elkind, E.; Faliszewski, P.; and Slinko, A. 2015. Distance rationalization of voting rules. *Social Choice and Welfare* 45(2):345–377.

Elkind, E.; Lackner, M.; and Peters, D. 2016. Preference restrictions in computational social choice: Recent progress. In *Proceedings of IJCAI-2016*, 4062–4065.

Enelow, J. M., and Hinich, M. J. 1984. *The spatial theory of voting: An introduction*. Cambridge University Press.

Enelow, J. M., and Hinich, M. J. 1990. *Advances in the spatial theory of voting*. Cambridge University Press.

Eğecioğlu, Ö., and Giritligil, A. 2013. The impartial, anonymous, and neutral culture model: A probability model for sampling public preference structures. *Journal of Mathematical Sociology* 37(4):203–222.

Grohe, M.; Rattan, G.; and Woeginger, G. J. 2018. Graph similarity and approximate isomorphism. *arXiv preprint arXiv:1802.08509*.

Hashemi, V., and Endriss, U. 2014. Measuring diversity of preferences in a group. In *Proceedings of ECAI-2014*, 423–428.

Litvak, B. 1983. Distances and consensus rankings. *Cybernetics and systems analysis* 19(1):71–81. Translated from Kibernetika, No. 1, pp. 57–63, January–February, 1983.

Mattei, N., and Walsh, T. 2013. Preflib: A library for preferences. In *Proceedings of ADT-2013*, 259–270.

Meskanen, T., and Nurmi, H. 2008. Closeness counts in social choice. In Braham, M., and Steffen, F., eds., *Power, Freedom, and Voting*. Springer-Verlag.

Mirrlees, J. 1971. An exploration in the theory of optimal income taxation. *Review of Economic Studies* 38:175–208.

Monjardet, B. 2009. Acyclic domains of linear orders: A survey. In Brams, S.; Gehrlein, W.; and Roberts, F., eds., *The Mathematics of Preference, Choice and Order*, Studies in Choice and Welfare. Springer Berlin Heidelberg. 139–160.

Narodytska, N., and Walsh, T. 2014. The computational impact of partial votes on strategic voting. In *Proceedings of ECAI-2014*, 657–662.

Nitzan, S. 1981. Some measures of closeness to unanimity and their implications. *Theory and Decision* 13(2):129–138.

Puppe, C., and Slinko, A. 2017. Condorcet domains, median graphs and the single-crossing property. *Economic Theory* 1–34.

Roberts, K. 1977. Voting over income tax schedules. *Journal of Public Economics* 8(3):329–340.