The Nearest Neighbor Spearman Footrule Distance for Bucket, Interval, and Partial Orders

Franz J. Brandenburg, Andreas Gleißner, and Andreas Hofmeier

Department of Informatics and Mathematics, University of Passau
{brandenb;gleissner;hofmeier}@fim.uni-passau.de

Technical Report, Number MIP-1101
Department of Informatics and Mathematics
University of Passau, Germany
February 2011

The Nearest Neighbor Spearman Footrule Distance for Bucket, Interval, and Partial Orders

Franz J. Brandenburg, Andreas Gleißner, and Andreas Hofmeier

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

Abstract. Comparing and ranking information is an important topic in social and information sciences, and in particular on the web. Its objective is to measure the difference of the preferences of voters on a set of candidates and to compute a consensus ranking. Commonly, each voter provides a total order of all candidates. Recently, this approach has been generalized to bucket orders, which allow ties.
In this work we further generalize and consider total, bucket, interval and partial orders. The disagreement between two orders is measured by the nearest neigbor Spearman footrule distance, which has not been studied so far. We show that the nearest neighbor Spearman footrule distance of two bucket orders and of a total and an interval order can be computed in linear time, whereas the computation is **NP**-complete and 6-approximable for a total and a partial order. Moreover, we establish the **NP**-completeness and the 4-approximability of the rank aggregation problem for bucket orders. This sharply contrasts the well-known efficient solution of this problem for total orders.

1 Introduction

The rank aggregation problem consists in finding a consensus ranking on a set of candidates, based on the preferences of individual voters. The problem has many applications including meta search, biological databases, similarity search, and classification [2, 7, 12, 16, 18–20, 23]. It has been mathematically investigated by Borda [6] and Condorcet [8] (18th century) and even by Lullus [17] and Cusanus [10] (13th century) in the context of voting theory.

The formal treatment of the rank aggregation problem is determined by the strictness of the preferences. It is often assumed that each voter makes clear and unambiguous decisions on all candidates, i. e. the preferences are given by total orders. However, the rankings encountered in practice often have deficits against the complete information provided by a total order, as voters often come up with unrelated candidates, which they consider as tied ("I consider x and y coequal.") or incomparable ("I cannot compare x (apples) and y (oranges)".). Voters considering all unrelated pairs of candidates as tied are represented by *bucket orders*, such that ties define an equivalence relation on candidates within a bucket. They are also known as partial rankings or weak orders [1, 13].

As incomparable pairs of candidates come into play, more general orders are needed: A ranking is an *interval order* if the voters specify their preferences by associating an interval with each candidate. Candidate x is then preferred over y if the interval of x ends before the one of y begins, while overlapping intervals represent incomparabilities or ties. In the most general case the voters describe their preferences by *partial orders*. In this case unrelatedness (ties and incomparabilities) is not transitive and the preference relation is not negatively transitive. In all orders for two unrelated candidates, no matter if they are tied or incomparable, the voter accepts any local order on them without penalty or cost. Nevertheless, we will stress the different intuition behind unrelated candidates by speaking of tied candidates ($\cong$) in bucket orders and of unrelated ($\not\precsim$, meaning tied or incomparable) candidates in interval or partial orders.

The common distance measures for two total orders σ and τ are the Kendall tau and the Spearman footrule distance, $K(\sigma, \tau)$ and $F(\sigma, \tau)$. $K(\sigma, \tau)$ counts the number of disagreements of candidates, while $F(\sigma, \tau)$ accumulates the mismatches, summing the distances of the positions of each candidate.

Investigations on ranking problems have focused on total orders or permutations. Its generalization to bucket orders has been considered more recently by Ailon [1] and Fagin et al. [13]. The focus and main result in [13] is the equivalence of several distance measures, especially the Hausdorff versions of the Kendall tau and Spearman footrule distances, introduced by Critchlow [9]. Ailon [1] studied the nearest neighbor Kendall tau distance for bucket orders.

In this work we generalize rankings to partial and interval orders, and measure the distance by the nearest neighbor Spearman footrule distance. Our emphasis is on the complexity of computing distances and rank aggregations. We establish a sharp separation between efficient algorithms and **NP**-completeness. In particular, we show that the nearest neighbor Spearman footrule distance can be computed in linear time for two bucket orders and for a total and an interval order. In contrast, these computations are **NP**-complete for a total and a partial order, and hence for the more general cases. These results (and some open problems) are summarized in Tab. 1. Concerning the Spearman footrule distance and total orders, the rank aggregation problem can be solved efficiently using a weighted bipartite matching [12]. This sharply contrasts our **NP**-completeness result for bucket orders. Furthermore, we establish the equivalence between the nearest neighbor Spearman footrule distance and the nearest neighbor Kendall tau distance. Finally, we achieve constant factor approximations for the computation of the nearest neighbor Spearman footrule distance of a total and a partial order as well as for the rank aggregation problem for bucket orders.

This work is organized as follows: In Sect. 2 we introduce orders and distances. In Sect. 3 and Sect. 4 we consider the complexity of distance and rank aggregation problems. Sect. 5 addresses the equivalence of the nearest neighbor Kendall tau and Spearman footrule distances of partial orders and establishes the constant factor approximability for some problems, which we have shown to be **NP**-complete. We conclude with some open problems in Sect. 6.

Table 1. Computation of F_{NN} between two orders

	total	bucket	interval	partial
partial	**NP**-C (Th. 3) 6-appr.	**NP**-C (Th. 3) appr. open	**NP**-C (Th. 3) appr. open	**NP**-C (Th. 3) appr. open
interval	$\mathcal{O}(n)$ (Th. 2)	compl. open appr. open	compl. open appr. open	
bucket	$\mathcal{O}(n)$ (Th. 1)	$\mathcal{O}(n)$ (Th. 1)		
total	$\mathcal{O}(n)$ (obv.)			

Table 2. Rank aggregation problems with F_{NN} for different types of orders

total	bucket	interval	partial
$\mathcal{O}(n^3)$ ([12])	**NP**-C (Th. 4) 4-appr.	**NP**-C (Th. 4) appr. open	**NP**-C (Th. 4) appr. open

2 Preliminaries

For a binary relation R on a domain $\mathcal{D}$ and for each $x, y \in \mathcal{D}$, we denote $x \prec_R y$ if $(x, y) \in R$ and $x \not\prec_R y$ if $(x, y) \notin R$. A binary relation κ is a (strict) *partial order* if it is *irreflexive*, *asymmetric* and *transitive*, i. e., $x \not\prec_\kappa x$, $x \prec_\kappa y \Rightarrow y \not\prec_\kappa x$, and $x \prec_\kappa y \wedge y \prec_\kappa z \Rightarrow x \prec_\kappa z$ for all $x, y, z \in \mathcal{D}$. Candidates x and y are called *unrelated* if $x \not\prec_\kappa y \wedge y \not\prec_\kappa x$, which we denote by $x \not\precsim_\kappa y$. The intuition of $x \prec_\kappa y$ is that κ *ranks* x *before* y, which means a preference for x. A partial order α is an *interval order* if there is a bijection I from $\mathcal{D}$ into a set of intervals with $I(x) = [l_x, r_x]$ and $x \prec_\alpha y \Leftrightarrow r_x < l_y$. W. l. o. g., the boundaries of the intervals are integers between 1 and $|\mathcal{D}|$. A partial order π is a *bucket order* if it is irreflexive, asymmetric, transitive and *negatively transitive*, which says that for each $x, y, z \in \mathcal{D}$, $x \prec_\pi y \Rightarrow x \prec_\pi z \vee z \prec_\pi y$. Hence, the domain is partitioned into a sequence of buckets $\mathcal{B}_1, \ldots, \mathcal{B}_t$ such that $x \prec_\pi y$ if there are i, j with $i < j$ and $x \in \mathcal{B}_i$ and $y \in \mathcal{B}_j$. Note that x and y are unrelated if they are in the same bucket. Thus, unrelatedness is an equivalence relation on *tied* candidates $x \cong_\pi y$ within a bucket. Finally, a partial order τ is a *total order* if it is irreflexive, asymmetric, transitive and *complete*, i. e., $x \prec_\tau y \vee y \prec_\tau x$ for all $x, y \in \mathcal{D}$ with $x \neq y$. Then τ is a permutation of the elements of $\mathcal{D}$. τ can also be considered as a bijection $\tau : \mathcal{D} \to \{1, \ldots, |\mathcal{D}|\}$. Clearly, total $\subset$ bucket $\subset$ interval $\subset$ partial, where $\subset$ expresses a generalization.

For two total orders σ and τ the *Kendall tau distance* counts the disagreements or inversions of pairs of candidates, $K(\sigma, \tau) = |\{\{x, y\} \subseteq \mathcal{D} : x \prec_\sigma y \wedge y \prec_\tau x\}|$. The *Spearman footrule distance* is the L_1-norm taking the difference of the positions of the candidates into account, $F(\sigma, \tau) = \sum_{x \in \mathcal{D}} |\sigma(x) - \tau(x)|$.

We consider distances between generalized orders based on their sets of total extensions. A total order τ is a *total extension* of a partial order κ if τ does not contradict κ, i. e., $x \prec_\kappa y \Rightarrow x \prec_\tau y$ for all $x, y \in \mathcal{D}$.

Definition 1. *For partial orders κ and μ on a domain $\mathcal{D}$ define the* nearest neighbor Spearman footrule *and* Kendall tau distance *via their extensions,*

$$F_{NN}(\kappa, \mu) = \min\{F(\tau, \sigma) : \tau \in \mathrm{Ext}(\kappa), \sigma \in \mathrm{Ext}(\mu)\}$$

$$K_{NN}(\kappa, \mu) = \min\{K(\tau, \sigma) : \tau \in \mathrm{Ext}(\kappa), \sigma \in \mathrm{Ext}(\mu)\}$$

Observe that the nearest neighbor distances fail the axioms of a metric. They do neither satisfy the identity of indiscernible $d(x, y) = 0 \Leftrightarrow x = y$ nor does the triangle inequality hold.

Proposition 1. *The nearest neighbor Kendall tau and Spearman footrule distances coincide with their mates on total orders τ and σ, i. e. $K_{NN}(\tau, \sigma) = K(\tau, \sigma)$ and $F_{NN}(\tau, \sigma) = F(\tau, \sigma)$.*

Definition 2. *Given two orders κ and μ on a domain $\mathcal{D}$ and an integer k, the* distance problem *is whether or not $d(\kappa, \mu) \leq k$.*

Accordingly, the rank aggregation problem *is the problem whether or not for orders $\kappa_1, \ldots, \kappa_r$ and an integer k, there exists a total order τ such that $\sum_{i=1}^{r} d(\kappa_i, \tau) \leq k$. A total order τ^* minimizing k is the* consensus ranking.

For a partial order κ on a domain $\mathcal{D}$ and a set $\mathcal{X} \subseteq \mathcal{D}$ we write $[\mathcal{X}]$ if $\mathcal{X}$ is totally ordered by κ in a way that is clear from the respective context. For sets $\mathcal{X}, \mathcal{Y} \subseteq \mathcal{D}$, if $x \prec_\kappa y$ for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, we write $\mathcal{X} \prec_\kappa \mathcal{Y}$. We call $\mathcal{X}$ *unrelated by* κ if $x_i \not\preccurlyeq_\kappa x_j$ for all $x_i, x_j \in \mathcal{X}$.

In the following proofs we use shifting and switching operations on total orders. For two total orders σ_1 and σ_2 on a domain $\mathcal{D}$ and candidates $x, y \in \mathcal{D}$ we say that σ_2 is derived from σ_1 by *shifting x up to position p* if $\sigma_2(c) = \sigma_1(c)$ for all $c \in \mathcal{D}$ with $\sigma_1(c) < \sigma_1(x)$ or with $\sigma_1(c) > p$, and if $\sigma_2(c) = \sigma_1(c) - 1$ for all $c \in \mathcal{D}$ with $\sigma_1(x) < \sigma_1(c) \leq p$, and if $\sigma_2(x) = p$. *Shifting x down to position p* is defined analogously. We say that σ_2 is derived from σ_1 by *switching x and y*, if $\sigma_2(c) = \sigma_1(c)$ for all $c \in \mathcal{D} \setminus \{x, y\}$, and if $\sigma_2(x) = \sigma_1(y)$, and if $\sigma_2(y) = \sigma_1(x)$.

3 Distance Problems

In this section we address the computation of the nearest neighbor Spearman footrule distance of two bucket orders, of a total and an interval order and of a total and a partial order.

3.1 Nearest Neighbor Spearman Footrule Distance of Bucket Orders

Theorem 1. *The nearest neighbor Spearman footrule distance of two bucket orders can be computed in linear time.*

We start with the definition of an operation, that breaks ties within a bucket order. The *refinement* of a bucket order γ by a bucket order π is the bucket order $\pi * \gamma$ such that $x \prec_{\pi*\gamma} y \Leftrightarrow x \prec_\gamma y \vee x \cong_\gamma y \wedge x \prec_\pi y$ holds for all $x, y \in \mathcal{D}$. Hence, a tie in γ may be broken by π. Clearly, if π is a total order then $\pi * \gamma$ is a total order. $*$ is an associative operation, so for a third bucket order η on $\mathcal{D}$, $\eta * \pi * \gamma$ makes sense. Note that refinement is only defined for bucket orders, but not for interval or partial orders.

Fagin et al. [13] have characterized the Hausdorff Spearman footrule distance of two bucket orders in terms of refinements. Adopting techniques from [13] we obtain the corresponding characterization for the nearest neighbor Spearman footrule distance. From [13] we can directly reuse Lemma 1, Lemma 2 and Lemma 3, which we state here without proof, and rephrase Lemma 4 to serve our purposes.

Lemma 1. *[13] For positive integers $a, b, c, d \in \mathbb{N}$, suppose $a \leq b$ and $c \leq d$. Then $|a - c| + |b - d| \leq |a - d| + |b - c|$.*

Lemma 2. *[13] Let τ be a total order and let γ be a bucket order on the domain $\mathcal{D}$. Suppose that $\tau \neq \gamma$. Then there exist $x, y \in \mathcal{D}$ such that $\tau(y) = \tau(x) + 1$ and $y \prec_\gamma x$ or $y \cong_\gamma x$. If γ is a total order, then $\gamma(y) < \gamma(x)$.*

Lemma 3. *[13] Let τ be a total order and let γ be a bucket order on the domain $\mathcal{D}$. Then the quantity $F(\tau, \sigma)$ taken over all $\sigma \in \mathrm{Ext}(\gamma)$ is minimized for $\sigma = \tau * \gamma$.*

Lemma 4. *(adapted from [13]) Let π and γ be bucket orders and let ρ be an arbitrary total order on the domain $\mathcal{D}$. Then the quantity $F(\sigma, \sigma * \gamma)$, taken over all $\sigma \in \mathrm{Ext}(\pi)$, is minimzed if $\sigma = \rho * \gamma * \pi$.*

Proof. Note that for any $\sigma \in \mathrm{Ext}(\pi)$ there is some total order τ, such that $\sigma = \tau * \pi$. We now show that $\rho * \gamma$ is among the best choices for τ with regard to the minimization of $F(\sigma, \sigma * \gamma)$. That means for all total orders τ,

$$F(\rho * \gamma * \pi, \rho * \gamma * \pi * \gamma) \leq F(\tau * \pi, \tau * \pi * \gamma)$$

from which the lemma follows.

Let U be the set of total orders with $U = \{\tau : F(\rho * \gamma * \pi, \rho * \gamma * \pi * \gamma) > F(\tau * \pi, \tau * \pi * \gamma))\}$. If U is empty, we are done, so suppose U is not empty.

Over all total orders in U, choose τ to be the total order minimizing $K(\tau, \rho * \gamma)$. As clearly $\rho * \gamma \notin U$, $\tau \neq \rho * \gamma$. Therefore, Lemma 2 guarantees that we can find a pair $x, y \in \mathcal{D}$ such that $\tau(y) = \tau(x) + 1$, but $\rho * \gamma(y) < \rho * \gamma(x)$. Produce τ' by switching x and y in τ. Clearly, τ' has one inversion less than τ with respect to $\rho * \gamma$, so $K(\tau', \rho * \gamma) < K(\tau, \rho * \gamma)$. We now show that $\tau' \in U$ holds, which derives a contradiction as τ is supposed to be the total order in U having the minimum Kendall tau distance to $\rho * \gamma$.

Case 1: If $x \prec_\pi y$ or $y \prec_\pi x$, then $\tau' * \pi = \tau * \pi$. Hence $F(\tau' * \pi, \tau' * \pi * \gamma) = F(\tau * \pi, \tau * \pi * \gamma)$ and $\tau' \in U$.

Case 2: If $x \cong_\pi y$ and $x \cong_\gamma y$ then switching x and y in τ switches their positions in both $\tau * \pi$ and $\tau * \pi * \gamma$, while leaving all the other candidates in

their position. So we have $F(\tau' * \pi, \tau' * \pi * \gamma) = F(\tau * \pi, \tau * \pi * \gamma)$ and we again conclude that $\tau' \in U$.

Case 3: If $x \cong_\pi y$ and $x \prec_\gamma y$ or $y \prec_\gamma x$, we have the following situation: First $\tau' * \pi$ is just $\tau * \pi$ with the adjacent elements x and y switched. Second $\tau' * \pi * \gamma = \tau * \pi * \gamma$ as x and y are not tied in γ. Recall that we have chosen x and y with the property that $x \prec_\tau y$ and $y \prec_{\rho * \gamma} x$. From $x \cong_\pi y$ and $x \prec_\tau y$ we derive $\tau * \pi(x) < \tau * \pi(y)$. From $y \prec_{\rho * \gamma} x$ we derive $y \prec_{\tau * \rho * \gamma} x$. We now make use of Lemma 1. We substitute $a = \rho * \gamma(y)$, $b = \rho * \gamma(x)$, $c = \tau' * \pi(y)$ and $d = \tau' * \pi(x)$. Then by Lemma 1

$$\begin{aligned} &|\rho * \gamma(y) - \tau' * \pi(y)| + |\rho * \gamma(x) - \tau' * \pi(x)| \leq \\ \leq &|\rho * \gamma(y) - \tau' * \pi(x)| + |\rho * \gamma(x) - \tau' * \pi(y)| \ . \end{aligned}$$

From the fact that $\tau * \pi$ is just $\tau' * \pi$ with the adjacent elements x and y swapped and the fact that $\tau' * \pi * \gamma = \tau * \pi * \gamma$ we derive

$$\begin{aligned} &|\rho * \gamma(y) - \tau' * \pi(x)| + |\rho * \gamma(x) - \tau' * \pi(y)| = \\ = &|\rho * \gamma(y) - \tau * \pi(y)| + |\rho * \gamma(x) - \tau * \pi(x)| \ . \end{aligned}$$

Combining these two (in)equalities and using the fact that for all $z \in \mathcal{D}$ with $z \neq x, y$, $\tau * \pi(z) = \tau' * \pi(z)$, we immediately obtain $F(\tau' * \pi, \tau' * \pi * \gamma) \leq F(\tau * \pi, \tau * \pi * \gamma)$, from which we conclude that $\tau' \in U$. □

The correctness of Theorem 1 can now be verified by combining the results of Lemmas 3 and 4. Think for now of $\sigma \in \mathrm{Ext}(\gamma)$ as fixed. Then by Lemma 3 the quantity $F(\sigma, \tau)$ for every $\tau \in \mathrm{Ext}(\pi)$ is minimized for $\tau = \sigma * \pi$.

By Lemma 4 the quantity $F(\sigma, \sigma * \pi)$ for every $\sigma \in \mathrm{Ext}(\gamma)$ is minimized for $\sigma = \rho * \pi * \gamma$. Therefore

$$\min_{\sigma \in \mathrm{Ext}(\gamma)} \min_{\tau \in \mathrm{Ext}(\pi)} F(\sigma, \tau) = F(\rho * \pi * \gamma, \rho * \pi * \gamma * \pi).$$

Since $\rho * \pi * \gamma * \pi = \rho * \gamma * \pi$, we conclude

$$F_{NN}(\gamma, \pi) = F(\rho * \pi * \gamma, \rho * \gamma * \pi) \, .$$

Theorem 1 follows, since refinements as well as the Spearman footrule distance between two total orders can obviously be computed in linear time.

3.2 Nearest Neighbor Spearman Footrule Distance of a Total and an Interval Order

Theorem 2. *The nearest neighbor Spearman footrule distance of a total and an interval order can be computed in linear time.*

Let α be an interval order on a domain $\mathcal{D}$ with an interval $[l_x, r_x]$ for each candidate $x \in \mathcal{D}$, and let σ be a total order on $\mathcal{D}$. Then the following algorithm computes a total order $\tau^* \in \mathrm{Ext}(\alpha)$ with $F(\tau^*, \sigma) = F_{NN}(\alpha, \sigma)$.

The algorithm successively builds τ^* taking $|\mathcal{D}|$ steps. For $k = 1, \ldots, |\mathcal{D}|$ it determines $x \in \mathcal{D}$ with $\tau^*(x) = k$. We will refer to this as *x is placed at position k*.

In each step k the algorithm holds the set $\mathcal{A}_k$ of α-admissible candidates consisting of all not yet processed candidates x, for which all candidates y with $y \prec_\alpha x$ have already been processed. Due to the specification of the α-admissible candidates, $\tau^* \in \mathrm{Ext}(\alpha)$ holds. $\mathcal{L}_k$ contains all *late* candidates $x \in \mathcal{A}_k$, whose contribution to $F(\tau^*, \sigma)$ increases by one in the $k+1$-th step if x is not placed in the k-th step. $\mathcal{E}_k$ contains all *early* candidates $x \in \mathcal{A}_k$, whose contribution will decrease by one. If there are any late candidates, the algorithm places any at position k. Otherwise it chooses the early candidate x with the smallest right interval boundary r_x.

Input: Interval order α, total order σ on a domain $\mathcal{D}$
Output: Total order $\tau^* \in \mathrm{Ext}(\alpha)$ with $F(\tau^*, \sigma) = F_{NN}(\alpha, \sigma)$

foreach $x \in \mathcal{D}$ **do** set $\tau^*(x) \leftarrow \perp$;
for $k = 1, \ldots, |\mathcal{D}|$ **do**
  $\mathcal{A}_k = \{x \in \mathcal{D} : \tau^*(x) = \perp \wedge \forall_{y \prec_\alpha x} \tau^*(y) \neq \perp\}$;
  $\mathcal{L}_k = \{x \in \mathcal{A}_k : \sigma(x) \leq k\}$;
  $\mathcal{E}_k = \{x \in \mathcal{A}_k : \sigma(x) > k\}$;
  if $\mathcal{L}_k \neq \emptyset$ **then**
    choose an arbitrary $x \in \mathcal{L}_k$ and set $\tau^*(x) \leftarrow k$;
  else
    choose an arbitrary $x \in \mathcal{E}_k$ with $r_x = \min_{y \in \mathcal{E}_k} r_y$ and set $\tau^*(x) \leftarrow k$;
return τ^*;

Algorithm 1: Computing F_{NN} of an interval order and a total order

To prove the correctness of Algorithm 1, we consider the set of *optimal* orders $\tau \in \mathrm{Ext}(\alpha)$ with $F(\tau, \sigma) = F_{NN}(\alpha, \sigma)$.

Lemma 5. *The total order τ^* computed by Algorithm 1 is optimal.*

Proof. Choose any optimal order τ_1 that, considering τ_1 and τ^* as permutations on $\mathcal{D}$, coincides with τ^* in the longest prefix. That means, τ_1 maximizes the quantity z such that $s \leq z \Rightarrow \tau^{*-1}(s) = \tau_1^{-1}(s)$. If $z = |\mathcal{D}|$, we are done; so suppose by contradiction $z < |\mathcal{D}|$ and consider the candidate x having $\tau^*(x) = \tau_1(x) = z$, and the candidate y having $\tau^*(y) = z + 1$ and $\tau_1(y) > z + 1$. In the following, we show that a total order τ_2, which is derived from τ_1 by shifting and switching operations on y, thus having $s \leq z + 1 \Rightarrow \tau^{*-1}(s) = \tau_2^{-1}(s)$, is also optimal. This contradicts the fact that τ_1 maximizes z.

In the following let $\mathcal{X} = \{c \in \mathcal{D} : \tau_1(x) < \tau_1(c) < \tau_1(y)\}$, which intuitively means that $\mathcal{X}$ contains all candidates that are ranked between x and y by τ_1.

Case 1: $y \in \mathcal{L}_{z+1}$ holds, as Algorithm 1 placed y at position $z+1$ in τ^*. Thus $\sigma(y) \leq z + 1$. Now let τ_2 be the total order derived from τ_1 by shifting y down to position $z + 1$, causing each $c \in \mathcal{X}$ being shifted up by one position (see Fig. 1). As for each $c \in \mathcal{X}$, $y \prec_{\tau^*} c$, but $c \prec_{\tau_1} y$, and as $\tau^* \in \mathrm{Ext}(\alpha)$ and $\tau_1 \in \mathrm{Ext}(\alpha)$ both hold, clearly $y \not\asymp_\alpha c$. Therefore, shifting y did not cause τ_2 to contradict α and $\tau_2 \in \mathrm{Ext}(\alpha)$ holds.

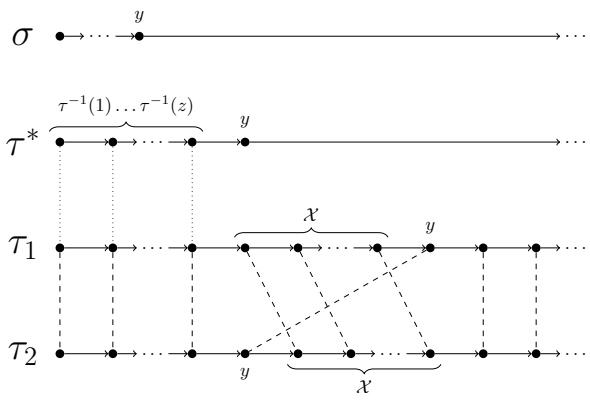

Fig. 1. σ, τ^*, τ_1 and τ_2 as they appear in Case 1 of Lemma 5.

Compare $F(\tau_2, \sigma)$ and $F(\tau_1, \sigma)$. We have $\tau_2(c) = \tau_1(c)$ for each $c \in \mathcal{D} \setminus (\mathcal{X} \cup \{y\})$, $\tau_2(c) = \tau_1(c) + 1$ for each $c \in \mathcal{X}$, and $\tau_2(y) = \tau_1(y) - |\mathcal{X}|$. Therefore the contribution of each $c \in \mathcal{X}$ to $F(\tau_2, \sigma)$ might increase by one compared to its contribution to $F(\tau_1, \sigma)$. On the other hand, as $\tau_2(y) = z+1$, $\tau_1(y) = z+1+|X|$ and $\sigma(y) \leq z+1$, the contribution of y to $F(\tau_2, \sigma)$ decreases by $|\mathcal{X}|$, such that $F(\tau_2, \sigma) \leq F(\tau_1, \sigma)$, and thus τ_2 is optimal, too.

Case 2: $\mathcal{L}_{z+1} = \emptyset$ and therefore $x \in \mathcal{E}_{z+1}$ held, as Algorithm 1 placed y at position $z+1$ in τ^*.

We first show that $\mathcal{X} \subseteq \mathcal{A}_{z+1}$, from which $\mathcal{X} \subseteq \mathcal{E}_{z+1}$ follows immediately. Suppose for contradiction that there exists some $c \in \mathcal{X}$ such that $c \notin \mathcal{A}_{z+1}$. That means, there exists at least one candidate c' which is α-admissible at step $z+1$, but prevents c from being α-admissible as $c' \prec_\alpha c$. Thus $c' \in \mathcal{E}_{z+1}$ as $\mathcal{L}_{z+1} = \emptyset$. As the algorithm picked y instead of c' at step $z+1$, $r_y \leq r_{c'}$ (see line 9 of Algorithm 1). But then $y \prec_\alpha c$, which yields a contradiction to $c \prec_{\tau_1} y$, although $\tau_1 \in \mathrm{Ext}(\alpha)$.

From that we derive two important facts: First $\sigma(c) > z+1$ for all $c \in \mathcal{X}$, and second all candidates from $\mathcal{X} \cup \{y\}$ are pairwise unrelated in α as otherwise they could not be within the α-admissible candidates at the same time.

We now derive τ_2 from τ_1 by a sequence of switching operations (see Fig. 2). Let $c_1 \in \mathcal{X}$ be the candidate having $\tau_1(c_1) = z+1$. Now switch y and c_1. If $\sigma(c_1) \geq z+1+|\mathcal{X}|$, we are done. If otherwise $z+1 < \sigma(c_1) < z+1+|\mathcal{X}|$, let $c_2 \in \mathcal{X}$ be the candidate $\sigma(c_1) = \tau_1(c_2)$ and switch c_1 and c_2. The repetition of this procedure will finish as soon as we find a candidate c_i having $\sigma(c_i) \geq z+1+|\mathcal{X}|$ (which according to the pidgeon hole principle will happen).

As we only performed switching operations concerning candidates from $\mathcal{X} \cup \{y\}$, which are pairwise unrelated in α, τ_2 does not contradict α, and thus $\tau_2 \in \mathrm{Ext}(\alpha)$.

Compare $F(\tau_2, \sigma)$ and $F(\tau_1, \sigma)$. For each $c \in \mathcal{D} \setminus (\mathcal{X} \cup \{y\})$ and for each $c \in \mathcal{X}$ which has not been moved by a switching operation, we have $\tau_2(c) = \tau_1(c)$. As y has been shifted down by $|\mathcal{X}|$ positions we have $\tau_2(y) = \tau_1(y) - |\mathcal{X}|$, which means that the contribution of y to $F(\tau_2, \sigma)$ might increase by $|\mathcal{X}|$ compared

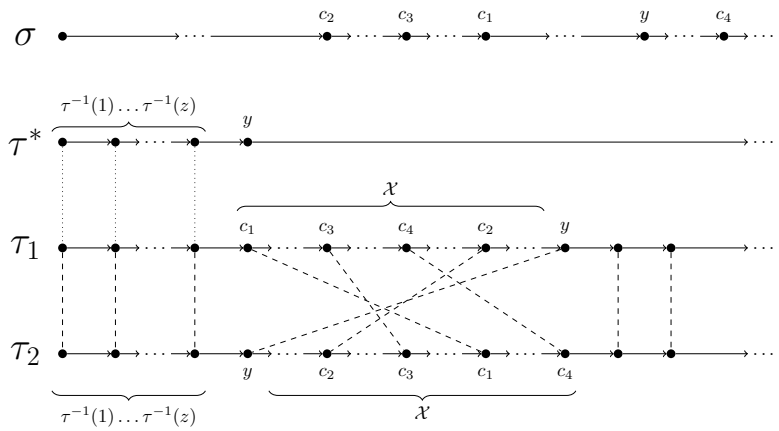

Fig. 2. σ, τ^*, τ_1 and τ_2 as they appear in Case 2 of Lemma 5. Note that $\tau_2(y) = \tau_1(c_1)$, $\tau_2(c_1) = \tau_1(c_2)$, $\tau_2(c_2) = \tau_1(c_3)$, $\tau_2(c_3) = \tau_1(c_4)$, $\tau_2(c_4) = \tau_1(y)$.

to its contribution to $F(\tau_1, \sigma)$. Finally, for each $c \in \mathcal{X}$ that has been moved i positions in a switching operation, we have $\tau_2(c) = \tau_1(c) \pm i$. As each of these candidates has been moved i positions closer to the position it is ranked by σ, its contribution to $F(\tau_2, \sigma)$ decreases by i compared to its contribution to $F(\tau_1, \sigma)$. Summing up the number of positions each $c \in \mathcal{X}$ has been moved, we clearly have a quantity larger than or equal to $|\mathcal{X}|$, as we start with candidate c_1 having $\tau_1(c_1) = z+1$ and place the candidate in the final switching operation at position $z + 1 + |\mathcal{X}|$. Thus $F(\tau_2, \sigma) \leq F(\tau_1, \sigma)$ and therefore τ_2 is optimal, too. □

For the linear run time, instead of rebuilding $\mathcal{A}_k$, $\mathcal{L}_k$ and $\mathcal{E}_k$ at each step, we hold them implicitly in an array $a[]$ of length $|\mathcal{D}|$, in which the beginning (resp. the end) of the interval of each not yet placed candidate $x \in \mathcal{D}$ is stored at $a[i]$ iff $l_x = i$ (resp. iff $r_x = i$), and a pointer p on the smallest r_x of all α-admissible candidates. Recall that the boundaries of the intervals of α are integers between 1 and $|\mathcal{D}|$, so that $a[]$ can be initialized via bucket sort. $a[]$ and p can be updated within each step in amortized $\mathcal{O}(1)$ time steps, as each candidate only once is removed from $a[i]$, becomes α-admissible, and switches from early to late during the execution of the algorithm.

Theorem 2 now follows immediately from Lemma 5 and from the fact that Algorithm 1 as well as the computation of the Spearman footrule distance on total orders can be implemented to run in linear time.

3.3 Nearest Neighbor Spearman Footrule Distance of a Total and a Partial Order

A partial order completely changes the picture, and shows a sharp separation between an interval and a partial order, when the distance to a total order is of concern. By a reduction from Clique [14] we show:

Theorem 3. *The distance problem for the nearest neighbor Spearman footrule distance of a total and a partial order is* **NP***-complete.*

Let a graph $G = (\mathcal{V}, \mathcal{E})$ with $\mathcal{V} = \{v_1, \ldots, v_n\}$ and $\mathcal{E} = \{e_1, \ldots, e_m\}$ and a positive integer k be an instance of Clique. Clearly Clique remains **NP**-complete for $n \geq 6$ and $k \geq 3$. For convenience let $k^* = k + \binom{k}{2}$. Furthermore, Clique remains **NP**-complete for $m \geq k^*$, as otherwise we add pairs of vertices v_i', v_i'' and edges $\{v_i', v_i''\}$ for $1 \leq i \leq k^*$ to $\mathcal{V}$ and $\mathcal{E}$. We will therefore assume $n > 3$, $k \geq 3$ and $m \geq k^*$.

We reduce to an instance of the distance problem, i. e., a domain $\mathcal{D}$, a partial order κ and a total order σ on $\mathcal{D}$, and a positive integer $k' \in \mathbb{N}$ as follows. We use $\mathcal{V}$ and $\mathcal{E}$ as sets of candidates, introduce two additional sets of candidates $\mathcal{B} = \{b_1, \ldots, b_{n^8}\}$ and $\mathcal{F} = \{f_1, \ldots, f_{m-k^*}\}$ and let $\mathcal{D} = \mathcal{V} \cup \mathcal{E} \cup \mathcal{B} \cup \mathcal{F}$.

Now construct $\sigma = [\mathcal{E}] \prec_\sigma [\mathcal{B}] \prec_\sigma [\mathcal{V}] \prec_\sigma [\mathcal{F}]$ with $\mathcal{V}$, $\mathcal{E}$, $\mathcal{B}$ and $\mathcal{F}$ each being consecutively totally ordered by σ. κ is constructed as follows: $\mathcal{F}$ is consecutively totally ordered by κ, while $\mathcal{V}$, $\mathcal{E}$ and $\mathcal{B}$ are each unrelated by κ. Furthermore $b \not\gtrless_\kappa c$ for each $b \in \mathcal{B}$ and $c \in \{\mathcal{V} \cup \mathcal{E} \cup \mathcal{F}\}$ and $f \prec_\kappa c$ for each $f \in \mathcal{F}$ and $c \in \{\mathcal{V} \cup \mathcal{E}\}$. Finally, the most important part of κ is the specification for $\mathcal{V}$ and $\mathcal{E}$. Here for each $v \in \mathcal{V}, e \in \mathcal{E}$, we set $v \prec_\kappa e$ if e is incident to v in G and $v \not\gtrless_\kappa e$, otherwise (we will refer to this as the incidence property). To complete the reduction we set $k' = \left(2m - 2\binom{k}{2}\right) n^8 + n^7$. For the specification of σ and κ see also Fig. 3.

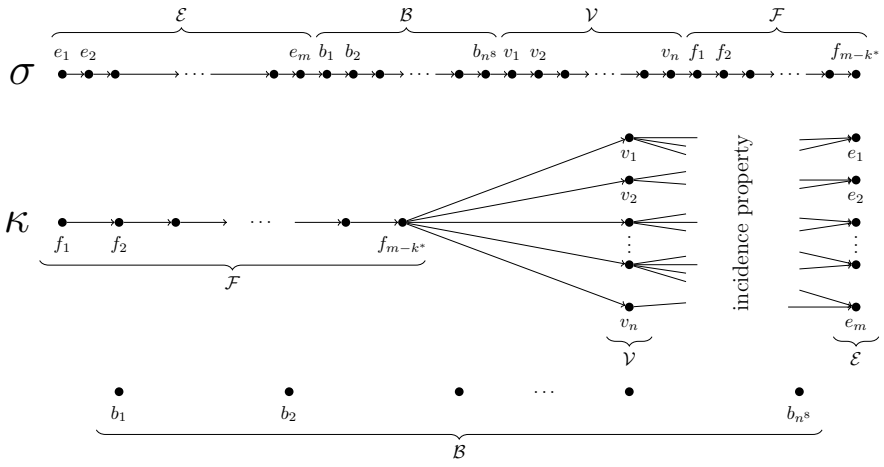

Fig. 3. κ and σ as they appear in Theorem 3.

We call a total order $\tau \in \mathrm{Ext}(\kappa)$ *optimal*, if $F(\tau, \sigma) = F_{NN}(\kappa, \sigma)$. Before verifying the correctness of the reduction, we start with a helpful lemma showing that there always is an optimal order τ which ranks each candidate of $\mathcal{B}$ at the same position as σ.

Lemma 6. *There exists an optimal order τ, such that $\tau^*(b) = \sigma(b)$ for all $b \in \mathcal{B}$.*

Proof. Choose any optimal order τ_1 that ranks the longest prefix of $b_1, \ldots, b_{n^8}$ in the same way as σ does, i. e., τ_1 maximizes the quantity z such that $s \leq z \Rightarrow \sigma(b_s) = \tau_1(b_s)$. If $z = n^8$, we are done, so suppose by contradiction $z < n^8$ and consider candidate b_{z+1}. In the following we show that a total order τ_2, which is derived from τ_1 by shifting and switching operations on candidate b_{z+1}, thus having $s \leq z+1 \Rightarrow \sigma(b_s) = \tau_1(b_s)$, is also optimal. This contradicts the fact that τ_1 maximizes z.

Case 1: Suppose $\tau_1(b_{z+1}) > \sigma(b_{z+1})$ and let $\mathcal{X} = \{c \in \mathcal{D} : \tau_1(b_z) < \tau_1(c) < \tau_1(b_{z+1})\}$, which intuitively means that $\mathcal{X}$ contains all candidates that are ranked between b_z and b_{z+1} by τ_1. Now let τ_2 be the total order derived from τ_1 by shifting b_{z+1} down to position $\tau_1(b_z) + 1 = \sigma(b_{z+1})$, causing each $c \in \mathcal{X}$ being shifted up by one position (see Fig. 4). As b_{z+1} is unrelated to all other candidates in κ, $\tau_2 \in \mathrm{Ext}(\kappa)$.

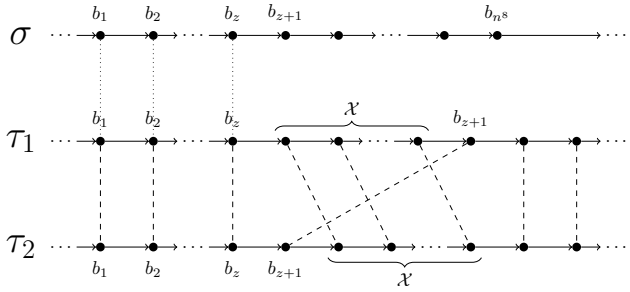

Fig. 4. σ, τ_1 and τ_2 as they appear in Case 1 of Lemma 6

Compare $F(\tau_2, \sigma)$ and $F(\tau_1, \sigma)$. We have $\tau_2(c) = \tau_1(c)$ for each $c \in \mathcal{D} \setminus (\mathcal{X} \cup \{b_{z+1}\})$, $\tau_2(c) = \tau_1(c)+1$ for each $c \in \mathcal{X}$, and $\tau_2(b_{z+1}) = \tau_1(b_{z+1})-|\mathcal{X}|$. Therefore the contribution of each $c \in \mathcal{X}$ to $F(\tau_2, \sigma)$ might increase by one compared to its contribution to $F(\tau_1, \sigma)$. On the other hand, as $\tau_2(b_{z+1}) = \sigma(b_{z+1})$, the contribution of b_{z+1} to $F(\tau_2, \sigma)$ decreases by $|\mathcal{X}|$, such that $F(\tau_2, \sigma) \leq F(\tau_1, \sigma)$ and thus τ_2 is optimal, too.

Case 2: Now suppose $\tau_1(b_{z+1}) < \tau_1(b_1)$ and let τ_1' be the total order derived from τ_1 by shifting b_{z+1} up to position $\tau_1(b_1) - 1$. With an argument analogous to Case 1 it can be shown that τ_1' is optimal.

Now let x be the element having $\tau_1'(x) = \sigma(b_{z+1})$ and let τ_2 be the total order derived from τ_1' by switching b_{z+1} and x (see Fig. 5). As the candidates ranked between b_{z+1} and x by τ_1' are exactly $b_1, \ldots, b_z$, which are each unrelated to all other candidates in κ, $\tau_2 \in \mathrm{Ext}(\kappa)$.

Comparing $F(\tau_2, \sigma)$ and $F(\tau_1', \sigma)$, we have $\tau_2(c) = \tau_1'(c)$ for each $c \in \mathcal{D} \setminus \{b_{z+1}, x\}$, $\tau_2(x) = \tau_1'(x) - (z+1)$ and $\tau_2(b_{z+1}) = \tau_1'(b_{z+1}) + z + 1$. Therefore the contribution of x to $F(\tau_2, \sigma)$ might increase by $z+1$ compared to its contribution to $F(\tau_1', \sigma)$. On the other hand, as $\tau_2(b_{z+1}) = \sigma(b_{z+1})$, the contribution of b_{z+1} to

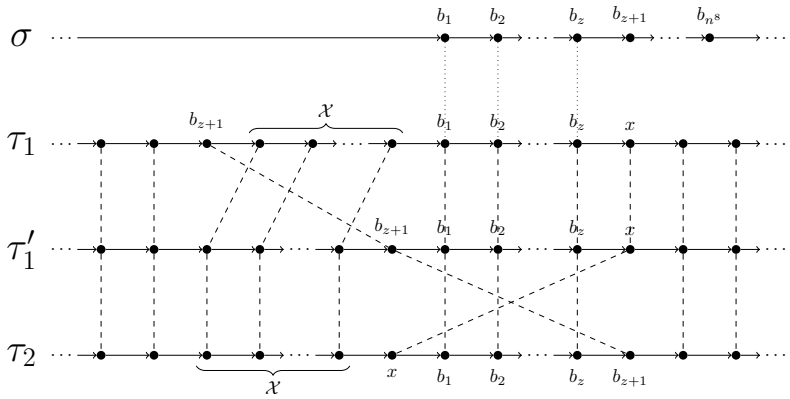

Fig. 5. σ, τ_1, τ_1' and τ_2 as they appear in Case 2 of Lemma 6

$F(\tau_2, \sigma)$ decreases by $z+1$, such that $F(\tau_2, \sigma) \leq F(\tau_1', \sigma)$ and thus τ_2 is optimal, too. □

Lemma 7. *G contains a clique of size at least k iff $F_{NN}(\kappa, \sigma) \leq k'$.*

Proof. "⇒": First suppose G contains a clique of size k, i. e., a complete subgraph $G' = (\mathcal{V}', \mathcal{E}')$ with $|\mathcal{V}'| = k$ and therefore $|\mathcal{E}'| = \binom{k}{2}$. We now compute a total order τ^* on $\mathcal{D}$ and show that $\tau^* \in \mathrm{Ext}(\kappa)$ and $F(\tau^*, \sigma) \leq k'$. Let

$$\tau^* = [\mathcal{F}] \prec_{\tau^*} [\mathcal{V}'] \prec_{\tau^*} [\mathcal{E}'] \prec_{\tau^*} [\mathcal{B}] \prec_{\tau^*} [\mathcal{V} \setminus \mathcal{V}'] \prec_{\tau^*} [\mathcal{E} \setminus \mathcal{E}']$$

with $\mathcal{B}$ and $\mathcal{F}$ being consecutively totally ordered and $\mathcal{V}'$, $\mathcal{E}'$, $\mathcal{V} \setminus \mathcal{V}'$ and $\mathcal{E} \setminus \mathcal{E}'$ being arbitrarily totally ordered (see Fig. 6).

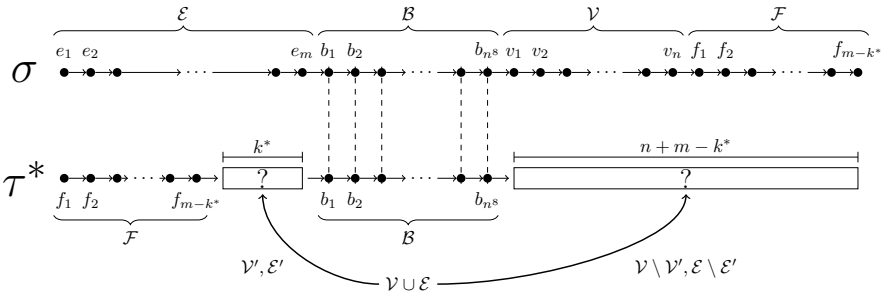

Fig. 6. σ and τ^* as they appear in Lemma 7.

To show that $\tau^* \in \mathrm{Ext}(\kappa)$, we have to verify that τ^* also has the incidence property, which means that no edge is ranked before its incident vertices by τ^*. This immediately follows from the fact that for each $e \in \mathcal{E}'$ both incident vertices

are within $\mathcal{V}'$. As both τ^* and κ consecutively totally order $\mathcal{F}$ and rank each $f \in \mathcal{F}$ before $\mathcal{V} \cup \mathcal{E}$ (which are the only remaining constraints of κ), we conclude $\tau^* \in \mathrm{Ext}(\kappa)$.

Considering $F(\tau^*, \sigma)$, it is easy to see that τ^* and σ both rank m candidates before b_1. As both consecutively totally order $\mathcal{B}$, we have $\tau^*(b) = \sigma(b)$ for all $b \in \mathcal{B}$ and thus the contribution of each $b \in \mathcal{B}$ to $F(\tau^*, \sigma)$ is zero. Due to its purpose in the proof, we will refer to $\mathcal{B}$ as *the blocker* in the following.

For all candidates $c \in \{\mathcal{V} \cup \mathcal{E} \cup \mathcal{F}\}$ we now distinguish whether they are ranked before the blocker by both τ^* and σ (*type 1*), ranked after the blocker by both τ^* and σ (*type 2*), or ranked before the blocker by τ^* and after the blocker by σ or vice versa (*type 3*). According to the definition of τ^* and σ (see again Fig. 6), all $e \in \mathcal{E}'$ are of type 1, all $v \in \mathcal{V} \setminus \mathcal{V}'$ are of type 2 and all $c \in \{\mathcal{F} \cup \mathcal{V}' \cup (\mathcal{E} \setminus \mathcal{E}')\}$ are of type 3. Summarized there are $n - k + \binom{k}{2} \leq n + m$ candidates of type 1 and 2, and $2m - 2\binom{k}{2}$ candidates of type 3. As both τ^* and σ rank m candidates before the blocker and $n + m - k^* \leq n + m$ candidates after the blocker, the contribution of a candidate of type 1 or 2 to $F(\tau^*, \sigma)$ is at most $n + m$, while the contribution of a single candidate of type 3 is at most $|\mathcal{D}| = n^8 + n + m + m - k^* \leq n^8 + n + 2m$. Summing up all these contributions and making use of the facts that $k \leq n$, $m \leq n^2$ and $n \geq 6$, we derive

$$F(\tau^*, \sigma) \leq (n+m)(n+m) + \left(2m - 2\binom{k}{2}\right)(n^8 + n + 2m) \leq k' \,.$$

As clearly $F_{NN}(\kappa, \sigma) \leq F(\tau^*, \sigma)$, we are done.

"$\Leftarrow$": Now suppose $F_{NN}(\kappa, \sigma) \leq k'$. Then there exists a total order $\tau^* \in \mathrm{Ext}(\kappa)$ with $F(\tau^*, \sigma) \leq k'$ and, according to Lemma 6, $\tau^*(b) = \sigma(b)$ for all $b \in \mathcal{B}$. Therefore, the contribution of each $b \in \mathcal{B}$ to $F(\tau^*, \sigma)$ is zero. Again we call $\mathcal{B}$ a blocker and classify the candidates of $\mathcal{V} \cup \mathcal{E} \cup \mathcal{F}$ into types 1, 2 and 3. Each candidate of type 3 contributes at least n^8 to $F(\tau^*, \sigma)$. As $F(\tau^*, \sigma) \leq k' = \left(2m - 2\binom{k}{2}\right) n^8 + n^7$, there are at most $\lfloor \frac{k'}{n^8} \rfloor = 2m - 2\binom{k}{2}$ candidates of type 3. All $m - k^*$ candidates of $\mathcal{F}$ are of type 3, because τ^*, being in $\mathrm{Ext}(\kappa)$, ranks all candidates from $\mathcal{F}$ before all candidates of $\mathcal{V} \cup \mathcal{E}$, of which some must be ranked before the blocker. Hence, there are at most $m + k - \binom{k}{2}$ candidates of type 3 within $\mathcal{V} \cup \mathcal{E}$. Again, according to the definition of κ and σ, we have that each $v \in \mathcal{V}$ is of type 3 iff τ^* ranks it before the blocker, while each $e \in \mathcal{E}$ is of type 3 iff τ^* ranks it after the blocker. Let $\mathcal{V}'$ be the set of candidates from $\mathcal{V}$ which are ranked before the blocker, and $\mathcal{E}'$ be the set of candidates from $\mathcal{E}$ which are ranked before the blocker by τ^*. As τ^* ranks m candidates before the blocker, of which $m - k^*$ are from $\mathcal{F}$, $|\mathcal{V}'| + |\mathcal{E}'| = k^*$.

Case 1: Suppose by contradiction that $|\mathcal{V}'| > k$ and $|\mathcal{E}'| < \binom{k}{2}$. Then there are $|\mathcal{V}'| + |\mathcal{E} \setminus \mathcal{E}'| = |\mathcal{V}'| + |\mathcal{E}| - |\mathcal{E}'| > k + m - \binom{k}{2}$ candidates of type 3, which yields a contradiction to the fact that there are at most $m + k - \binom{k}{2}$ candidates of type 3 within $\mathcal{V} \cup \mathcal{E}$.

Case 2: Suppose $|\mathcal{V}'| < k$ and $|\mathcal{E}'| > \binom{k}{2}$. As $\tau^* \in \mathrm{Ext}(\kappa)$, it has the incidence property and therefore each edge within $\mathcal{E}'$ is incident only to vertices within $\mathcal{V}'$.

This means that more than $\binom{k}{2}$ edges are only incident to less than k vertices – clearly a contradiction.

Thus, as $|\mathcal{V}'| = k$ and $|\mathcal{E}'| = \binom{k}{2}$ and as τ^* has the incidence property, each of the $\binom{k}{2}$ edges within $\mathcal{E}^*$ is incident to two of the k vertices within $\mathcal{V}^*$ and therefore $G' = (\mathcal{V}^*, \mathcal{E}^*)$ forms a clique of size k in G. □

Theorem 3 follows, since the above reduction runs in polynomial time and the containment of the distance problem in **NP** is straightforward.

4 Rank Aggregation Problem

The rank aggregation problem aims at finding a consensus ranking for a list of voters represented by partial orders. It is **NP**-hard for the Kendall tau distance [3] even for an even number of at least four voters represented by total orders [5, 12]. The **NP**-hardness also holds for related problems, such as computing top-k-lists [1] or determining winners [3, 4, 15, 22]. However, the rank aggregation problem for total orders under the Spearman footrule distance can be solved by a weighted bipartite matching, see [12]. We emphasize this result and show the **NP**-completeness for bucket orders by a reduction from Maximum Optimal Linear Arrangement (Max-Ola), which is reduced from Optimal Linear Arrangement (Ola) [14].

For a graph $G = (\mathcal{V}, E)$ with n vertices and m edges, and for a positive integer k, Ola asks whether or not there exists a permutation τ on $\mathcal{V}$ with $\sum_{\{u,v\}\in E} |\tau(u) - \tau(v)| \leq k$. Max-Ola is a modified version of Ola, in which we ask for a τ with $\sum_{\{u,v\}\in E} |\tau(u) - \tau(v)| \geq k$. It can be shown by induction that for a complete graph, $\sum_{\{u,v\}\in E} |\tau(u) - \tau(v)| = \frac{n^3-n}{6}$ for any τ. So we derive a reduction from Ola to Max-Ola, in which we make use of the complementary graph and ask for a τ' with $\sum_{\{u,v\}\in E} |\tau'(u) - \tau'(v)| \geq \frac{n^3-n}{6} - k$.

Theorem 4. *The rank aggregation problem for an arbitrary number of bucket orders under the Spearman footrule distance is* **NP***-complete.*

For the reduction from Max-Ola to the rank aggregation problem consider the vertices $\mathcal{V}$ as candidates and add two candidates x_1, x_2 with $x_1, x_2 \notin \mathcal{V}$, forming the domain $\mathcal{D} = \mathcal{V} \cup \{x_1, x_2\}$. Let $k' = 4nm + 4m - 2k$. There are two lists of bucket orders on $\mathcal{D}$, the *edge voters* Π_1 and the *dummy voters* Π_2. There are $k' + 1$ identical dummy voters π_s in Π_2. For $s \in \{1, \ldots, k' + 1\}$, $\pi_s = \{x_1\}\mathcal{V}\{x_2\}$. For each edge $\{u, v\} \in E$, Π_1 contains two bucket orders π_{uv} and π_{vu} with

$$\pi_{uv} = \{u\}(\mathcal{D} \setminus \{u, v\})\{v\} \text{ and } \pi_{vu} = \{v\}(\mathcal{D} \setminus \{u, v\})\{u\}\,.$$

Let the total order τ^* on $\mathcal{D}$ be any solution of the rank aggregation instance. The purpose of the dummy voters is to force any τ^* to rank x_1 and x_2 at the extremal positions 1 and $|\mathcal{D}|$. If $\tau^*(x_1) \neq 1$ or $\tau^*(x_2) \neq |\mathcal{D}|$, then for each dummy voter $\pi_s \in \Pi_2$ and for each total order $\sigma \in \mathrm{Ext}(\pi_s)$, we have $\sigma(x_1) = 1$ and $\sigma(x_2) = |\mathcal{D}|$, thus $F(\tau^*, \sigma) \geq 1$, which results in $\sum_{\pi_s \in \Pi_2} F_{NN}(\tau^*, \pi_s) > k'$. Thus

τ^* would violate the upper bound k' solely by considering the costs of the dummy voters. In the following suppose that τ^* satisfies the aforementioned necessary condition by $\tau^*(x_1) = 1$ and $\tau^*(x_2) = |\mathcal{D}|$. Then the dummy voters do not generate any costs, since $\tau^* \in \mathrm{Ext}(\pi_s)$, such that $F_{NN}(\tau^*, \pi_s) \leq F(\tau^*, \tau^*) = 0$.

Next we consider the costs contributed by the edge voters. Choose any single pair of edge-voters $\pi_{uv}, \pi_{vu} \in \Pi_1$. Following the proof of Theorem 1, $F_{NN}(\tau^*, \pi_{uv}) = F(\rho * \pi_{uv} * \tau^*, \rho * \tau^* * \pi_{uv})$ for an arbitrary total order ρ. As τ^* is a total order, we have $\rho * \pi_{uv} * \tau^* = \tau^*$ and $\rho * \tau^* * \pi_{uv} = \tau^* * \pi_{uv}$. Therefore $F_{NN}(\tau^*, \pi_{uv}) = F(\tau^*, \tau^* * \pi_{uv})$. With an analogous argument we get $F_{NN}(\tau^*, \pi_{vu}) = F(\tau^*, \tau^* * \pi_{vu})$. W. l. o. g. let $\tau^*(u) < \tau^*(v)$ (otherwise we switch the roles of u and v). Let $\mathcal{A} = \{w \in \mathcal{D} : 2 \leq \tau^*(w) < \tau^*(u)\}$, let $\mathcal{B} = \{w \in \mathcal{D} : \tau^*(u) < \tau^*(w) < \tau^*(v)\}$ and let $\mathcal{C} := \{w \in \mathcal{D} : \tau^*(v) < \tau^*(w) \leq |\mathcal{D}| - 1\}$. We use $[\mathcal{A}]$ to denote $\tau^{*-1}(2), \ldots, \tau^{*-1}(\tau^*(u) - 1)$ and use $[\mathcal{B}]$ and $[\mathcal{C}]$ in an analogous way. Then according to the definition of π_{uv} and π_{vu} in the above reduction, we have

$$\begin{aligned} \tau^* * \pi_{uv} &= u,\ x_1,\ [\mathcal{A}], [\mathcal{B}], [\mathcal{C}],\ x_2,\ v\,, \\ \tau^* * \pi_{vu} &= v,\ x_1,\ [\mathcal{A}], [\mathcal{B}], [\mathcal{C}],\ x_2,\ u\,, \text{ and} \\ \tau^* &= x_1, [\mathcal{A}],\ u,\ [\mathcal{B}],\ v,\ [\mathcal{C}], x_2\,. \end{aligned}$$

Thus we have a contribution of 2 to $F(\tau^*, \tau^* * \pi_{uv}) + F(\tau^*, \tau^* * \pi_{vu})$ for each $w \in \mathcal{A} \cup \mathcal{C} \cup \{x_1, x_2\}$, a contribution of 0 for each $w \in \mathcal{B}$, and a contribution of $|\mathcal{D}| - 1$ for each u and v. Observe that $|\mathcal{A}| = \tau^*(u) - 2$, $|\mathcal{B}| = \tau^*(v) - \tau^*(u) - 1$ and $|\mathcal{C}| = |\mathcal{D}| - \tau^*(v) - 1$.

Summing those quantities, considering $\tau^*(u) < \tau^*(v)$ and $|\mathcal{D}| = n + 2$, yields

$$\begin{aligned} F_{NN}(\tau^*, \pi_{uv}) + F_{NN}(\tau^*, \pi_{vu}) &= 2\,|\mathcal{A}| + 2\,|\mathcal{C}| + (|\mathcal{D}| - 1)\,|\{u, v\}| + 2\,|\{x_1, x_2\}| \\ &= 4\,|\mathcal{D}| - 4 + 2(\tau^*(u) - \tau^*(v)) \\ &= 4\,|\mathcal{D}| - 4 - 2\,|\tau^*(u) - \tau^*(v)| \\ &= 4n + 4 - 2\,|\tau^*(u) - \tau^*(v)|\ . \end{aligned}$$

Summing over all m pairs $\pi_{uv}, \pi_{vu} \in \Pi_1$ gives us

$$\sum_{\pi \in \Pi_1} F_{NN}(\tau^*, \pi) = 4nm + 4m - 2 \cdot \sum_{\pi_{uv}, \pi_{vu} \in \Pi_1} |\tau^*(u) - \tau^*(v)|\ .$$

Next we proof the correctness of the reduction.
"⇒": Suppose there is a permutation τ' on $\mathcal{V}$ such that $\sum_{\{u,v\} \in E} |\tau'(u) - \tau'(v)| \geq k$. From τ' we construct the permutation $\tau^* = x_1, \tau'^{-1}(1), \ldots, \tau'^{-1}(n), x_2$. As $\tau^*(x_1) = 1$ and $\tau^*(x_2) = |\mathcal{D}|$, $\sum_{\pi_s \in \Pi_2} F_{NN}(\tau^*, \pi_s) = 0$. Therefore,

$$\sum_{\pi \in \Pi} F_{NN}(\tau^*, \pi) = \sum_{\pi \in \Pi_1} F_{NN}(\tau^*, \pi) = 4nm + 4m - 2 \cdot \sum_{\pi_{uv}, \pi_{vu} \in \Pi_1} |\tau^*(u) - \tau^*(v)|\ .$$

Considering that $\tau^*(u) = \tau'(u) + 1$ and that $\tau^*(v) = \tau'(v) + 1$, and according to our assumption that $\sum_{\{u,v\} \in E} |\tau'(u) - \tau'(v)| \geq k$, we derive

$$\sum_{\pi \in \Pi_1} F_{NN}(\tau^*, \pi) = 4nm + 4m - 2 \cdot \sum_{\{u,v\} \in E} |\tau'(u) - \tau'(v)| \leq 4nm + 4m - 2k\,.$$

"⇐": Suppose there is a total order τ^* on $\mathcal{D}$ such that $\sum_{\pi\in\Pi} F_{NN}(\tau^*,\pi) \leq 4nm + 4m - 2k$. Due to the dummy voters, $\tau^*(x_1) = 1$ and $\tau^*(x_2) = |\mathcal{D}|$ and thus $\sum_{\pi_s\in\Pi_2} F_{NN}(\tau^*,\pi_s) = 0$ and $\sum_{\pi\in\Pi} F_{NN}(\tau^*,\pi) = \sum_{\pi\in\Pi_1} F_{NN}(\tau^*,\pi)$. We now construct a permutation τ' on $\mathcal{V}$ by setting $\tau'(u) = \tau^*(u) - 1$ for each $u \in \mathcal{V}$.

As $\tau^*(x_1) = 1$ and $\tau^*(x_2) = |\mathcal{D}|$,

$$\sum_{\pi\in\Pi_1} F_{NN}(\tau^*,\pi) = 4nm + 4m - 2 \cdot \sum_{\pi_{uv},\pi_{vu}\in\Pi_1} |\tau^*(u) - \tau^*(v)| \,.$$

According to our assumption on τ^* we derive

$$4nm + 4m - 2 \cdot \sum_{\pi_{uv},\pi_{vu}\in\Pi_1} |\tau^*(u) - \tau^*(v)| \leq 4nm + 4m - 2k$$

and from that

$$\sum_{\pi_{uv},\pi_{vu}\in\Pi_1} |\tau^*(u) - \tau^*(v)| \geq k \,.$$

Considering that $\tau^*(u) = \tau'(u) + 1$ and that $\tau^*(v) = \tau'(v) + 1$, we conclude

$$\sum_{\{u,v\}\in E} |\tau'(u) - \tau'(v)| \geq k \,.$$

From that we derive the correctness of the reduction.

Theorem 4 follows, since the reduction clearly runs in polynomial time and the containment of the rank aggregation problem in **NP** is straightforward.

5 Approximation algorithms

For total orders σ and τ, the Kendall tau and the Spearman footrule distances are related by the Diaconis-Graham inequality [11], which says that $K(\sigma,\tau) \leq F(\sigma,\tau) \leq 2K(\sigma,\tau)$. Fagin et al. [13] have extended this inequality to the Hausdorff distances on arbitrary sets (and thus for partial orders). With a proof similar to [13] we show that this inequality also holds for nearest neighbor distances of partial orders.

Theorem 5. *The Diaconis-Graham inequality holds for partial orders κ and μ under the nearest neighbor distances.*

$$K_{NN}(\kappa,\mu) \leq F_{NN}(\kappa,\mu) \leq 2K_{NN}(\kappa,\mu) \,.$$

Proof. Consider $\kappa',\kappa'' \in \mathrm{Ext}(\kappa)$ and $\mu',\mu'' \in \mathrm{Ext}(\mu)$, such that $F_{NN}(\kappa,\mu) = F(\kappa',\mu')$ and $K_{NN}(\kappa,\mu) = K(\kappa'',\mu'')$. Then

$$K_{NN}(\kappa,\mu) = K(\kappa'',\mu'') \leq K(\kappa',\mu') \leq F(\kappa',\mu') = F_{NN}(\kappa,\mu) \,.$$

where $K(\kappa'',\mu'') \leq K(\kappa',\mu')$ follows from the fact that $K_{NN}(\kappa,\mu) = K(\kappa'',\mu'')$ and $K(\kappa',\mu') \leq F(\kappa',\mu')$ is derived from the Diaconis-Graham inequality for total orders. Accordingly,

$$F_{NN}(\kappa,\mu) = F(\kappa',\mu') \leq F(\kappa'',\mu'') \leq 2K(\kappa'',\mu'') = 2K_{NN}(\kappa,\mu) \,.$$

Combining these inequalities completes the proof. □

Theorem 6. *Computing the nearest neighbor Spearman footrule distance between a partial and a total order is 6-approximable.*

Proof. We first consider the problem of computing the nearest neighbor Kendall tau distance between a partial order κ and a total order τ. Here, we intuitively ask for the total extension of κ, where as many ties as possible are broken according to τ. Thus, we transform κ and τ into a tournament graph as follows: For each candidate introduce a vertex and for each pair of vertices $u, v \in V$ introduce an edge $(u, v) \in E$ if $u \prec_\kappa v$ (κ-*edges*), or if $u \not\succ_\kappa v$ and $u \prec_\tau v$ (τ-*edges*). Clearly determining, whether the nearest neighbor Kendall tau distance of κ and τ is less or equal than k corresponds to asking whether there is a subset E' with $|E'| \leq k$ of the τ-edges, such that removing E' makes G acyclic. This is a special case of the constrained feedback arc set problem on tournament graphs, which is 3-approximable [21]. Theorem 5 now yields the result. □

Theorem 7. *The rank aggregation problem for bucket orders using the nearest neighbor Spearman footrule distance is 4-approximable by a deterministic algorithm and* 3*-approximable by a randomized algorithm.*

Proof. This follows immediately from Theorem 5 and a result of Ailon [1], who shows that the rank aggregation problem for bucket orders under the nearest neighbor Kendall tau distance is 2-approximable by a deterministic algorithm and $1,5$-approximable by a randomized algorithm. □

6 Conclusion and Open Problems

In this work we have investigated the nearest neigbor Spearman footrule distance on rankings with incomplete information. The incompleteness is expressed by bucket, interval and partial orders. The step from interval to partial implies a jump in the complexity from linear time to **NP**-completeness for the computation of the distance to a total order. Still open is the distance problem between two interval or an interval and a bucket order. Furthermore, there is the jump to **NP**-completeness for the rank aggregation problem from total to bucket orders. Our new **NP**-complete problems have good approximations. Our linear time algorithms, the **NP**-reductions, and the approximations used quite different techniques. It is left open to improve the given approximation ratios and to establish an approximation e.g., for the rank agregation problem for the general case with partial rankings. A further area of investigations addresses the Kendall tau distance and other measures, such as the Hausdorff distance [13].

References

1. N. Ailon. Aggregation of partial rankings, p-ratings and top-k lists. *Algorithmica*, 57:284–300, 2010.
2. J. A. Aslam and M. H. Montague. Models for metasearch. In *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 275–284. ACM, 2001.

3. J. J. Bartholdi III, C. A. Tovey, and M. A. Trick. Voting schemes for which it can be difficult to tell who won the election. *Social Choice and Welfare*, 6:157–165, 1989.
4. N. Betzler and B. Dorn. Towards a dichotomy for the possible winner problem in elections based on scoring rules. *Journal of Computer and System Sciences*, 76:812–836, 2010.
5. T. Biedl, F. J. Brandenburg, and X. Deng. On the complexity of crossings in permutations. *Discrete Mathematics*, 309:1813–1823, 2009.
6. J. C. Borda. Mémoire aux les élections au scrutin., 1781.
7. W. W. Cohen, R. E. Schapire, and Y. Singer. Learning to order things. *Journal of Artificial Intelligence Research (JAIR)*, 10:243–270, 1999.
8. M.-J. Condorcet. Éssai sur l'application de l'analyse à la probalité des décisions rendues à la pluralité des voix, 1785.
9. D. E. Critchlow. *Metric methods for analyzing partially ranked data.* Number 34 in Lecture notes in statistics. Springer, Berlin, 1985.
10. N. Cusanus. De arte eleccionis, 1299.
11. P. Diaconis and R. L. Graham. Spearman's footrule as a measure of disarray. *Journal of the Royal Statistical Society, Series B*, 39:262–268, 1977.
12. C. Dwork, R. Kumar, M. Naor, and D. Sivakumar. Rank aggregation methods for the web. In *Proceedings of the 10th International World Wide Web Conference (WWW10)*, pages 613–622, 2001.
13. R. Fagin, R. Kumar, M. Mahdian, D. Sivakumar, and E. Vee. Comparing partial rankings. *SIAM Journal on Discrete Mathematics*, 20:628–648, 2006.
14. M. R. Garey and D. S. Johnson. *Computers and Intractability; A Guide to the Theory of NP-Completeness.* W. H. Freeman & Co., New York, 1990.
15. E. Hemaspaandra, L. A. Hemaspaandra, and J. Rothe. Exact analysis of dodgson elections: Lewis Carroll's 1876 voting system is complete for parallel access to NP. *Journal of the ACM (JACM)*, 44:806–825, 1997.
16. G. Lebanon and J. D. Lafferty. Cranking: Combining rankings using conditional probability models on permutations. In *Machine Learning, Proceedings of the 19th International Conference (ICML)*, pages 363–370. Morgan Kaufmann, 2002.
17. R. Lullus. Artifitium electionis personarum, 1283.
18. M. H. Montague and J. A. Aslam. Condorcet fusion for improved retrieval. In *Proceedings of the 2002 ACM CIKM International Conference on Information and Knowledge Management*, pages 538–548. ACM, 2002.
19. M. E. Renda and U. Straccia. Web metasearch: Rank vs. score based rank aggregation methods. In *Proceedings of the 2003 ACM Symposium on Applied Computing (SAC)*, pages 841–846. ACM, 2003.
20. J. Sese and S. Morishita. Rank aggregation method for biological databases. *Genome Informatics*, 12:506–507, 2001.
21. A. van Zuylen and D. P. Williamson. Deterministic pivoting algorithms for constrained ranking and clustering problems. *Mathematics of Operations Research*, 34:594–620, 2009.
22. L. Xia and V. Conitzer. Determining possible and necessary winners under common voting rules given partial orders. In *Proceedings of the 23rd AAAI Conference on Artificial Intelligence*, pages 196–201. AAAI Press, 2008.
23. R. R. Yager and V. Kreinovich. On how to merge sorted lists coming from different web search tools. *Soft Computing Research Journal*, 3:83–88, 1999.