Computing Kemeny Rankings From d-Euclidean Preferences

Thekla Hamm, Martin Lackner, and Anna Rapberger

TU Wien, Vienna, Austria
⟨firstname⟩.⟨lastname⟩@tuwien.ac.at

Abstract. Kemeny's voting rule is a well-known and computationally intractable rank aggregation method. In this work, we propose an algorithm that finds an embeddable Kemeny ranking in d-Euclidean elections. This algorithm achieves a polynomial runtime (for a fixed dimension d) and thus demonstrates the algorithmic usefulness of the d-Euclidean restriction. We further investigate how well embeddable Kemeny rankings approximate optimal (unrestricted) Kemeny rankings.

1 Introduction

Rank aggregation is the problem of combining a collection of rankings into a social "consensus" ranking, with applications ranging from multi-agent planning [22] and collaborative filtering [32] to internet search [5, 17]. The classic application of rank aggregation is voting and thus rank aggregation methods are extensively studied in social choice theory, where rankings correspond to voters' preferences. A prominent rank aggregation method is *Kemeny's voting rule*, also known as Kemeny-Young method. This method is based on the Kendall-tau distance between rankings and outputs a *consensus ranking (or Kemeny ranking)* that minimizes the sum of distances to the input rankings.

Kemeny's voting rule is of particular importance for two reasons: First, it is the only rank aggregation method satisfying three desirable properties (neutrality, consistency, and being a Condorcet method) [39]. Second, it is the maximum likelihood estimator for the "correct" ranking if the input is viewed as noisy perceptions of a ground truth (assuming a very natural noise model) [40]. However, Kemeny's rule has a main disadvantage: its computational complexity [7, 27]. In particular, computing the Kemeny score is NP-hard even for four voters [17].

Due to the importance of Kemeny's rule, much algorithmic research has been conducted with the goal to overcome this computational barrier. The majority of this work has focused on approximation algorithms, parameterized algorithms and heuristical methods (see related work below). In this paper, we take an approach that is widely used in computational social choice: to restrict the input to a smaller preference domain [20]. If the input rankings possess a favorable structure, it may be possible to circumvent hardness results that hold in the general case. For Kemeny's rule, this is the case if the input has a certain 1-dimensional structure; more specifically, Kemeny's rule is polynomial-time computable for single-peaked rankings [11] and for rankings with bounded single-peaked or single-crossing width [14]. In contrast, Kemeny's rule remains NP-hard for preferences that are single-peaked on a circle [34] and, as very recently shown in [24], for d-Euclidean preferences with $d \geq 2$. In fact, both preference domains admit an interesting connection: In [38] it has been shown that preferences that are single-peaked on a circle can capture specific 2-Euclidean preferences.

The d-Euclidean preference domain [10, 21] is a d-dimensional spatial model based on the assumption that voters and candidates can be placed in $\mathbb{R}^d$ and a voter's preference ranking is derived from the Euclidean distance between her coordinates and the candidates—closer candidates being more preferable. This model captures situations where voters' preferences are mainly determined by real-valued attributes of candidates (e.g., a political candidate may be placed in a two-dimensional space with axes corresponding to her position on economic and social issues, or a textbook might be judged on its focus on theory/applications and on its complexity level). It is intuitively clear that a one-dimensional model is too simplistic to capture most real-world situations, and more dimensions greatly increase the applicability of this domain. However, as mentioned before, it is not the case that simply restricting the input to d-Euclidean preferences yields a computational advantage as the problem remains NP-hard [24].

The goal of our paper is to find an efficient algorithm for Kemeny's voting rule given d-Euclidean preferences (for $d \geq 2$) by additionally imposing reasonable restrictions on the output. We work under the assumption that an embedding witnessing the d-Euclidean property is known and that the consensus ranking (i.e., the output) has to be embeddable via the same embedding. The embeddability of the consensus ranking is a sensible assumption as it extends the explanation of the preference structure to the consensus ranking, i.e., if voters' preferences can be understood as points in a d-dimensional space, then also the output should be explainable via this space. Our main result is that this problem can be solved in time in $\mathcal{O}(|\mathcal{C}|^{4d})$ for strict orders and $\tilde{\mathcal{O}}(|\mathcal{C}|^{4.746 \cdot d+2})$ for weak orders (with ties), i.e., it is solvable in polynomial time for a fixed dimension d. This algorithm makes use of a correspondence between embeddable rankings and faces of a hyperplane arrangement in which each hyperplane is equidistant to two embedded candidates. The determination of an embeddable consensus ranking is then performed on an appropriately constructed vertex- and edge-weighted graph, which is extracted from the arrangement.

We further show that this algorithm can be adapted to an egalitarian variant of the Kemeny problem, which minimizes the maximum Kendall-tau distance. Finally, we study the restriction of requiring an embeddable consensus ranking in more detail. We prove that an embeddable consensus ranking has at most twice the Kemeny score of the optimal, unrestricted Kemeny ranking. In numerical experiments, we show that the embeddable Kemeny ranking and the optimal Kemeny ranking coincide in most small instances.

Related work. In addition to the results by Escoffier et al. [24] who showed NP-hardness of Kemeny's voting rule given d-Euclidean preferences for $d \geq 2$, the work of Peters [33] on the recognition of d-Euclidean elections is of particular importance to our problem. Peters shows that this problem is NP-hard for $d \geq 2$ [33] (it is even $\exists\mathbb{R}$-complete). Thus, one cannot hope for a polynomial-time algorithm for our problem if the embedding is removed from the input. Instead, we assume that the embedding is either found in a preprocessing stage (with sufficient time available) or is known due to understanding the origin of preferences (which adhere to a d-dimensional geometry). In contrast, recognizing 1-Euclidean elections is possible in polynomial time [16, 29].

As mentioned before, Kemeny's rule has attracted much attention from an algorithmic perspective: exponential-time search-based techniques [6, 13, 15], approximation algorithms [1, 28], parameterized algorithms [8, 14], and heuristical algorithm [2, 36]. As Kemeny's voting rule is of practical importance, much work has also been invested in runtime benchmarks [3].

2 Preliminaries

A weak order $\succeq$ over a set X is a complete ($x \succeq y$ or $y \succeq x$ for all $x, y \in X$) and transitive binary relation. We write $x \succ y$ if $x \succeq y$ but not $y \succeq x$. Further, we write $x \sim y$ if $x \succeq y$ and $y \succeq x$. A weak order $\succeq$ is a *strict order* if it has no ties, i.e., if $x \neq y$ then either $x \succ y$ or $y \succ x$.

We define an *election* $(\mathcal{C}, \mathcal{V}, (\succeq_v)_{v\in\mathcal{V}})$ as a set of *candidates* $\mathcal{C}$, a set of *voters* $\mathcal{V}$, and for each $v \in \mathcal{V}$, a weak order $\succeq_v$ over the candidates called the *preference (order) of* v. Whenever $c \succeq_v c'$, we say that v *prefers* c over c'

Let d be positive integer and let $p : \mathcal{C} \cup \mathcal{V} \to \mathbb{R}^d$ be an *embedding* in the d-dimensional space. Further, let $\|\cdot\|_d$ denote the Euclidean norm in $\mathbb{R}^d$. We say that a voter's preference order $\succeq_v$ for $v \in \mathcal{V}$ on $\mathcal{C}$ is *p-embeddable* if for all $c, c' \in \mathcal{C}$, $c \succeq c'$ if and only if $\|p(v) - p(c)\|_d \leq \|p(v) - p(c')\|_d$. Generally for a weak order $\succeq$ on $\mathcal{C}$ that do not coincide with a voter's preference order, we say $\succeq$ is *p-embeddable* if there is some $x \in \mathbb{R}^d$ such that for all $c, c' \in \mathcal{C}$, $c \succeq c'$ if and only if $\|x - p(c)\|_d \leq \|x - p(c')\|_d$. An election $(\mathcal{C}, \mathcal{V}, (\succeq_v)_{v\in\mathcal{V}})$ is said to be p-embeddable if $\succeq_v$ for all $v \in \mathcal{V}$ are p-embeddable. Finally, an election is *d-Euclidean* if it it is p-embeddable for some p.

We define the *Kendall-tau distance* of two weak orders $\succeq, \succeq'$ over $\mathcal{C}$ as

$$\mathrm{K}(\succeq, \succeq') = \sum_{\{x,y\}\subseteq\mathcal{C}} d_{\succeq,\succeq'}(x, y), \quad \text{where}$$

$$d_{\succeq,\succeq'}(x, y) = \begin{cases} 2 & \text{if } (x \succ y \text{ and } y \succ' x) \text{ or } (y \succ x \text{ and } x \succ' y) \\ 1 & \text{if } (x \sim y \text{ and } x \not\sim' y) \text{ or } (x \not\sim y \text{ and } x \sim' y) \\ 0 & \text{otherwise (i.e., } \succ \text{ and } \succ' \text{ agree on the order of } x \text{ and } y). \end{cases}$$

Equivalently,

$$\begin{aligned} \mathrm{K}(\succeq, \succeq') = &\ |\{\{x, y\} \subseteq \mathcal{C} \mid (x \succeq y \wedge y \succ' x) \vee (y \succeq x \wedge x \succ' y)\}| \\ &+ |\{(x, y\} \subseteq \mathcal{C} \mid (x \succeq' y \wedge y \succ x) \vee (y \succeq' x \wedge x \succ y)\}|. \end{aligned}$$

For strict orders $\succ$ and $\succ'$, this definition simplifies to the number of ordered candidate pairs on which the two orders disagree, i.e., $\mathrm{K}(\succ, \succ') = |\{(x, y) \in \mathcal{C}^2 \mid (x \succ y \wedge y \succ' x) \vee (y \succ x \wedge x \succ' y)\}|$.

We can now define Kemeny's voting rule and the corresponding consensus rankings, which we refer to as *optimal Kemeny rankings* in the following.

Definition 1. *Given an election $(\mathcal{C}, \mathcal{V}, (\succeq_v)_{v\in\mathcal{V}})$, a strict order $\succ$ on $\mathcal{C}$ is an* optimal Kemeny ranking *if there is no other strict order $\succ'$ on $\mathcal{C}$ with*

$$\sum_{v\in\mathcal{V}} \mathrm{K}(\succ', \succeq_v) < \sum_{v\in\mathcal{V}} \mathrm{K}(\succ, \succeq_v),$$

i.e., an optimal Kemeny ranking minimizes the sum of Kendall-tau distances to the preference orders. We refer to $\sum_{v\in\mathcal{V}} \mathrm{K}(\succ, \succeq_v)$ as the Kemeny score of $\succ$.

We note that Definition 1 could be adapted to define Kemeny rankings as weak orders; this would not change our results.

From a computational viewpoint, Kemeny's voting rule is captured by the following NP-hard decision problem [7, 17, 27]:

KEMENY SCORE
Instance: An election $(\mathcal{C}, \mathcal{V}, (\succeq_v)_{v\in\mathcal{V}})$ and an objective value $z \in \mathbb{N}$.
Question: Is there a strict order $\succ$ on $\mathcal{C}$ such that $\sum_{v\in\mathcal{V}} \mathrm{K}(\succ, \succeq_v) \leq z$?

We furthermore consider an *egalitarian variant* which minimizes the maximal dissatisfaction of each voter.

Definition 2. *Given an election $(\mathcal{C}, \mathcal{V}, (\succeq_v)_{v\in\mathcal{V}})$, we say that a strict order $\succ$ on $\mathcal{C}$ is an* egalitarian Kemeny ranking *if there is no other strict order $\succ' \neq \succ$ on $\mathcal{C}$ with $\max_{v\in\mathcal{V}} \mathrm{K}(\succ', \succeq_v) < \max_{v\in\mathcal{V}} \mathrm{K}(\succ, \succeq_v)$.*

Like for KEMENY SCORE, the corresponding decision problem EGALITARIAN KEMENY SCORE, i.e., given $(\mathcal{C}, \mathcal{V}, (\succeq_v)_{v\in\mathcal{V}})$, $z \in \mathbb{N}$, decide whether there is a strict order $\succ$ on $\mathcal{C}$ such that $\max_{v\in\mathcal{V}} \mathrm{K}(\succ, \succeq_v) \leq z$, is NP-hard even for four voters which was independently proved by Biedl et al. [9] and Popov [35].

3 Embeddable Kemeny Rankings

The main focus of this paper is on the constrained setting of d-Euclidean elections, that is, we assume that the input is an embedding p as well as a p-embeddable election. In addition, we require that the output (i.e., the Kemeny ranking) is also p-embeddable.

Definition 3. *Given an embedding $p : \mathcal{C} \cup \mathcal{V} \to \mathbb{R}^d$ and a p-embeddable election $(\mathcal{C}, \mathcal{V}, (\succeq_v)_{v\in\mathcal{V}})$, a strict order $\succ$ on $\mathcal{C}$ is a* p-embeddable Kemeny ranking *if $\succ$ is p-embeddable and there is no other p-embeddable strict order $\succ'$ on $\mathcal{C}$ such that $\sum_{v\in\mathcal{V}} \mathrm{K}(\succ', \succeq_v) < \sum_{v\in\mathcal{V}} \mathrm{K}(\succ, \succeq_v)$.*

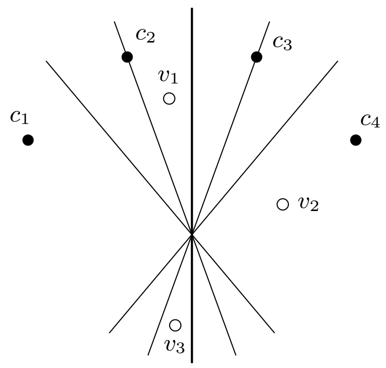


Fig. 1: Election from Example 1.

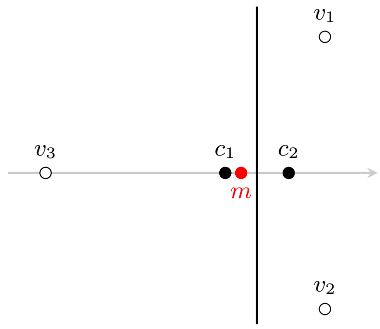


Fig. 2: Election from Example 2

A *p-embeddable egalitarian Kemeny ranking* is defined analogously.

First we observe that a p-embeddable Kemeny ranking does not need to coincide with any optimal Kemeny rankings for a given p-embeddable election.

Example 1. Consider the voting setting depicted in Figure 1. The preferences of voter v_1 are given by $c_2 \succ_1 c_3 \succ_1 c_1 \succ_1 c_4$, the preferences of voter v_2 are $c_4 \succ_2 c_3 \succ_2 c_2 \succ_2 c_1$ and v_3 prefers $c_1 \succ_3 c_4 \succ_3 c_2 \succ_3 c_3$. The unique Kemeny ranking is $c_4 \succ c_2 \succ c_3 \succ c_1$ (with a Kemeny score of 14) since $\mathrm{K}(\succ, \succ_1) = 6$, $\mathrm{K}(\succ, \succ_2) = 2$, $\mathrm{K}(\succ, \succ_3) = 6$, and $\sum_{i \leq 3} \mathrm{K}(\succ', \succ_{v_i}) > 14$ for all $\succ' \neq \succ$. Now observe that $\succ$ is not embeddable in Figure 1. Among embeddable rankings, the Kemeny score is minimized by $\succ_1$, $\succ_2$, and $\succ_3$, all of which achieve a Kemeny score of 16. These are the embeddable Kemeny rankings.

One may ask whether it is sensible to use an ordinal voting rule such as Kemeny's rule in our setting where voters and candidates can be represented in a coordinate space. It is important to note that we do *not* assume that a voter's position in $\mathbb{R}^d$, given by an embedding, is actually a correct representation of this voter's preferences. In particular, we do not assume that distances between voters and candidates is an accurate measure of *intensities*. That is, a voter prefers a candidate with distance 1 over a candidate with distance 2, but not necessarily twice as much. Hence, our assumption of embeddability in d-Euclidean space is significantly weaker than assuming a model where distances correspond to voters' utilities. In such a model, ordinal voting rules indeed are less useful and choosing the geometric median of the set of voter points[1] is more natural than computing a Kemeny ranking (in contrast to Kemeny's rule, the geometric median can be computed efficiently [12]). The next example shows that these two concepts differ.

[1] The geometric median of a set of points S is a point that minimizes the sum of distances to points in S (as does the Kemeny ranking albeit for a different metric).

Example 2. Consider a 2-Euclidean election with two candidates $\mathcal{C} = \{c_1, c_2\}$ and three voters $\mathcal{V} = \{v_1, v_2, v_3\}$. The embedding p is given by $p(c_2) = -p(c_1) = (1, 0)$; $p(v_1) = (3, 6)$, $p(v_2) = (3, -6)$, $p(v_3) = (-10, 0)$. Voters v_1, v_2 prefer c_2 over c_1 while voter v_3 prefers c_1 over c_2. The optimal Kemeny ranking is thus $c_2 \succ c_1$ (which is clearly p-embeddable). In contrast, the geometric median m is the point $\approx (-0.46, 0)$ which lies on the side of $p(c_1)$ and thus corresponds to the ordering $c_1 \succ c_2$. The crucial point here is that if we changed the embedding so that $p(v_1) = (4, 6)$, $p(v_2) = (4, -6)$, the geometric median would lie at $\approx (0.54, 0)$ and thus correspond to the Kemeny ranking.

A similar observation can be made in the case of the egalitarian Kemeny ranking; minimizing the maximum Euclidean distance is known as the 1-center problem or smallest enclosing ball problem.

For the 1-dimensional case, the question is easy to answer.

Proposition 1. *In a p-embeddable 1-Euclidean election, any optimal Kemeny ranking is also p-embeddable and coincides with the geometric median.*

As we have seen before, Proposition 1 does not extend to higher dimensions: Examples 1 and 2 are counter-examples for $d = 2$.

4 Computing Embeddable Kemeny Rankings

In this section, we give a brute-force algorithm to determine all p-embeddable Kemeny rankings of a given p-embeddable election. In order to traverse all strict p-embeddable orders, we observe their correspondence to faces of the hyperplane arrangement that contains all hyperplanes consisting of points equidistant to any two embedded candidates. This correspondence is also important for our main algorithm (Section 5), which drastically improves the asymptotic runtime.

Consider a d-Euclidean election $(\mathcal{C}, \mathcal{V}, (\succeq_v)_{v \in \mathcal{V}})$ embedded via $p : \mathcal{C} \cup \mathcal{V} \to \mathbb{R}^d$. For any pair $c, c' \in \mathcal{C}$ of candidates we consider the hyperplane $S_{c,c'} = \{x \in \mathbb{R}^d \mid \|x - p(c)\|_d = \|x - p(c')\|_d\}$. Each $S_{c,c'}$ divides $\mathbb{R}^d$ into two halfspaces — one containing $p(c)$, we also say this halfspace *lies on the same side* of $S_{c,c'}$ as c; and one containing $p(c')$. Each halfspace is assumed to be closed, that is, it contains its bounding hyperplane. A *face* of the hyperplane arrangement $\{S_{c,c'} \mid c, c' \in \mathcal{C}\}$ is a connected non-empty subspace of $\mathbb{R}^d$ obtained by intersecting halfspaces of the arrangement with at least one halfspace chosen for each hyperplane $S_{c,c'}$. We write $\mathcal{P}$ to denote the set of all faces of the arrangement.

Let $f \in \mathcal{P}$ be a face. For any pair of candidates $c, c' \in \mathcal{C}$, we say that f *lies on the same side* of $S_{c,c'}$ as c, if it is a subset of the halfspace that lies on the same side of $S_{c,c'}$ as c. This allows us to identify f by the set $X = \{(c, c') \in \mathcal{C}^2 \mid c \text{ and the subspace lie on the same side of } S_{c,c'}\}$; we write f_X to denote the face identified by X, i.e., $f_X = f$. A face f is called k-face if it has dimension k. Observe that for every face f_X, either $(c, c') \in X$ or $(c', c) \in X$ for every pair $c, c' \in \mathcal{C}$. Further note that X can also contain both tuples (c, c'), (c', c)—in that case, $f_X \subseteq S_{c,c'}$. For a face f_X, if $(c, c') \in X$ then $f_X \subseteq \{x \in$

$\mathbb{R}^d \mid \|x-p(c)\|_d \leq \|x-p(c')\|_d\}$. Additionally we denote the set of d-dimensional faces as $\mathcal{R}$ and refer to them as *regions*. In the following, we use the standard notation f° for the interior of a set f.

Intuitively, each face f_X corresponds to a weak p-embeddable order for the given d-Euclidean election and embedding p. This correspondence is formally captured by the following result.

Lemma 1. *Let $\Phi\colon \mathcal{P} \to \{\succeq \subseteq \mathcal{C}^2 \mid \succeq$ is a p-embeddable weak order$\}$ be a function defined by $\Phi(f_X) = \succeq$ where $c \succeq c' \Leftrightarrow (c,c') \in X$. Then Φ is a bijection.*

Since we require that Kemeny rankings are strict, the following observation showing that each region corresponds to a strict p-embeddable ordering for the given embedded d-Euclidean election is also useful.

Lemma 2. *Let $\Phi'\colon \mathcal{R} \to \{\succeq \subseteq \mathcal{C}^2 \mid \succeq$ is a p-embeddable strict order$\}$ be the restriction of Φ (from Lemma 1) to regions. Also Φ' is a bijection.*

For a face $f \in \mathcal{P}$, we write $\succeq_f$ instead of $\Phi(f)$ (this is a weak order). Further, for a region R, we write $\succ_R$ instead of $\Phi'(R)$ (this is a strict order).

We can now use the preceding correspondences to give a straightforward polynomial time algorithm that enumerates all p-embeddable strict orders.

Theorem 1. *Determining all p-embeddable Kemeny rankings for a d-Euclidean election $(\mathcal{C}, \mathcal{V}, (\succeq_v)_{v\in\mathcal{V}})$ given by $p : \mathcal{C} \cup \mathcal{V} \to \mathbb{R}^d$ is possible in time polynomial in $|\mathcal{C}|$, more specifically in time in $\mathcal{O}(|\mathcal{C}|^{6d})$.*

Proof. Consider the d-Euclidean preference profile given by the function $p : \mathcal{C} \cup \mathcal{V} \to \mathbb{R}^d$. For every $f \in \mathcal{P}$, let $\#(f)$ denote the number of voters in f, i.e. $\#(f) = |\{v \in \mathcal{V} \mid p(v) \in f\}|$. By comparing the corresponding values for each $R \in \mathcal{R}$, we can determine $R \in \mathcal{R}$ which minimizes $\sum_{f'\in\mathcal{P}} \#(f')\cdot\mathrm{K}(\succeq_{f'}, \succ_R)$, and denote such an R by $R_{\min}$. We return $\succ_{R_{\min}}$ as p-embeddable Kemeny ranking.

Correctness. For $R \in \mathcal{R}$ and $f' \in \mathcal{P}$,

$$\begin{aligned}\sum_{f'\in\mathcal{P}} \#(f')\cdot\mathrm{K}(\succeq_{f'}, \succ_R) &= \sum_{f'\in\mathcal{P}} \sum_{\substack{v\in\mathcal{V}\\ p(v)\in f'}} \mathrm{K}(\succeq_{f'}, \succ_R)\\ &= \sum_{f'\in\mathcal{P}} \sum_{\substack{v\in\mathcal{V}\\ p(v)\in f'}} \mathrm{K}(\succeq_{v}, \succ_R)\\ &= \sum_{v\in\mathcal{V}} \mathrm{K}(\succeq_{v}, \succ_R)\end{aligned}$$

Since we are looking for a p-embeddable Kemeny ranking, it has to have the form $\succ_R$ for some $R \in \mathcal{R}$ by Lemma 2, which implies correctness.

Running time. The hyperplane arrangement induces $\mathcal{O}(|\mathcal{C}|^{2d})$ faces (by [26, Corollary 28.1.2] as we consider at most $\binom{|\mathcal{C}|}{2}$ distinct hyperplanes) and can be computed in time in $\mathcal{O}(|\mathcal{C}|^{2d})$ [18, Theorem 7.6]. For each face $R \in \mathcal{R}$, the computation and comparison of the objective function naively requires time in $\mathcal{O}(|\mathcal{P}|^2) \subseteq \mathcal{O}(|\mathcal{C}|^{4d})$. Thus the overall complexity of the procedure lies in $\mathcal{O}(|\mathcal{C}|^{6d})$. □

An analogous procedure works for the egalitarian variant.

5 Increasing Efficiency

To achieve a better runtime—in particular for large d—we conduct a more in-depth graphical analysis of the relation of p-embeddable orders to each other. Specifically, this section is dedicated to proving our following main result.

Theorem 2 (Main Theorem). *Determining all p-embeddable Kemeny rankings for a d-Euclidean election $(\mathcal{C}, \mathcal{V}, (\succeq_v)_{v\in\mathcal{V}})$ given by $p : \mathcal{C} \cup \mathcal{V} \to \mathbb{R}^d$ is possible in time in $\tilde{\mathcal{O}}(|\mathcal{C}|^{2(d\cdot\omega+1)})$, where $\omega < 2.373$ [4] is the exponent of matrix multiplication.*

5.1 Preference Graph

We define the preference graph G_{pref} as the edge-weighted graph given by setting

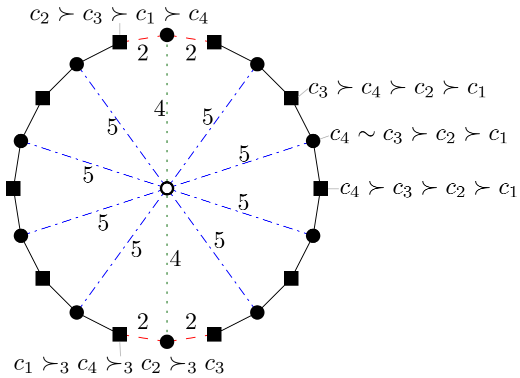
- $V(G_{\text{pref}}) = \{v_f \mid f \in \mathcal{P}\}$;
- $E(G_{\text{pref}}) = \{\{v_f, v_{f'}\} \mid (\dim(f) = \dim(f') - 1 \wedge f \subset f') \vee (\dim(f') = \dim(f) - 1 \wedge f' \subset f)\}$; and
- $w : E(G_{\text{pref}}) \to \mathbb{N}$, $\{v_f, v_{f'}\} \mapsto |\{\{c, c'\} \subseteq \mathcal{C} \mid (\dim(f' \cap S_{c,c'}) = \dim(f') \wedge \dim(f \cap S_{c,c'}) \neq \dim(f)) \vee (\dim(f \cap S_{c,c'}) = \dim(f) \wedge \dim(f' \cap S_{c,c'}) \neq \dim(f'))\}|$.

In other words, vertices corresponding to faces one of which is contained in the other are connected to each other by edges in G_{pref} whenever the dimension of one face differs from the other by exactly one. The edge weights correspond to the number of pairs (c, c') of candidates inducing this respective hyperplane. An example is given in Figure 3. By a bound on the number of faces [26, Corollary 28.1.2] and since we consider at most $\binom{|\mathcal{C}|}{2}$ different hyperplanes, we can bound the number of vertices by $|V(G_{\text{pref}})| \in \mathcal{O}(|\mathcal{C}|^{2d})$.

G_{pref} without weights coincides with the *incidence graph* of a hyperplane arrangement as defined in [18] which is constructed in $\mathcal{O}(|\mathcal{C}|^{2d})$-time [18, Theorem 7.6]. We modify this procedure to include appropriate edge weights for G_{pref}.

Lemma 3. G_{pref} *can be constructed in time in* $\mathcal{O}(|\mathcal{C}|^{2d})$.

Fig. 3: G_{pref} for candidates as given in Example 1. Vertex shapes encode the dimensions of the corresponding faces, and dash-styles encode weights where edges without weight labels have unit-weight. Exemplary vertices are annotated with the corresponding p-embeddable orders.



Note that at this point, we have set up or shown natural bijective correspondences between: the vertices of G_{pref}, the faces in $\mathcal{P}$, all p-embeddable orders of $\mathcal{C}$ and sets of pairs of candidates in $\mathcal{C}$ which explicitly encode the pairwise comparisons according to such p-embeddable orders. In this way, it will be natural to write any $v \in V(G_{\text{pref}})$ as v_f for some $f \in \mathcal{P}$, any p-embeddable order of $\mathcal{C}$ as $\succeq_f$ for some $f \in \mathcal{P}$, and any $f \in \mathcal{P}$ as f_X for some $X \subseteq \mathcal{C}^2$.

5.2 Shortest Paths in the Preference Graph

The crucial property of the preference graph, apart from capturing p-embeddable orders through its vertices, is that the chosen edge weights reflect the Kendall-tau distance between embeddable orders. We first can show this for single edges.

Lemma 4. *For* $\{v_f, v_{f'}\} \in E(G_{\text{pref}})$, $w(\{v_f, v_{f'}\}) = \mathrm{K}(\succeq_f, \succeq_{f'})$.

This previous lemma acts as the base case for the general correspondence of distances in G_{pref} and the Kendall-tau distance between the orders associated to the vertices of G_{pref} (i.e. the p-embeddable orders). We denote by $\mathrm{dist}_{G_{\text{pref}}}(v, w)$ the length of a shortest (in terms of summed edge weights) v-w-path in G_{pref}.

Lemma 5. *For* $f, f' \in \mathcal{P}$, $\mathrm{K}(\succeq_f, \succeq_{f'}) = \mathrm{dist}_{G_{\text{pref}}}(v_f, v_{f'})$.

Proof Sketch. We present a proof by induction over the length ℓ of cardinality-minimal shortest v_f-$v_{f'}$-paths (i.e., a path having minimum number of vertices among all weight-minimal paths between v_f, $v_{f'}$). The proof makes use of the observation that the Kendall-Tau distance between two faces f_X, f_Y corresponds to the symmetric difference $|X \Delta Y|$. The base case $\ell = 2$ is covered by Lemma 4.

Now assume that the statement holds for any cardinality-minimal shortest path of length $\ell - 1$ and observe that each proper subpath of a cardinality-minimal shortest v_f-$v_{f'}$-path consisting of ℓ vertices in G_{pref} is cardinality-minimal; otherwise one can replace the subpath with a cardinality-minimal shortest path, contradicting the assumption on $v_f \ldots v_{f'}$. Together with the triangle-inequality for the Kendall-tau distance, we get $\mathrm{K}(\succeq_f, \succeq_{f'}) \leq \mathrm{dist}_{G_{\text{pref}}}(v_f, v_{f'})$.

To show $\mathrm{dist}_{G_{\mathrm{pref}}}(v_f, v_{f'}) \leq \mathrm{K}(\succeq_f, \succeq_{f'})$, we construct a v_f-$v_{f'}$-path of weight $\mathrm{K}(\succeq_f, \succeq_{f'})$ by connecting two arbitrary points $p_f \in f^\circ$ and $p_{f'} \in f'^\circ$ via a straight line l and extracting a path along the traversal of l from p_f to $p_{f'}$. The path consists of vertices v_g with $l \cap g \neq \emptyset$ such that $g \in \mathcal{P}$ satisfies $\dim(g) < \dim(g')$ for all $g' \in \mathcal{P}$ with $l \cap g = l \cap g'$; also, we connect every two vertices v_i, v_{i+1} which are—w.r.t. the ordering along the line traversal —"adjacent" but not connected via an edge (i.e., $|\dim(f_i) - \dim(f_{i+1})| > 1$ for the corresponding faces f_i, f_{i+1}) via a weight- and vertex-minimal path.

Let $v_f = v_{f_1} \ldots v_{f_s} = v_{f'}$ denote the constructed v_f-$v_{f'}$-path P and let $X_1, \ldots, X_s \subseteq \mathcal{C}^2$ denote the pairs of candidates such that $f_i = f_{X_i}$ according to our notation introduced in Section 4. We verify that the constructed path P has the desired weight $K(\succeq_f, \succeq_{f'}) = |X_1 \triangle X_s|$ by showing that a pair $(c, c') \in \mathcal{C}^2$ contributes to the weight of P exactly once if and only if $(c, c') \in X_1 \triangle X_s$. Indeed, it can be shown that there is at most one edge $\{v_{f_i}, v_{f_{i+1}}\} \in P$ satisfying $f_i \cap S_{c,c'} = \emptyset$ but $f_{i+1} \cap S_{c,c'} \neq \emptyset$; also there is at most one edge $\{v_{f_i}, v_{f_{i+1}}\} \in P$ satisfying $f_i \cap S_{c,c'} \neq \emptyset$ but $f_{i+1} \cap S_{c,c'} = \emptyset$; i.e., P "enters" and "exists" a hyperplane $S_{c,c'}$ only once. This follows from the construction and by the fact that a straight line intersects a hyperplane at most once. □

5.3 The Algorithm

Having established the correspondence between the Kendall-tau distance and the shortest paths in the edge-weighted graph G_{pref} we obtain the following result.

Theorem 2 (Main Theorem). *Determining all p-embeddable Kemeny rankings for a d-Euclidean election $(\mathcal{C}, \mathcal{V}, (\succeq_v)_{v \in \mathcal{V}})$ given by $p : \mathcal{C} \cup \mathcal{V} \to \mathbb{R}^d$ is possible in time in $\tilde{\mathcal{O}}(|\mathcal{C}|^{2(d \cdot \omega + 1)})$, where $\omega < 2.373$ [4] is the exponent of matrix multiplication.*

Proof. Consider the d-Euclidean preference profile given by the function $p : \mathcal{C} \cup \mathcal{V} \to \mathbb{R}^d$. We construct the corresponding preference graph G_{pref} using Lemma 3. We then apply the Shoshan-Zwick all-pairs shortest path algorithm for undirected graphs with integer weights (proposed in [37] and corrected in [19]) which returns a matrix $M_{\mathrm{dist}} \in \mathbb{N}^{V(G_{\mathrm{pref}}) \times V(G_{\mathrm{pref}})}$ containing the length of the shortest path between every pair of vertices in G_{pref}. For every vertex $v_f \in V(G_{\mathrm{pref}})$, let $\#(v_f)$ denote the number of voters in f, i.e. $\# : V(G_{\mathrm{pref}}) \to \mathbb{N}$ with $\#(v_f) = |\{v \in \mathcal{V} \mid p(v) \in f\}|$, or equivalently $\#(v_f) = |\{v \in \mathcal{V} \mid \succeq_v = \succeq_f\}|$. By comparing the corresponding values for each $R \in \mathcal{R}$, we can determine all $R \in \mathcal{R}$ which minimize $\sum_{f' \in \mathcal{P}} \#(v_{f'}) \cdot \mathrm{dist}_{G_{\mathrm{pref}}}(v_{f'}, v_R)$, and denote such an R by $R_{\min}$. We return the (set of) all such $\succ_{R_{\min}}$ as p-embeddable Kemeny rankings.

Correctness follows from the Lemmas 5, 2, and 1.

Running time. The construction of the preference graph takes time in $\mathcal{O}(|\mathcal{C}|^{2d})$ by Lemma 3. By [19, 37], the all-pairs shortest path algorithm for undirected graphs with integer weights runs in time in $\tilde{\mathcal{O}}(M \cdot |V(G_{\mathrm{pref}})|^\omega)$ where M is the largest edge weight and $\omega < 2.373$ is the exponent of matrix multiplication.

Since $M \leq \binom{|\mathcal{C}|}{2}$ we get $\tilde{\mathcal{O}}(M \cdot |V(G_{\text{pref}})|^{\omega}) = \tilde{\mathcal{O}}(|\mathcal{C}|^{2(d\omega+1)})$ The computation and comparison of the objective function for each $f \in \mathcal{P}$ naively requires time in $\mathcal{O}(|\mathcal{P}|^2) \subseteq \mathcal{O}(|\mathcal{C}|^{4d})$. Thus the overall complexity lies in $\tilde{\mathcal{O}}(|\mathcal{C}|^{2(d\omega+1)})$. □

Weak Kemeny Rankings. We remark that whenever we allow p-embeddable Kemeny rankings to be weak rather than strict, we can easily adapt our algorithm by comparing the values of $\sum_{f' \in \mathcal{P}} \#(v_{f'}) \cdot \text{dist}_{G_{\text{pref}}}(v_{f'}, v_f)$ for each $f \in \mathcal{P}$, denoting an f that minimizes this value by $f_{\min}$, and returning $\succeq_{f_{\min}}$ as Kemeny ranking. Correctness then follows immediately from Lemma 1.

Egalitarian Kemeny rankings. An analogous result for the p-embeddable egalitarian Kemeny method can be obtained by an appropriate adaption of the objective function in the proof of Theorem 2.

Strict Preferences. Conversely whenever we restrict ourselves to instances in which all voters have only strict p-embeddable orders as preferences, we can focus on a proper minor of G_{pref} rather than the whole graph. More specifically we can restrict ourselves to the vertex set given by $\{v \in V(G_{\text{pref}}) \mid \exists R \in \mathcal{R} \quad v = v_R\}$; where edges between the vertices correspond to traversals of single hyperplanes: We contract paths of length 2 in G_{pref} between such vertices to single edges while summing up the weight of contracted edges. More explicitly instead of G_{pref} we can consider the graph H_{pref} given by the following information:

- $V(H_{\text{pref}}) = \{v_R \mid R \in \mathcal{R}\}$;
- $E(H_{\text{pref}}) = \{\{v_R, v_{R'}\} \mid \exists c, c' \in \mathcal{C} \quad \dim(R \cap R' \cap S_{c,c'}) = d - 1\}$; and
- $w : E(H_{\text{pref}}) \to \mathbb{N}$, $\{v_R, v_{R'}\} \mapsto 2|\{\{c, c'\} \subseteq \mathcal{C} \mid \dim(R \cap R' \cap S_{c,c'}) = d - 1\}|$.

Without weights, this graph is also known as the *region graph* or the *dual graph* of the embedded election induced hyperplane arrangement. Using the representation of the region graph as *medium*, i.e., as a system of *states* and transitions between states via *tokens*[23], we can employ a faster quadratic time all-pairs-shortest-paths algorithm [23] to achieve a better runtime for strict orders.

Theorem 3. *Determining all p-embeddable Kemeny rankings for a d-Euclidean election $(\mathcal{C}, \mathcal{V}, (\succ_v)_{v \in \mathcal{V}})$ in which all voters have strict preferences given by $p : \mathcal{C} \cup \mathcal{V} \to \mathbb{R}^d$ is possible in time in $\mathcal{O}(|\mathcal{C}|^{4d})$.*

6 Approximating the Kemeny Score

Our main algorithm fundamentally rests on the assumption that we are interested in an *embeddable* Kemeny ranking. As we have already seen in Example 1, such an embeddable Kemeny ranking may differ from an optimal Kemeny ranking. It is thus natural to ask

1. how often embeddable Kemeny rankings differ from optimal Kemeny rankings; and

2. how far these rankings can be apart (measured by their Kendall-tau distance).

We investigate these questions via numerical experiments and prove a bound on the worst-case approximation ratio of embeddable Kemeny rankings.

6.1 Approximation

Our goal is to quantify how much an embeddable Kemeny ranking and an optimal Kemeny ranking may differ. This can be phrased as an approximability results for computing Kemeny's voting rule in d-Euclidean elections. We can show that p-embeddable Kemeny rankings 2-approximate optimal Kemeny rankings.

Proposition 2. *Let $\prec$ be an optimal Kemeny ranking, and $\prec_{res}$ be a p-embeddable Kemeny ranking for a given embedding p. Then* $\frac{\sum_{v\in\mathcal{V}} \mathrm{K}(\prec_{res},\prec_v)}{\sum_{v\in\mathcal{V}} \mathrm{K}(\prec,\prec_v)} \leq 2$.

However, it is unclear whether our ratio 2 is tight (even for $d = 2$). The largest ratio we are aware of is $8/7$ and arises, e.g., in Example 1.

6.2 Experiments

We conducted numerical experiments on randomly generated 2-Euclidean elections to test the approximation quality of embeddable Kemeny rankings and to record how often embeddable Kemeny rankings do not achieve an optimal Kemeny score. In brief, our experiments suggest that the optimal Kemeny ranking is p-embeddable in 98.9% of the cases when considering up to 7 candidates.

To compute optimal Kemeny scores, we implemented Kemeny's rule with a trivial brute-force algorithm. The implementation for the p-embeddable Kemeny score used in these experiments[2] does not exploit all runtime improvements from the algorithm for strict orderings described in Section 5.3; its runtime currently inhibits experiments on larger instances. We randomly generated instances of 2-Euclidean elections with n voters, $3 \leq n \leq 15$, with strict preferences and m candidates, $4 \leq m \leq 7$, both of which we identify with points in $[0, 1000]^2$. For each pair (m, n), we generated 150 instances: 50 each assuming that (a) candidates and voters are component-wise uniformly distributed; that (b) candidates and voters are component-wise truncated normally distributed with mean 500 and variance 150; and that (c) candidates are uniformly distributed and voters are truncated normally distributed with mean 500 and variance 150.

In total, we ran 7800 tests; among them, only 84 exhibited a p-embeddable Kemeny ranking that differs from the optimal Kemeny ranking. In these 84 instances, the ratio r of embeddable and optimal Kemeny rankings is between 1.0077 and 1.11. A difference in the scores of the optimal and the p-embeddable

[2] We construct the preference graph H_{pref} by adapting the dual arrangement construction from CGAL (*The CGAL Project*, `https://www.cgal.org`) and apply Johnson's all-pairs shortest path algorithm to determine the p-embeddable Kemeny rankings.

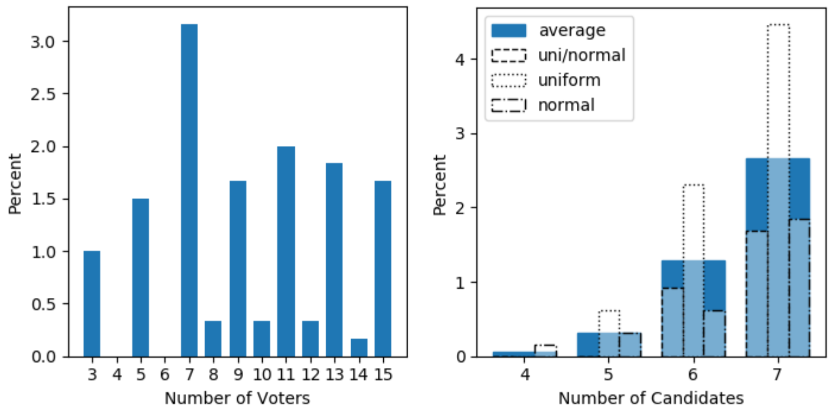


Fig. 4: Percentage of instances with ratio $r > 1$.

Kemeny rankings occurred slightly more often in uniformly distributed instances — 1.85% of uniformly distributed instances have ratios $r > 1$, which is the case for only $\approx 0.7\%$ for other distributions. Figure 4 gives an overview of the percentage of instances where $r > 1$. The results indicate that an increasing number of voters does not cause a significant rise in the numbers of instances with suboptimal p-embeddable Kemeny rankings. Interestingly, instances with an odd number of voters have suboptimal p-embeddable Kemeny rankings significantly more often (77 out of 84), possibly due to fewer ties. On the other hand, the results indicate a positive correlation between the number of candidates and the number of instances with suboptimal p-embeddable Kemeny ranking (for $m = 4$, there is only one of 1950 instances with $r > 1$ ($\approx 0.05\%$), while for $m = 7$, 52 instances out of 1950 admit ratio $r > 1$ ($\approx 2.66\%$)). This suggests that the low overall percentage is due to the choice of the candidate range. Further tests with a larger number of candidates remains—due to limited computational power and, in terms of runtime, suboptimal implementation of the p-embeddable Kemeny ranking computation—a point on our future agenda.

7 Conclusions and Open Problems

We have shown that p-embeddable Kemeny rankings can be computed in time in $\mathcal{O}(|\mathcal{C}|^{4d})$ for strict orders and $\tilde{\mathcal{O}}(|\mathcal{C}|^{4.746 \cdot d+2})$ for weak orders. Apart from improving these runtimes, it would be interesting to provide lower bounds on the computational complexity. In particular, a W-hardness result for computing p-embeddable Kemeny rankings could show that the dimension d has to occur in the exponent.

Further, our polynomial time solvability result juxtaposes the NP-hardness for the Kemeny Score problem on d-Euclidean elections, i.e., when one assumes p-embeddable preferences (given by p) but allows non-embeddable Kemeny rankings. To slightly relax our embeddability requirement on solutions with the hope of still remaining in P it would also be interesting to consider the problem where one requires a solution to be embeddable together with all voter preferences in the same dimension as the input, but allows the embedding to differ from the input embedding.

Let us end with a conceptual note. While d-Euclidean preferences are well-motivated and used in applications [21, 30, 31], there have been no successful attempts to leverage their structural properties for tractability results for $d \geq 2$, to the best of our knowledge. A likely reason for this is that combinatorial properties implied by d-Euclidean preferences seem to be difficult to derive. Our constructions of G_{pref} (and H_{pref} for strict preferences) in Section 5 may thus be of independent interest as a concise representation of d-Euclidean preferences and their mutual Kendall-tau distances under a fixed embedding. We would like to encourage the study of d-Euclidean preferences also for other computationally hard voting rules (such as Dodgson, Young).On this note, very recently many approval based multiwinner voting rules which are polynomial times solvable on 1-Euclidean elections were shown to be NP-hard on 2-Euclidean elections [25].

Acknowledgments. This work was supported by the Austrian Science Foundation FWF through projects P31890, P31336, W1255-N23, and Y1329.

Bibliography

[1] Ailon, N., Charikar, M., Newman, A.: Aggregating inconsistent information: ranking and clustering. Journal of the ACM **55**(5), 1–27 (2008)

[2] Aledo, J.A., Gámez, J.A., Molina, D.: Tackling the rank aggregation problem with evolutionary algorithms. Applied Mathematics and Computation **222**, 632–644 (2013)

[3] Ali, A., Meilă, M.: Experiments with Kemeny ranking: What works when? Mathematical Social Sciences **64**(1), 28–40 (2012)

[4] Alman, J., Williams, V.V.: A refined laser method and faster matrix multiplication. In: SODA. pp. 522–539. SIAM (2021)

[5] Altman, A., Tennenholtz, M.: Axiomatic foundations for ranking systems. J. Artif. Intell. Res. **31**, 473–495 (2008)

[6] Azzini, I., Munda, G.: A new approach for identifying the Kemeny median ranking. European Journal of Operational Research **281**(2), 388–401 (2020)

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

[8] Betzler, N., Fellows, M.R., Guo, J., Niedermeier, R., Rosamond, F.A.: Fixed-parameter algorithms for Kemeny rankings. Theoretical Computer Science **410**(45), 4554–4570 (2009)

[9] Biedl, T., Brandenburg, F.J., Deng, X.: On the complexity of crossings in permutations. Discrete Mathematics **309**(7), 1813–1823 (2009)

[10] Bogomolnaia, A., Laslier, J.F.: Euclidean preferences. Journal of Mathematical Economics **43**(2), 87–98 (February 2007)

[11] Brandt, F., Brill, M., Hemaspaandra, E., Hemaspaandra, L.A.: Bypassing combinatorial protections: Polynomial-time algorithms for single-peaked electorates. J. Artif. Intell. Res. **53**, 439–496 (2015)

[12] Cohen, M.B., Lee, Y.T., Miller, G., Pachocki, J., Sidford, A.: Geometric median in nearly linear time. In: STOC. pp. 9–21 (2016)

[13] Conitzer, V., Davenport, A., Kalagnanam, J.: Improved bounds for computing Kemeny rankings. In: AAAI. pp. 620–626 (2006)

[14] Cornaz, D., Galand, L., Spanjaard, O.: Kemeny elections with bounded single-peaked or single-crossing width. In: IJCAI. pp. 76–82 (2013)

[15] Davenport, A., Kalagnanam, J.: A computational study of the Kemeny rule for preference aggregation. In: AAAI. pp. 697–702 (2004)

[16] Doignon, J.P., Falmagne, J.C.: A polynomial time algorithm for unidimensional unfolding representations. J. Algorithms **16**(2), 218–233 (1994)

[17] Dwork, C., Kumar, R., Naor, M., Sivakumar, D.: Rank aggregation methods for the web. In: WWW. pp. 613–622 (2001)

[18] Edelsbrunner, H.: Algorithms in combinatorial geometry. Springer (2012)

[19] Eirinakis, P., Williamson, M., Subramani, K.: On the Shoshan-Zwick algorithm for the all-pairs shortest path problem. Journal of Graph Algorithms and Applications **21**(2), 177–181 (2017)

[20] Elkind, E., Lackner, M., Peters, D.: Structured preferences. In: Trends in Computational Social Choice, chap. 10, pp. 187–207. AI Access (2017)
[21] Enelow, J., Hinich, M. (eds.): Advances in the spatial theory of voting. Cambridge University Press (1990)
[22] Ephrati, E., Rosenschein, J.S., et al.: Multi-agent planning as a dynamic search for social consensus. In: IJCAI. pp. 423–429 (1993)
[23] Eppstein, D., Falmagne, J.C.: Algorithms for media. Discrete applied mathematics **156**(8), 1308–1320 (2008)
[24] Escoffier, B., Spanjaard, O., Tydrichova, M.: Kemeny ranking is NP-hard for 2-dimensional Euclidean preferences. arXiv:2106.13054, preprint (2021)
[25] Godziszewski, M.T., Batko, P., Skowron, P., Faliszewski, P.: An analysis of approval-based committee rules for 2D-Euclidean elections. In: AAAI. pp. 5448–5455 (2021)
[26] Goodman, J., ORourke, J., Tth, C.: Handbook of discrete and computational geometry, third edition (01 2017)
[27] Hemaspaandra, E., Spakowski, H., Vogel, J.: The complexity of Kemeny elections. Theoretical Computer Science **349**(3), 382 – 391 (2005)
[28] Kenyon-Mathieu, C., Schudy, W.: How to rank with few errors. In: STOC. pp. 95–103 (2007)
[29] Knoblauch, V.: Recognizing one-dimensional Euclidean preference profiles. Journal of Mathematical Economics **46**(1), 1–5 (2010)
[30] Laslier, J.F.: Spatial approval voting. Pol. Analysis **14**(2), 160–185 (2006)
[31] Londregan, J.: Estimating legislators' preferred points. Pol. Analysis **8**(1), 35–56 (1999)
[32] Pennock, D.M., Horvitz, E., Giles, C.L.: Social choice theory and recommender systems: Analysis of the axiomatic foundations of collaborative filtering. In: AAAI. pp. 729–734 (2000)
[33] Peters, D.: Recognising multidimensional euclidean preferences. In: AAAI. pp. 642–648 (2017)
[34] Peters, D., Lackner, M.: Preferences single-peaked on a circle. J. Artif. Intell. Res. **68**, 463–502 (2020)
[35] Popov, V.: Multiple genome rearrangement by swaps and by element duplications. Theoretical Computer Science **385**(1), 115–126 (2007)
[36] S Badal, P., Das, A.: Efficient algorithms using subiterative convergence for Kemeny ranking problem. Comp. Oper. Res. **98**, 198–210 (2018)
[37] Shoshan, A., Zwick, U.: All pairs shortest paths in undirected graphs with integer weights. In: FOCS. pp. 605–614. IEEE (1999)
[38] Szufa, S., Faliszewski, P., Skowron, P., Slinko, A., Talmon, N.: Drawing a map of elections in the space of statistical cultures. In: AAMAS. pp. 1341–1349 (2020)
[39] Young, H.P., Levenglick, A.: A consistent extension of condorcets election principle. SIAM Journal on applied Mathematics **35**(2), 285–300 (1978)
[40] Young, P.: Optimal voting rules. J. Econ. Persp. **9**(1), 51–64 (1995)