On Maximum Rank Aggregation Problems*

Christian Bachmaier, Franz J. Brandenburg,
Andreas Gleißner, and Andreas Hofmeier

University of Passau
94030 Passau, Germany
{bachmaier,brandenb,gleissner,hofmeier}@fim.uni-passau.de

Abstract. The rank aggregation problem consists in finding a consensus ranking on a set of alternatives, based on the preferences of individual voters. These are expressed by permutations, whose distance can be measured in many ways.

In this work we study a collection of distances, including the Kendall tau, Spearman footrule, Spearman rho, Cayley, Hamming, Ulam, and Minkowski distances, and compute the consensus against the maximum, which attempts to minimize the discrimination against any voter.

We provide a general schema from which we can derive the **NP**-hardness of the maximum rank aggregation problems under the aforementioned distances. This reveals a dichotomy for rank aggregation problems under the Spearman footrule and Minkowski distances: the common sum version is solvable in polynomial time whereas the maximum version is **NP**-hard. Moreover, the maximum rank aggregation problems are proved to be 2-approximable under all pseudometrics and fixed-parameter tractable under the Kendall tau, Hamming, and Minkowski distances.

1 Introduction

The task of ranking a list of alternatives is encountered in many situations. One of the underlying goals is to find the best consensus. This task is known as the rank aggregation problem, and was widely studied in the past decade [1, 13]. The problem has numerous applications in sports, voting systems for elections, search engines and evaluation systems on the web.

From mathematical and computational perspectives, the rank aggregation problem is given by a set of m permutations on a set of size n, and the goal is to find a consensus permutation with minimum distance to the given permutations. There are many ways to measure the distance between two permutations and to aggregate the cost by an objective function. Kemeny [19] proposed to count the pairwise disagreements between the orderings of two items, which is commonly known as the Kendall tau distance. For permutations it is the 'bubble sort' distance, i. e., the number of pairwise adjacent transpositions needed to transform one permutation into the other, or the number of crossings in a two-layered

* Supported by the Deutsche Forschungsgemeinschaft (DFG), grant Br835/16-1.

drawing [6]. Another popular measure is the Spearman footrule distance [11], which is the L_1-norm of two n-dimensional vectors.

The geometric median of the input permutations is commonly taken for the optimal aggregation, which means the *sum* of the cost of the comparison of each input permutation and the consensus. From the computational perspective this makes a difference between the Spearman footrule and the Kendall tau distance, since the further allows a polynomial time solution via weighted bipartite matching [13], whereas the latter leads to an **NP**-hard rank aggregation problem [3], even for four voters [6,13]. It has an expected $\frac{11}{7}$ randomization [2], a PTAS [20], and is fixed-parameter tractable [5,18].

Here we study the *maximum version*, which attempts to avoid a discrimination of a single voter or permutation against the consensus. The objective is a minimum k such that all permutations are within distance k from the consensus. Biedl et al. [6] studied this version for the Kendall tau distance and showed that determining whether there is a permutation τ which is within distance at most k to all input permutations, is **NP**-hard, even for any $m \geq 4$ permutations.

There are other distance measures on permutations than the Kendall tau and the Spearman footrule distances. These can be derived from steps in sorting algorithms. In their fundamental study Diaconis and Graham [11] relate the Kendall tau and Spearman footrule distance, and the Spearman rho and Cayley distance. Critchlow [9] added the Hamming and edit distances.

Our main contribution is a general schema for the complexity analysis, which allows us to prove that the maximum rank aggregation problem is **NP**-hard and fixed-parameter tractable under any metric d which satisfies some requirements. These are granted by the aforementioned distances. For the **NP**-hardness results we provide a simpler reduction from the Closest Binary String problem and from the Hitting String problem. Previous reductions used the Feedback Arc Set problem (see [6,13]).

The paper is organized as follows. After some preliminaries in Sect. 2 we show in Sect. 3 that Maximum Ranking (MR) is tractable under the Maximum distance, whereas MR is intractable under many other distances as shown in Sect. 4. In Sect. 5 we establish that MR is 2-approximable for pseudometrics. Finally, in Sect. 6, we present fixed-parameter algorithms to solve MR under various distances.

2 Preliminaries

For a binary relation R on a domain $\mathcal{D}$ and for each $x, y \in \mathcal{D}$, we write $x <_R y$ if $(x, y) \in R$ and $x \not<_R y$ if $(x, y) \notin R$. A binary relation κ is a (strict) *partial order* if it is *irreflexive*, *asymmetric* and *transitive*, i. e., $x \not<_\kappa x$, $x <_\kappa y \Rightarrow y \not<_\kappa x$, and $x <_\kappa y \wedge y <_\kappa z \Rightarrow x <_\kappa z$ for all $x, y, z \in \mathcal{D}$. Candidates x and y are called *unrelated by* κ if $x \not<_\kappa y \wedge y \not<_\kappa x$, which we denote by $x \not\gtrless_\kappa y$. The intuition of $x <_\kappa y$ is that κ *ranks* x *before* y, which means a preference for x. If $x <_\kappa y$ or $y <_\kappa x$, we speak of a *constraint of* κ *on* x *and* y. For $\mathcal{X}, \mathcal{Y} \subseteq \mathcal{D}$ we denote $\mathcal{X} <_\kappa \mathcal{Y}$ if $\underset{x \in \mathcal{X}}{\forall} \underset{y \in \mathcal{Y}}{\forall} x <_\kappa y$ and define $x <_\kappa \mathcal{Y}$ and $\mathcal{X} <_\kappa y$ accordingly.

A *total order* is a *complete* partial order, i. e., $x <_\tau y \vee y <_\tau x$ for all $x, y \in \mathcal{D}$ with $x \neq y$. Let $n = |\mathcal{D}|$ and $\underline{n} = \{1, \ldots, n\}$. For every total order τ there is a unique *permutation*, i. e., a bijection $\tau' : \mathcal{D} \to \underline{n}$ such that $x <_\tau y \Leftrightarrow \tau'(x) < \tau'(y)$. In the rest of the paper we identify total orders and their corresponding permutations, taking the view whichever comes in more handy. The set of all permutations on $\mathcal{D}$ is denoted by $\mathrm{Perm}(\mathcal{D})$. We denote the permutation $\{x_1, \ldots, x_n\} \to \underline{n} : x_i \mapsto i$ by $[x_1 x_2 \ldots x_n]$.

A total order $\tau \in \mathrm{Perm}(\mathcal{D})$ is a *total extension* of a partial order κ if τ does not contradict κ, i. e., $x <_\kappa y$ implies $x <_\tau y$ for all $x, y \in \mathcal{D}$. We denote the set of total extensions of a partial order κ by $\mathrm{Ext}(\kappa)$.

A *bucket order* is a partial order κ for which unrelatedness $\ngtrless_\kappa$ is transitive. Then $\ngtrless_\kappa$ is an equivalence relation whose equivalence classes are called *buckets*. In other words, κ induces a total order order on the buckets while candidates of the same bucket are unrelated, see [1, 2, 15].

A *transposition* is a permutation on $\underline{n}$ switching the positions of two candidates. Hence, for positions $i, j \in \underline{n}$, we define the transposition $T_{i,j} \in \mathrm{Perm}(\underline{n})$ by $T_{i,j}(i) = j$, $T_{i,j}(j) = i$ and $T_{i,j}(k) = k$ for $k \notin \{i, j\}$. Transpositions can also be considered as operations acting on permutations on $\mathcal{D}$. For $x, y \in \mathcal{D}$ and $\sigma \in \mathrm{Perm}(\mathcal{D})$ we say $T_{\sigma(x),\sigma(y)} \circ \sigma \in \mathrm{Perm}(\mathcal{D})$ is the *transposition of x and y in* σ. Transpositions $T_{i,j}$ of adjacent candidates with $|i - j| = 1$ are called *swaps*.

A binary function $d : \mathrm{Perm}(\mathcal{D}) \times \mathrm{Perm}(\mathcal{D}) \to \mathbb{R}$ is called a *pseudometric* if $d(\sigma, \tau) \geq 0$, $d(\sigma, \tau) = d(\tau, \sigma)$, $\sigma = \tau \Rightarrow d(\sigma, \tau) = 0$, and $d(\sigma, \tau) + d(\tau, \rho) \geq d(\sigma, \rho)$ for all $\sigma, \tau, \rho \in \mathrm{Perm}(\mathcal{D})$. It is a *metric* if, additionally, $\sigma = \tau \Leftrightarrow d(\sigma, \tau) = 0$.

Next we introduce the main concepts of this work: The maximum version of the rank aggregation problem under various distances [9, 10].

Definition 1 (Maximum Ranking (MR)).
Instance: A set $\mathcal{D}$ of n candidates, m voters $\sigma_1, \ldots, \sigma_m \in \mathrm{Perm}(\mathcal{D})$, $k \in \mathbb{N}$.
Question: Does there exist a permutation $\tau \in \mathcal{D}$ with $\max_{j=1}^m d(\sigma_j, \tau) \leq k$?

Then permutation τ is called *k-consensus*. Observe that this is equivalent to say that $d(\sigma_j, \tau) \leq k$ for all voters σ_j, $j \in \underline{m}$.

Let $\sigma, \tau \in \mathrm{Perm}(\mathcal{D})$. Define the set of *dirty* pairs $\mathcal{K}(\sigma, \tau) = \{\{x, y\} \subseteq \mathcal{D} : x <_\sigma y \wedge y <_\tau x\}$ as the set of pairs of candidates $x, y \in \mathcal{D}$ where σ and τ disagree on their order. Then the *Kendall tau distance* K is defined by $K(\sigma, \tau) = |\mathcal{K}(\sigma, \tau)|$. It coincides with the minimum number k of swaps $T_1, \ldots, T_k$ such that $\tau = T_k \circ \ldots \circ T_1 \circ \sigma$. If we also allow switching non-adjacent candidates, we obtain the *Cayley distance* $C(\sigma, \tau)$, which is the minimum number of transpositions $T_1, \ldots, T_k$ such that $\tau = T_k \circ \ldots \circ T_1 \circ \sigma$. A permutation on $\underline{n}$ can also be specified by its constituent cycles. A cycle $\mathcal{C} = (x_1 x_2 \ldots x_{|\mathcal{C}|})$ of $\rho \in \mathrm{Perm}(\underline{n})$ is a (cyclic) sequence of distinct candidates such that $\rho(x_i) = x_{i+1}$ for $1 \leq i < |\mathcal{C}|$ and $\rho(x_{|\mathcal{C}|}) = x_1$. The cycles form a partition of $\underline{n}$. Denote by $\sharp\mathcal{C}(\rho)$ the number of cycles of ρ. The Cayley distance can be expressed as $C(\sigma, \tau) = n - \sharp\mathcal{C}(\tau \circ \sigma^{-1})$ [10].

Define the set of *displaced* candidates by $\mathcal{H}(\sigma, \tau) = \{x \in \mathcal{D} : \sigma(x) \neq \tau(x)\}$ as the set of candidates $x \in \mathcal{D}$ where σ and τ disagree on their position. The *Hamming distance* H is defined by $H(\sigma, \tau) = |\mathcal{H}(\sigma, \tau)|$, which is the number of positions $i \in \underline{n}$ where $\sigma^{-1}(i) \neq \tau^{-1}(i)$. This view is also taken by the

Hamming distance between strings $s, t \in \{0,1\}^n$, which is defined as $H(s,t) = |\{i \in \underline{n} : s(i) \neq t(i)\}|$ where $s(i)$ denotes the i-th character of s.

Let $\sigma, \tau \in \mathrm{Perm}(\mathcal{D})$. A tuple $(x_1, \ldots, x_l)$ with $x_i \in \mathcal{D}$ is a *common subsequence* of σ and τ if $i < j \Leftrightarrow x_i <_\sigma x_j \wedge x_i <_\tau x_j$. Let $\mathrm{lcs}(\sigma,\tau) = \max\{\, l\colon (x_1, \ldots, x_l)$ is a common subsequence of σ and $\tau\,\}$. Then the *Ulam distance* is $U(\sigma,\tau) = n - \mathrm{lcs}(\sigma,\tau)$.

Finally, the *Minkowski distance* F_p is defined as $F_p(\sigma,\tau) = \left(\sum_{x \in \mathcal{D}} |\sigma(x) - \tau(x)|^p\right)^{\frac{1}{p}}$ for $p \in \mathbb{N} \setminus \{0\}$. F_1 is also known as the Spearman Footrule distance or taxicab metric. F_2 is the Euclidean metric and also known as the Spearman rho distance [10]. To simplify proofs we introduce the notion of the *raised Minkowski distance* $\hat{F}_p$ defined by $\hat{F}_p(\tau,\sigma) = (F_p(\tau,\sigma))^p = \sum_{x \in \mathcal{D}} |\tau(x) - \sigma(x)|^p$.

One can also consider the limit for $p \to \infty$ and $p \to -\infty$. The Chebyshev or *Maximum distance* is $F_\infty(\sigma,\tau) = \max_{x \in \mathcal{D}} |\sigma(x) - \tau(x)|$. Define the *Minimum distance* $F_{-\infty}(\sigma,\tau) = \min_{x \in \mathcal{D}} |\sigma(x) - \tau(x)|$. Note that $F_{-\infty}$ is not a metric and satisfies only non-negativity and symmetry.

3 Efficient Algorithms

Theorem 1. *MR is efficiently solvable under the Maximum distance F_∞.*

Proof. To find a permutation τ satisfying $\max_{j=1}^m \max_{x \in \mathcal{D}} |\sigma_j(x) - \tau(x)| \leq k$, we solve a maximum matching problem in the bipartite graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with vertices $\mathcal{V} = \mathcal{D} \cup \underline{n}$ and an edge $(x,i) \in \mathcal{E}$ if $\max_{j=1}^m |\sigma_j(x) - i| \leq k$. Every matching of size n corresponds to a k-consensus τ and vice versa. As $|\mathcal{E}| < n(2k+1)$, this can be done in $\mathcal{O}(n^2 \cdot k)$ time. For an improvement observe that the suitable positions for each candidate are consecutive, thus form an interval. Assign to each candidate $x \in \mathcal{D}$ the interval $I_x = \{i \in \underline{n} : \max_{j=1}^m |\sigma_j(x) - i| \leq k\}$. Then iterate over the positions $i \in \underline{n}$. In step i, select the candidate to place at position i. Choose from those candidates x with $i \in I_x$ and which have not been placed before. If there are multiple suitable candidates, prefer a candidate whose interval has the least upper endpoint. In the case that there are no suitable candidates, reject the instance. We use a heap to manage the intervals of unplaced candidates, inserting the interval once we reach its lower endpoint. Determining the endpoints of the intervals can be done in $\mathcal{O}(n \cdot m)$ and the iteration is done in $\mathcal{O}(n \log n)$, resulting in a total running time of $\mathcal{O}(n(\log n + m))$. □

4 Intractability Results

We show that MR is **NP**-complete under the Hamming, Minkowski, Kendall tau, Cayley, Ulam and the Minimum distances. As these distances can be efficiently computed between total orders [4,6,9,21], membership is in **NP**. For the **NP**-hardness proofs we develop a general schema. First we proof that the **NP**-complete Closest Binary String problem [16] can be reduced to a special case of MR under any metric subject to Requirements 1 and 2 defined below. Then we show that these requirements are satisfied by all of the aforementioned metrics except the Minimum distance, for which we provide a reduction from the **NP**-complete Hitting String problem [14].

Definition 2 (Closest Binary String [16]).
Instance: $k, n \in \mathbb{N}$, *a list* $s_1, \ldots, s_m \in \{0,1\}^n$ *of* m *binary strings of length* n.
Question: Does there exist a string $t \in \{0,1\}^n$ *with* $\max_{j=1}^m H(s_j, t) \leq k$?

For the rest of this section, we introduce distinct elements a_i, b_i and sets $\mathcal{B}_i = \{a_i, b_i\}$ for $i \in \underline{n}$ and let $\mathcal{D} = \bigcup_{i=1}^n \mathcal{B}_i$. Let κ be the bucket order on $\mathcal{D}$ with buckets $\mathcal{B}_i$ ordered by $\mathcal{B}_1 <_\kappa \ldots <_\kappa \mathcal{B}_n$. We call a permutation *local* if it is an extension of κ. We state the following properties to be met by a metric d in order to be applicable in the forthcoming reduction.

Requirement 1 (Optimality of local permutations). *Let* $\sigma_1, \ldots, \sigma_m \in \mathrm{Ext}(\kappa)$ *and* $k \in \mathbb{N}$. *If there is a* k*-consensus* $\tau \in \mathrm{Perm}(\mathcal{D})$ *with* $\max_{j=1}^m d(\sigma_j, \tau) \leq k$, *then there also is a local permutation* $\tau' \in \mathrm{Ext}(\kappa)$ *with* $\max_{j=1}^m d(\sigma_j, \tau') \leq k$.

In other words, if all voters are local and our metric meets Requirement 1, then we can safely demand that the consensus is local, too, without impairing its chance to satisfy the upper bound k. Note that d satisfies Requirement 1 if for every $\sigma \in \mathrm{Ext}(\kappa)$ and $\tau \in \mathrm{Perm}(\mathcal{D})$ we can find $\tau' \in \mathrm{Ext}(\kappa)$ such that $d(\tau', \sigma) \leq d(\tau, \sigma)$. The second requirement puts tight constraints on the distance of local permutations.

Requirement 2 (Distance constraints). *There is a constant* $c > 0$ *such that for all local permutations* $\sigma, \tau \in \mathrm{Ext}(\kappa)$ *the distance is* $d(\sigma, \tau) = c \cdot |\mathcal{K}(\sigma, \tau)|$.

Note that all local permutations agree on the order of candidates from different buckets. Thus, a distance satisfying Requirement 2 is exactly a constant multiple of the number of buckets $\mathcal{B}_i$ where one permutation ranks a_i before b_i and the other ranks b_i before a_i.

Theorem 2. *MR under a metric* d *is* **NP***-hard if* d *satisfies Requirements 1 and 2.*

Proof. Consider an instance of Closest Binary String consisting in a list $s_1, \ldots, s_m \in \{0,1\}^n$ of m binary strings of length n and an upper bound $k \in \mathbb{N}$ as in Definition 2. We choose the candidate set $\mathcal{D}$ as defined above. Consider the bijective mapping $f : \{0,1\}^n \to \mathrm{Ext}(\kappa)$, which encodes strings of length n as a local permutation where $a_i <_{f(s)} b_i$ if $s(i) = 0$ and $b_i <_{f(s)} a_i$ if $s(i) = 1$. More formally, $f(s)(a_i) = 2i - 1 + s(i)$ and $f(s)(b_i) = 2i - s(i)$ for all strings $s \in \{0,1\}^n$. For instance, $f(\texttt{"010"}) = [\underbrace{a_1 b_1}_{\mathcal{B}_1} \underbrace{b_2 a_2}_{\mathcal{B}_2} \underbrace{a_3 b_3}_{\mathcal{B}_3}]$. Observe that for strings $s, t \in \{0,1\}^n$ and $i \in \underline{n}$ we have $s(i) \neq t(i)$ if and only if $\{a_i, b_i\} \in \mathcal{K}(f(s), f(t))$. For each string s_j we introduce the voter $\sigma_j = f(s_j)$ and let $k' = c \cdot k$, where c is the constant from Requirement 2.

Suppose that a string $t^* \in \{0,1\}^n$ satisfies $\max_{j=1}^m H(s_j, t^*) \leq k$. Let $j \in \underline{m}$. We have

$$\begin{aligned} k' = c \cdot k \geq c \cdot H(s_j, t^*) &= c \cdot |\{i \in \underline{n} : s_j(i) \neq t^*(i)\}| = c \cdot |\mathcal{K}(f(s_j), f(t^*))| \\ &= d(\sigma_j, f(t^*)) \end{aligned}$$

by Requirement 2. Therefore, $f(t^*)$ is a k'-consensus for the MR problem.

Conversely suppose that τ^* satisfies $\max_{j=1}^m d(\sigma_j, \tau^*) \leq k'$. W. l. o. g. assume that τ^* is local by Requirement 1. Again, let $j \in \underline{m}$. By Requirement 2 we obtain

$$\begin{aligned} k = \frac{k'}{c} \geq \frac{1}{c} \cdot d(\sigma_j, \tau^*) &= |\mathcal{K}(\sigma_j, \tau^*)| = \left|\{i \in \underline{n} : f^{-1}(\sigma_j) \neq f^{-1}(\tau^*)\}\right| \\ &= H(s_j, f^{-1}(\tau^*))\,, \end{aligned}$$

i. e., the string $t^* = f^{-1}(\tau^*) \in \{0,1\}^n$ satisfies $\max_{j=1}^m H(s_j, t^*) \leq k$. □

Lemma 1. *Let $\sigma, \tau \in \mathrm{Perm}(\mathcal{D})$ and $\{x, y\} \in \mathcal{K}(\sigma, \tau)$ be a dirty pair between σ and τ. Then the Kendall tau distance strictly decreases if we transpose x and y in τ, i. e., $K(\sigma, T_{\tau(x),\tau(y)} \circ \tau) < K(\sigma, \tau)$.*

Proof. Let $\tau' = T_{\tau(x),\tau(y)} \circ \tau$. W. l. o. g. assume $x <_\tau y$. We compare the set $\mathcal{K}^+ = \mathcal{K}(\tau', \sigma) \setminus \mathcal{K}(\tau, \sigma)$ with the set $\mathcal{K}^- = \mathcal{K}(\tau, \sigma) \setminus \mathcal{K}(\tau', \sigma)$. Then $K(\tau', \sigma) < K(\tau, \sigma)$ if $|\mathcal{K}^+| < |\mathcal{K}^-|$. Now, let $Z_<, Z_|$ and $Z_>$ be the candidates that are ranked by σ before, between, and after x and y, respectively. Formally, $Z_< = \{z \in \mathcal{D} : x <_\tau z <_\tau y \wedge z <_\sigma y <_\sigma x\}$, $Z_| = \{z \in \mathcal{D} : x <_\tau z <_\tau y \wedge y <_\sigma z <_\sigma x\}$, and $Z_> = \{z \in \mathcal{D} : x <_\tau z <_\tau y \wedge y <_\sigma x <_\sigma z\}$. By a simple but cumbersome distinction of cases we obtain

$$\begin{aligned} \mathcal{K}^+ &= \bigcup_{z \in Z_<} \{\{y, z\}\} \cup \bigcup_{z \in Z_>} \{\{x, z\}\}\,, \text{ and} \\ \mathcal{K}^- &= \bigcup_{z \in Z_<} \{\{x, z\}\} \cup \bigcup_{z \in Z_>} \{\{y, z\}\} \cup \bigcup_{z \in Z_|} \{\{x, z\}, \{y, z\}\} \cup \{\{x, y\}\}\,. \end{aligned}$$

Hence, $K(\tau', \sigma) = K(\tau, \sigma) - \left|Z_|\right| - 1$. □

Next we show that Requirements 1 and 2 hold for the Kendall tau, Cayley, Hamming, Ulam, and Minkowski distances.

Lemma 2. *Let τ^* be an optimal consensus for the MR problem under the Kendall tau distance K with voters $\sigma_1, \ldots, \sigma_m$. Let $\mu = \bigcap_{j=1}^m \sigma_j$ be the partial order with $x <_\mu y$ if and only if $x <_{\sigma_j} y$ for all $j \in \underline{m}$. Then $\tau^* \in \mathrm{Ext}(\mu)$.*

Proof. Assume by contradiction that there are candidates $x, y \in \mathcal{D}$ with $x <_\mu y$ and $y <_{\tau^*} x$. Then $x <_{\sigma_j} y$ and $\{x, y\} \in \mathcal{K}(\sigma_j, \tau^*)$ for every $j \in \underline{m}$. Thus, $\max_{j=1}^m d(\sigma_j, T_{\tau^*(x),\tau^*(y)} \circ \tau^*) < \max_{j=1}^m d(\sigma_j, \tau^*)$ by Lemma 1, which is a contradiction to the optimality of τ^*. □

Corollary 1. *The Kendall tau distance K satisfies Requirements 1 and 2.*

Proof. Let $\sigma_1, \ldots, \sigma_m \in \mathrm{Ext}(\kappa)$ be local permutations and $\mu = \bigcap_{j=1}^{m} \sigma_j$. Every extension of μ is also an extension of κ since $\kappa \subseteq \mu$. Hence, Requirement 1 follows immediately from Lemma 2. Let $c = 1$. Then Requirement 2 is just the definition of the Kendall tau distance restricted to local permutations. $\square$

Lemma 3. *The Cayley distance C satisfies Requirement 2.*

Proof. Let $\sigma, \tau \in \mathrm{Ext}(\kappa)$ be local permutations. Since σ and τ agree on the order of candidates in different buckets, $\mathcal{K}(\sigma, \tau) \subseteq \{\, \mathcal{B}_i : i \in \underline{n} \,\}$. Consider a bucket $\mathcal{B}_i = \{a_i, b_i\}$. If $\mathcal{B}_i \in \mathcal{K}(\sigma, \tau)$, then a_i and b_i form a single cycle $(a_i b_i)$ of size 2 in $\tau \circ \sigma^{-1}$ as $\sigma(a_i) = \tau(b_i)$ and $\sigma(b_i) = \tau(a_i)$. If otherwise $\mathcal{B}_i \notin \mathcal{K}(\sigma, \tau)$, a_i and b_i each form a cycle of size 1. Thus, $C(\sigma, \tau) = 2n - \sharp\mathcal{C}(\tau \circ \sigma^{-1}) = 2n - |\mathcal{K}(\sigma, \tau)| - 2 \cdot |\{\, \mathcal{B}_i : i \in \underline{n} \,\} \setminus \mathcal{K}(\sigma, \tau)| = |\mathcal{K}(\sigma, \tau)| = K(\sigma, \tau)$. $\square$

Lemma 4. *The Cayley distance C satisfies Requirement 1, i. e., $C(l(\tau), \sigma) \leq C(\tau, \sigma)$ for every $\sigma \in \mathrm{Ext}(\kappa)$ and $\tau \in \mathrm{Perm}(\mathcal{D})$.*

Proof. For $x \in \mathcal{D}$, denote by $\left|\mathcal{C}_x(\tau \circ \sigma^{-1})\right|$ the size of the cycle in $\tau \circ \sigma^{-1}$ containing x. If $\sigma(x) \neq \tau(x)$, then $\left|\mathcal{C}_x(\tau \circ \sigma^{-1})\right| \geq 2$, but $\left|\mathcal{C}_x(l(\tau) \circ \sigma^{-1})\right| \leq 2$ as shown in the proof of Lemma 3. If otherwise $\sigma(x) = \tau(x)$, then $\left|\mathcal{C}_x(\tau \circ \sigma^{-1})\right| = \left|\mathcal{C}_x(l(\tau) \circ \sigma^{-1})\right| = 1$. Observe that $\sum_{x \in \mathcal{D}} \frac{1}{|\mathcal{C}_x(\tau \circ \sigma^{-1})|} = \sharp\mathcal{C}(\tau \circ \sigma^{-1})$. Hence, $\sharp\mathcal{C}(\tau \circ \sigma^{-1}) \leq \sharp\mathcal{C}(l(\tau) \circ \sigma^{-1})$. $\square$

Proposition 1. *Let $\sigma, \tau \in \mathrm{Perm}(\mathcal{D})$ and $x \in \mathcal{H}(\sigma, \tau)$ be a displaced candidate. Then $H(\sigma, T_{\sigma(x),\tau(x)} \circ \tau) < H(\sigma, \tau)$.*

Proof. Let $y \in \mathcal{D}$ such that $\tau(y) = \sigma(x)$. Note that $y \in \mathcal{H}(\sigma, \tau)$ and the transposition of x and y in τ does not affect other candidates. Thus, $\mathcal{H}(\sigma, T_{\sigma(x),\tau(x)} \circ \tau) = \mathcal{H}(\sigma, \tau) \setminus \{x\}$ or even $\mathcal{H}(\sigma, T_{\sigma(x),\tau(x)} \circ \tau) = \mathcal{H}(\sigma, \tau) \setminus \{x, y\}$ if $\tau(x) = \sigma(y)$. $\square$

Lemma 5. *If $p \in \mathbb{N} \setminus \{0\}$, then the raised Minkowski distance $\hat{F}_p$ satisfies Requirement 1, i. e., $\hat{F}_p(l(\tau), \sigma) \leq \hat{F}_p(\tau, \sigma)$ for every $\sigma \in \mathrm{Ext}(\kappa)$ and $\tau \in \mathrm{Perm}(\mathcal{D})$.*

Proof. Let $x \in \mathcal{D}$. If $\tau(x) = \sigma(x)$ then $l(\tau)(x) = \sigma(x)$ since $x \notin \mathcal{A}_\tau$. Otherwise, $|\tau(x) - \sigma(x)| \geq 1$ implies $|\tau(x) - \sigma(x)|^p \geq 1$, but $|l(\tau)(x) - \sigma(x)| \leq 1$. In both cases $|l(\tau)(x) - \sigma(x)|^p \leq |\tau(x) - \sigma(x)|^p$. $\square$

Proposition 2. *MR under the raised Minkowski distance $\hat{F}_p$ for $p \in \mathbb{N} \setminus \{0\}$ and under the Hamming distance H satisfies Requirement 2.*

Proof. Let $\sigma, \tau \in \mathrm{Ext}(\kappa)$ be local permutations. Recall that $\mathcal{K}(\sigma, \tau) \subseteq \{\, \mathcal{B}_i : i \in \underline{n} \,\}$ as σ and τ agree on the order of candidates in different buckets. Hence, $|\tau(x) - \sigma(x)| = |\tau(y) - \sigma(y)| = 1$ for every bucket $\{x, y\} \in \mathcal{K}(\sigma, \tau)$, i. e., both x and y contribute 1 to the distance. Members of the remaining buckets $\{x, y\} \in \{\, \mathcal{B}_i : i \in \underline{n} \,\} \setminus \mathcal{K}(\sigma, \tau)$ contribute 0. $\square$

By a similar proof we obtain:

Lemma 6. *The Ulam distance U satisfies Requirement 2.*

For the proof of the following lemma we define the *refinement* of a bucket by a total order as in [7,15]. The *refinement* of a bucket order κ by a total order τ is the total order $\tau * \kappa$ such that $x <_{\tau * \kappa} y \Leftrightarrow x <_\kappa y \vee x \ngtrless_\kappa y \wedge x <_\tau y$ holds for all $x, y \in \mathcal{D}$. Note that $\tau * \kappa \in \mathrm{Ext}(\kappa)$.

Lemma 7. *The Ulam distance U satisfies Requirement 1, i. e., $U(\tau * \kappa, \sigma) \leq U(\tau, \sigma)$ for every $\sigma \in \mathrm{Ext}(\kappa)$ and $\tau \in \mathrm{Perm}(\mathcal{D})$.*

Proof. Let $\sigma \in \mathrm{Ext}(\kappa)$, $\tau \in \mathrm{Perm}(\mathcal{D})$, and $(x_1, \ldots, x_l)$ be a *longest* common subsequence of τ and σ, i. e., $l = \mathrm{lcs}(\sigma, \tau)$. As $\sigma \in \mathrm{Ext}(\kappa)$, all elements $x_1, \ldots, x_l$ are ordered by both σ and τ according to κ. Hence, $(x_1, \ldots, x_l)$ is also a common subsequence for $\tau * \kappa$ and σ and thus, $\mathrm{lcs}(\tau * \kappa, \sigma) \geq \mathrm{lcs}(\tau, \sigma)$. □

Theorem 3. *Requirements 1 and 2 are satisfied by the Kendall tau, Cayley, Hamming, Ulam, and Minkowski distances F_p for $p \in \mathbb{N} \setminus \{0\}$. Thus, MR is* **NP***-complete under these distances.*

In consequence we have a dichotomy between the sum and the maximum versions of the rank aggregation problem, in particular for the Spearman footrule distance.

Corollary 2. *For the Minkowski distances F_p and $p \in \mathbb{N} \setminus \{0\}$ (i) the common rank aggregation problem taking the sum is efficiently solvable, and (ii) the maximum rank aggregation problem MR is* **NP***-complete.*

Proof. The common rank aggregation problem can be solved by weighted bipartite matching, where the weights $w_{x,i}$ express the cost of placing x at position i [13], and (ii) follows from Theorem 3. □

Since the Minimum distance does not satisfy Requirement 2, we provide a different reduction from Hitting String.

Definition 3 (Hitting String [14]).
Instance: $n \in \mathbb{N}$, a list $s_1, \ldots, s_m \in \{0, 1, *\}^n$ of m strings of length n.
Question: Does there exist a string $t \in \{0, 1\}^n$ such that every string s_j is hit *by t in at least one position, i. e., $\forall j \in \underline{m} : \exists i \in \underline{n} : s_j(i) = t(i)$.*

Theorem 4. *MR under the Minimum distance $F_{-\infty}$ is* **NP***-complete even for $k = 0$.*

Proof. There is a consensus $\tau \in \mathrm{Perm}(\mathcal{D})$ with $\max_{j=1}^{m} \min_{x \in \mathcal{D}} |\sigma_j(x) - \tau(x)| = 0$ if and only if for every σ_j there is a candidate x such that $\sigma_j(x) = \tau(x)$. Then τ *hits* σ_j *at position* $\tau(x)$ and we call τ *hitting consensus*.

First we show how to construct an instance with $2n$ voters of length n which has no hitting consensus. Let $\mathcal{D} = \{u_1, \ldots, u_n\}$ and $\sigma_1 : \mathcal{D} \to \underline{n} : u_i \mapsto i$. We obtain n *primary voters* $\sigma_1, \ldots, \sigma_n$ by rotating σ_1, i. e., for every $j \in \underline{n}$ let $\sigma_j(u_i) = (i + j - 2) \mod n + 1$. Additionally, we introduce the *secondary voters* $\sigma'_1, \ldots, \sigma'_n$ defined by $\sigma'_j = T_{1,2} \circ \sigma_j$. For instance if $\mathcal{D} = \{a, b, c, d, e\}$, then the list of voters is

$$\begin{array}{ll} \sigma_1 = [abcde] & \sigma'_1 = [bacde] \\ \sigma_2 = [eabcd] & \sigma'_2 = [aebcd] \\ \sigma_3 = [deabc] & \sigma'_3 = [edabc] \\ \sigma_4 = [cdeab] & \sigma'_4 = [dceab] \\ \sigma_5 = [bcdea] & \sigma'_5 = [cbdea] . \end{array}$$

Assume for contradiction that this list of voters has a hitting consensus τ. Since there are n primary voters and no two primary voters place any candidate at the same position, every primary voter is hit at exactly one position and τ hits exactly one primary voter at position 1. Let σ be the primary voter hit at position 1 by a candidate x. Then τ cannot hit the secondary voter $\sigma' = T_{1,2} \circ \sigma$ at the positions 1 or 2 as $\tau(x) = \sigma(x) = 1 \neq 2 = \sigma'(x)$. Thus, it cannot hit σ' at all since σ and σ' agree in all other positions $\underline{n} \setminus \{1, 2\}$, a contradiction. We call the above list of voters the *n-anti-pattern*. With this in mind, we reduce from the **NP**-complete Hitting String to MR under the Minimum distance.

As in the proof of Theorem 2, let $\mathcal{D} = \bigcup_{i=1}^{n} \{a_i, b_i\}$ be the set of candidates and let $f : \{0,1\}^n \to \mathrm{Perm}(\mathcal{D})$ with $f(s)(a_i) = 2i - 1 + s(i)$ and $f(s)(b_i) = 2i - s(i)$. For every string s_j, $j \in \underline{m}$, we introduce a list of voters Σ_j in two steps. The instance of MR is then the concatenation of all $\Sigma_j, j \in \underline{m}$ and $k = 0$. In the first step we create a *template* $\rho_j : \mathcal{D} \to \underline{n} \cup \{*\}$ from which the actual list is obtained in the second step. Let $\rho_j(a_i) = f(s_j)(a_i)$ and $\rho(b_i) = f(s_j)(b_i)$ if $s(i) \in \{0, 1\}$ and $\rho_j(a_i) = \rho_j(b_j) = *$, otherwise. If none of the strings s_j did contain $*$, then we could establish a one-to-one correspondence between a hitting consensus for voters $\rho_1, \ldots, \rho_m$ and a hitting string for $s_1, \ldots, s_m$ as in Theorem 2 and would be done. Let $\mathcal{U}_j = \{x \in \mathcal{D} : \rho_j = *\}$ be the set of candidates which are not assigned a position by ρ_j. In Hitting String the $*$ marks a position where an input string cannot be hit however the hitting string looks alike. We reproduce this situation for MR by making $2\,|\mathcal{U}_j|$ copies $\sigma_j^{(1)}, \ldots, \sigma_j^{(2|\mathcal{U}_j|)}$ of ρ_j such that all copies agree on the candidates $\mathcal{D} \setminus \mathcal{U}_j$ but form a $|\mathcal{U}_j|$-anti-pattern if the candidate set is restricted to $\mathcal{U}_j$.

Suppose that t^* is a hitting string for $s_1, \ldots, s_m$. Then $f(t^*)$ is a hitting consensus since for every $j \in \underline{m}$ there is an $i \in \underline{n}$ with $t^*(i) = s_j(i)$, thus $f(t^*)(a_i) = \sigma_j(a_i)$. Conversely suppose that τ^* is a hitting consensus. Consider the string $t^* \in \{0,1\}^n$ defined by $t^*(i) = 0$ if $\tau^*(a_i) = 2i - 1 \vee \tau^*(b_i) = 2i$ and $t^*(i) = 1$, otherwise. For every $j \in \underline{m}$ there must be a candidate $x \notin \mathcal{U}_j$ with $\tau(x) = \sigma_j^{(r)}(x)$ for all $\sigma_j^{(r)} \in \Sigma_j$ since they form a $|\mathcal{U}_j|$-anti-pattern when restricted to $\mathcal{U}_j$. The position of $x \in \mathcal{D} \setminus \mathcal{U}_j$ in all $\sigma_j^{(r)} \in \Sigma_j$ is identical and determined by s_j. Therefore, $x = a_i$ or $x = b_i$ for a position i where $s_j(i) \neq *$, and thus, $t^*(i) = s_j(i)$. Hence, t^* is a hitting string. □

5 Approximability

We shortly discuss approximations.

Lemma 8. *The associated minimization problem of MR is 2-approximable for any pseudometric d.*

Proof. Let $\tau^* \in \mathrm{Perm}(\mathcal{D})$ be the optimal consensus for the MR problem under pseudometric d with voters $\sigma_1, \ldots, \sigma_m \in \mathcal{D}$. Then the *pick-a-perm* method [1] with $\tau = \sigma_j$ for $j \in \underline{m}$ yields a 2-approximation since for all $i \in \underline{m}$ we have

$$d(\sigma_i, \tau) \leq d(\sigma_i, \tau^*) + d(\tau^*, \tau) \leq 2 \cdot \max\{d(\sigma_i, \tau^*), d(\tau^*, \tau)\} \leq 2 \cdot \max_{j=1}^{m} d(\sigma_j, \tau^*) \square$$

Note that this approximation ratio for pick-a-perm is tight for all metrics satisfying Requirements 1 and 2. For instance, consider the voters $f(\texttt{"}1000\ldots\texttt{"})$, $f(\texttt{"}0100\ldots\texttt{"})$, $f(\texttt{"}0010\ldots\texttt{"})$ with f as defined in the last section. The distance between each pair of voters is $2c$, while the optimal consensus would be $f(\texttt{"}0000\ldots\texttt{"})$ with a distance of c.

6 Fixed-Parameter Tractability

The reduction in Sect. 4 demonstrates a close relationship between Closest Binary String and MR. We strengthen this observation by extending a fixed parameter algorithm for Closest Binary String [17, 22] such that it can be applied to MR under several metrics. For an introduction to fixed-parameter tractability see [12, 22].

The notion of the *modification set* $M(\tau, \sigma) \subseteq \mathrm{Perm}(\mathcal{D})$ is at the heart of our generalized algorithm. Intuitively, it captures the idea of going "one step" from τ to σ. The structure of the modification set must be chosen individually for each metric d. We state a sufficient condition, which we call the *δ-improving* of M, such that the algorithm actually finds the optimal consensus.

Requirement 3 (δ-improving). *Let $\delta \in \mathbb{N} \setminus \{0\}$. Let $\sigma, \tau, \tau^* \in \mathrm{Perm}(\mathcal{D})$ and $k \in \mathbb{N}$ such that $d(\tau^*, \sigma) \leq k$ and $d(\tau, \tau^*) \leq k$. If $k < d(\tau, \sigma) \leq 2k$, then there exists a $\tau' \in M(\tau, \sigma)$ such that $d(\tau', \tau^*) \leq d(\tau, \tau^*) - \delta$.*

Input: Voters $\sigma_1, \ldots, \sigma_m \in \mathrm{Perm}(\mathcal{D})$, bound $k \in \mathbb{N}$.
Output: k-consensus $\tau^* \in \mathrm{Perm}(\mathcal{D})$ or reject.

search(σ_1, k);

function *search*($\tau, \Delta k$)
    if $\forall j \in \underline{m} : d(\tau, \sigma_j) \leq k$ **then return** $\{\tau\}$;
    if $\exists j \in \underline{m} : d(\tau, \sigma_j) > k + \Delta k$ **then return** $\emptyset$;
    if $\Delta k > 0$ **then**
        let $j \in \underline{m}$ such that $k < d(\tau, \sigma_j) \leq k + \Delta k$;
        foreach $\tau' \in M(\tau, \sigma_j)$ **do**
            $R \leftarrow$ search($\tau', \Delta k - \delta$);
            if $R \neq \emptyset$ **then return** R;
    return $\emptyset$

Algorithm 1: Fixed-parameter algorithm for MR

Lemma 9. *Suppose that there is a k-consensus τ^*, i. e., $\max_{j=1}^{m} d(\tau^*, \sigma_j) \leq k$. If M is δ-improving, then at recursion depth i, search in Algorithm 1 has either already found a k-consensus, or is called at least once with a parameter τ such that $d(\tau, \tau^*) \leq k - \delta i$.*

Proof. We proof by induction on the recursion depth i. Induction basis: Since $\tau = \sigma_1$ in depth 0, we have $d(\tau, \tau^*) \leq k - 0$ by definition. Induction step: Suppose the program has not found the solution yet and that at recursion depth $i \leq \lceil \frac{k}{\delta} \rceil$ search is called with τ' having $d(\tau, \tau^*) \leq k - \delta i$. If $d(\tau, \sigma_j) \leq k$ we have found a k-consensus and are done. Otherwise, $d(\tau, \sigma_j) > k$. The break condition in line 4 does not hold since $d(\tau, \sigma_j) \leq d(\tau, \tau^*) + d(\tau^*, \sigma_j) \leq k - \delta i + k = \Delta k + k$. As τ' iterates over $M(\tau, \sigma_j)$, by Requirement 3 there is at least one iteration where search is called with a τ' where $d(\tau', \tau^*) \leq k - \delta i - \delta$. □

Theorem 5. *If M is δ-improving, then Algorithm 1 finds a k-consensus τ^* or correctly reports that no such consensus exists. Its running time is $\mathcal{O}((f(k))^{\lceil \frac{k}{\delta} \rceil} \cdot g(k, n))$, where $f(k)$ is the maximum size of the constructed modification sets and $g(k, n)$ is the time required for the construction of a modification set.*

Proof. The recursion depth is bounded by $\lceil \frac{k}{\delta} \rceil$ and the branching factor is limited by the maximum size of the modification set. The running time is worst if no k-consensus exists, in which case search returns the empty set. Otherwise, suppose that τ^* is a k-consensus. Then, by Lemma 9, search finds a different k-consensus or is eventually called with a τ such that $d(\tau, \tau^*) = 0$ which implies $\tau = \tau^*$. □

For fixed-parameter results it remains to construct a suitable modification set for each distance.

Lemma 10. *The modification set $M(\tau, \sigma) = \{T_{\tau(x),\tau(y)} \circ \tau : \{x, y\} \in \mathcal{K}(\tau, \sigma)\}$ is 1-improving under the Kendall tau distance K.*

Proof. Let $k \in \mathbb{N}$ and $\sigma, \tau, \tau^* \in \mathrm{Perm}(\mathcal{D})$ such that $K(\tau^*, \sigma) \leq k < K(\tau, \sigma) \leq 2k$. We show that $\mathcal{K}(\tau, \sigma) \cap \mathcal{K}(\tau, \tau^*) \neq \emptyset$ since for any dirty pair $\{x, y\} \in \mathcal{K}(\tau, \sigma) \cap \mathcal{K}(\tau, \tau^*)$ we have $d(T_{\tau(x),\tau(y)} \circ \tau, \tau^*) < d(\tau, \tau^*)$ by Lemma 1. Assume for contradiction that $\mathcal{K}(\tau, \sigma)$ and $\mathcal{K}(\tau, \tau^*)$ are disjoint. Let $\{x, y\} \in \mathcal{K}(\tau, \sigma)$. As $\{x, y\} \notin \mathcal{K}(\tau, \tau^*)$, we know that τ and τ^* agree on the relative order of x and y, which implies that $\{x, y\} \in \mathcal{K}(\sigma, \tau^*)$. Hence, $\mathcal{K}(\tau, \sigma) \subseteq \mathcal{K}(\sigma, \tau^*)$. Now let $\{x, y\} \in \mathcal{K}(\tau, \tau^*)$. As $\{x, y\} \notin \mathcal{K}(\tau, \sigma)$, τ and σ agree on the relative order of x and y, implying $\{x, y\} \in \mathcal{K}(\sigma, \tau^*)$. Hence, $\mathcal{K}(\tau, \tau^*) \subseteq \mathcal{K}(\sigma, \tau^*)$. We conclude that $K(\sigma, \tau^*) = |\mathcal{K}(\sigma, \tau^*)| \geq |\mathcal{K}(\tau, \sigma)| + |\mathcal{K}(\tau, \tau^*)| \geq k + 1$, a contradiction. □

Corollary 3. *MR under the Kendall tau distance K can be computed in $\mathcal{O}((2k)^k \cdot (mn \log n + k)\})$ time.*

Proof. Consider the modification set of Lemma 10, whose size is $|M(\tau, \sigma)| = |\mathcal{K}(\tau, \sigma)| = K(\tau, \sigma) \leq 2k$. The distance of two permutations can be computed in $\mathcal{O}(n \log n)$ time [21]. Hence, lines 3, 4 and 6 of Algorithm 1 need $\mathcal{O}(mn \log n)$ time. For efficiency, we represent the modification set $M(\tau, \sigma)$ only implicitly by the set $\mathcal{K}(\tau,\sigma)$ of at most $2k$ dirty pairs, which can be computed in $\mathcal{O}(n\log n+k)$ time [6]. We iterate τ' over $M(\tau,\sigma)$ by transposing the next dirty pair in $\mathcal{K}(\tau,\sigma)$, descent recursively, and undo the transposition after the recursive call returns. Thus, excluding the recursion, the loop requires $\mathcal{O}(n\log n+k)$ time. □

Lemma 11. *The modification set $M(\tau,\sigma) = \{T_{\tau(x),\sigma(x)} \circ \tau : x \in \mathcal{H}(\tau,\sigma)\}$ is 1-improving under the Hamming distance H.*

Proof. Let $k \in \mathbb{N}$ and $\sigma,\tau,\tau^* \in \mathrm{Perm}(\mathcal{D})$ such that $H(\tau^*,\sigma) \le k < H(\tau,\sigma) \le 2k$. The size of the modification set is $|M(\tau,\sigma)| = |\mathcal{H}(\tau,\sigma)| = H(\tau,\sigma) \le 2k$. σ and τ^* agree in the position of at least $|D|-k$ candidates. As $H(\tau,\sigma) > k$, there is at least one candidate x with $\tau(x) \ne \sigma(x) = \tau^*(x)$. Hence, $H(T_{\tau(x),\sigma(x)}\circ\tau,\tau^*) < H(\tau,\tau^*)$ (see Proposition 1). □

Corollary 4. *MR under the Hamming distance H can be computed in $\mathcal{O}((2k)^k \cdot mn)$ time.*

Proof. Consider the modification set of Lemma 11. The Hamming distance between two permutations can be computed in linear time. Thus, lines 3, 4 and 6 of Algorithm 1 need $\mathcal{O}(mn)$ time. Similarly to the proof of Corollary 3, the iteration of τ' over the modification set $M(\tau,\sigma)$ with $|M(\tau,\sigma)| \le 2k$ is done in place and needs only $\mathcal{O}(k)$ time. □

Lemma 12. *The modification set $M(\tau,\sigma) = \{T_{\tau(x),i} \circ \tau : x \in \mathcal{H}(\tau,\sigma) \wedge i \in \{\tau(x) + j\cdot \mathrm{sgn}(\sigma(x)-\tau(x)) : j \in \underline{\lceil k^{\frac{1}{p}}\rceil}\} \cap \underline{n}\}$ is $(p+1)$-improving under the raised Minkowski distance $\hat{F}_p$ for $p \in \mathbb{N}\setminus\{0\}$.*

Proof. Let $k \in \mathbb{N}$ and $\sigma,\tau,\tau^* \in \mathrm{Perm}(\mathcal{D})$ such that $\hat{F}_p(\tau^*,\sigma) \le k < \hat{F}_p(\tau,\sigma) \le 2k$. We take every displaced candidate $x \in \mathcal{H}(\tau,\sigma)$ and try all possibilities to transpose it with candidates placed at most k positions to its right or left, depending on whether $\sigma(x) > \tau(x)$ or $\sigma(x) < \tau(x)$, respectively. Suppose we have a candidate $x \in \mathcal{H}(\tau,\sigma)$ with $|\sigma(x)-\tau^*(x)| < |\sigma(x)-\tau(x)|$. There must be at least one such candidate since $\hat{F}_p(\tau^*,\sigma) < \hat{F}_p(\tau,\sigma)$. W.l.o.g. assume $\sigma(x) > \tau(x)$. Otherwise, the following arguments apply symmetrically. Let $\mathcal{Y} = \{\tau^{*-1}(i) : i \le \tau(x)\}$ be the set of candidates which are placed in τ^* to the left of or on the same position where x is placed in τ. As $\tau^*(x) > \tau(x)$, $x \notin \mathcal{Y}$, so by a counting argument there must be some $y \in \mathcal{Y}$ with $\tau(y) > \tau(x)$. We know that $\tau' = T_{\tau(x),\tau(y)} \circ \tau$ is contained in the modification set because $\tau(y) - \tau^*(y) \le k^{\frac{1}{p}}$ due to $\hat{F}_p(\tau,\tau^*) \le k$ by Requirement 3. We distinguish two cases whether or not $\tau(y) \le \tau^*(x)$.

Case 1: $\boldsymbol{\tau^*(y) \le \tau(x) < \tau(y) \le \tau^*(x)}$. Then both $\tau^*(x) - \tau'(x) = \tau^*(x) - \tau(y) < \tau^*(x) - \tau(x)$ and $\tau'(y) - \tau^*(y) = \tau(x) - \tau^*(y) < \tau(y) - \tau^*(y)$. Hence, by the Binomial Theorem, $|\tau(x)-\tau^*(x)|^p - |\tau'(x)-\tau^*(x)|^p =$

$$\underbrace{(|\tau(x)-\tau^*(x)| - |\tau'(x)-\tau^*(x)|)}_{\ge 1} \cdot \sum_{i=0}^{p-1} \underbrace{|\tau(x)-\tau^*(x)|^i}_{\ge 1} \cdot \underbrace{|\tau'(x)-\tau^*(x)|^{p-i-1}}_{\ge 1}$$

and thus $|\tau'(x) - \tau^*(x)|^p \leq |\tau(x) - \tau^*(x)|^p - p$. We obtain $|\tau'(y) - \tau^*(y)|^p \leq |\tau(y) - \tau^*(y)|^p - p$ symmetrically. In sum $|\tau'(x) - \tau^*(x)|^p + |\tau'(y) - \tau^*(y)|^p \leq |\tau(x) - \tau^*(x)|^p + |\tau(y) - \tau^*(y)|^p - 2p$.

Case 2: $\boldsymbol{\tau^*(y) \leq \tau(x) < \tau^*(x) < \tau(y)}$. Then $\tau'(x)-\tau^*(x)+\tau'(y)-\tau^*(y) = \tau(y) - \tau^*(x) + \tau(x) - \tau^*(y) < \tau(y) - \tau^*(y)$. By the Binomial Theorem we derive

$$(\tau'(x) - \tau^*(x) + \tau'(y) - \tau^*(y))^p \leq (\tau(y) - \tau^*(y))^p - p$$
$$|\tau'(x) - \tau^*(x)|^p + |\tau'(y) - \tau^*(y)|^p \leq |\tau(y) - \tau^*(y)|^p - p + \underbrace{|\tau(x) - \tau^*(x)|^p}_{\geq 1} - 1$$

Recall that the positions of candidates $\mathcal{D} \setminus \{x, y\}$ are unaffected. Hence, in both cases $\hat{F}_p(\tau', \tau^*) \leq \hat{F}_p(\tau, \tau^*) - (p + 1)$. □

Corollary 5. *MR under the Minkowski distance F_p for $p \in \mathbb{N} \setminus \{0\}$ can be computed in $\mathcal{O}((2k^{p+1})^{\lceil \frac{k^p}{p+1} \rceil} \cdot mn)$ time.*

Proof. Let $\hat{k} = k^p$. Finding a k-consensus for F_p is equivalent to finding a $\hat{k}$-consensus for $\hat{F}_p$. Consider the modification set of Lemma 12. Its size is $|M(\tau, \sigma)| \leq 2\hat{k}^{1+\frac{1}{p}}$ since there are most $2\hat{k}$ displaced candidates which are each tested on at most $\hat{k}^{\frac{1}{p}}$ positions. The Minkowski distance between two permutations can be computed in linear time. Thus, lines 3, 4 and 6 of Algorithm 1 need $\mathcal{O}(mn)$ time. Finding the up to $2\hat{k}$ displaced candidates to build the modification set needs $\mathcal{O}(n)$ time. Each displaced candidate is tested on $\hat{k}^{\frac{1}{p}}$ positions. Then the total running time is in $\mathcal{O}((2\hat{k}^{1+\frac{1}{p}})^{\lceil \frac{\hat{k}}{p+1} \rceil} \cdot mn) = \mathcal{O}((2k^{p+1})^{\lceil \frac{k^p}{p+1} \rceil} \cdot mn)$. □

There are tractable algorithms for Closest Binary String parameterizing the number of strings m [22]. However, parameterizing MR by the number of voters m does not lead to efficient algorithms since MR under the Kendall tau distance is **NP**-hard for $m = 4$ [6].

Note that the **NP**-hardness of MR under the Minimum distance even for $k = 0$ implies that this problem is not fixed-parameter tractable by k unless $\mathbf{P} = \mathbf{NP}$.

7 Conclusion

We explored the complexity of MR by stating sufficient conditions for metrics under which MR is **NP**-complete and fixed-parameter tractable. Considering **NP**-hardness, the Requirements 1 and 2 should also hold for other distances, e. g., Damerau-Levenshtein, Block-Transpositions, or Reversals [9, 10]. Finding a suitable modification set (Requirement 3) for Cayley and Ulam distances is still open. Another extension of MR is to allow the voters providing partial orders. The distance is then measured by the *Nearest Neighbor distance*, which we studied for Spearman footrule and Kendall tau in [7, 8].

References

1. Ailon, N.: Aggregation of partial rankings, p-ratings and top-m lists. Algorithmica 57(2), 284–300 (2010)
2. Ailon, N., Charikar, M., Newman, A.: Aggregating inconsistent information: Ranking and clustering. J. ACM 55(5), 23:1–23:27 (2008)
3. Bartholdi, J.J., Tovey, C.A., Trick, M.A.: Voting schemes for which it can be difficult to tell who won the election. Soc. Choice Welfare 6(2), 157–165 (1989)
4. Bergroth, L., Harri, H., Timo, R.: A survey of longest common subsequence algorithms. In: Proc. String Processing and Information Retrieval, SPIRE 2000. pp. 39–48. IEEE (2000)
5. Betzler, N., Fellows, M.R., Guo, J., Niedermeier, R., Rosamond, F.A.: Fixed-parameter algorithms for Kemeny rankings. Theor. Comput. Sci. 410(8), 4554–4570 (2009)
6. Biedl, T., Brandenburg, F.J., Deng, X.: On the complexity of crossings in permutations. Discrete Math. 309(7), 1813–1823 (2009)
7. Brandenburg, F.J., Gleißner, A., Hofmeier, A.: The nearest neighbor Spearman footrule distance for bucket, interval, and partial orders. J. Comb. Optim. pp. 1–23 (2012), published online March 2012
8. Brandenburg, F.J., Gleißner, A., Hofmeier, A.: Comparing and aggregating partial orders with Kendall tau distances. Discrete Math. Algorithms Appl. 5(2) (2013)
9. Critchlow, D.E.: Metric Methods for Analyzing Partially Ranked Data, LNS, vol. 34. Springer (1985)
10. Deza, M., Huang, T.: Metrics on permutations, a survey. J. Combin. Inform. System Sci. 23, 173–185 (1998)
11. Diaconis, P., Graham, R.L.: Spearman's footrule as a measure of disarray. J. Roy. Statist. Soc. B 39(2), 262–268 (1977)
12. Downey, R.G., Fellows, M.R.: Parameterized Complexity. Monographs in Computer Science, Springer (1999)
13. Dwork, C., Kumar, R., Naor, M., Sivakumar, D.: Rank aggregation methods for the web. In: Proc. World Wide Web Conference, WWW 2010. pp. 613–622 (2001)
14. Fagin, R.: Generalized first-order spectra and polynomial-time recognizable sets. In: Karp, R. (ed.) Complexity of Computation. SIAM-AMS Proceedings, vol. 7, pp. 43–73. ACM (1974)
15. Fagin, R., Kumar, R., Mahdian, M., Sivakumar, D., Vee, E.: Comparing partial rankings. SIAM J. Discrete Math. 20(3), 628–648 (2006)
16. Frances, M., Litman, A.: On covering problems of codes. Theor. Comput. Syst. 30(2), 113–119 (1997)
17. Gramm, J., Niedermeier, R., Rossmanith, P.: Fixed-parameter algorithms for Closest String and related problems. Algorithmica 37(1), 25–42 (2003)
18. Karpinski, M., Schudy, W.: Faster algorithms for feedback arc set tournament, Kemeny rank aggregation and betweenness tournament. In: Proc. ISAAC 2010. LNCS, vol. 6506, pp. 3–14. Springer (2010)
19. Kemeny, J.: Mathematics without numbers. Daedalus 88, 577–591 (1959)
20. Kenyon-Mathieu, C., Schudy, W.: How to rank with few errors. In: Proc. Symposium on Theory of Computing, STOC 2007. pp. 95–103. ACM (2007)
21. Knight, W.R.: A computer method for calculating Kendall's tau with ungrouped data. J. Amer. Statist. Assoc. 61(314), 436–439 (1966)
22. Niedermeier, R.: Invitation to Fixed-Parameter Algorithms. Oxford University Press (2006)