Comparing and Aggregating Partial Orders with Kendall Tau Distances

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-1102
Department of Informatics and Mathematics
University of Passau, Germany
February 2011

Comparing and Aggregating Partial Orders with Kendall Tau Distances

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.
In this work we consider the generalization of total orders and bucket orders to partial orders and compare them by the nearest neighbor and the Hausdorff Kendall tau distances. First, we establish an $\mathcal{O}(n \log n)$ algorithm for the computation of the nearest neighbor and the Hausdorff Kendall tau distances of two bucket orders. The computation of the nearest neighbor Kendall tau distance is **NP**-complete and 2-approximable for a total and a partial order. For the Hausdorff Kendall tau distance this problem is **coNP**-complete.
Considering rank aggregation problems with partial orders, we establish a significant discrepancy between the two distances. For the nearest neighbor Kendall tau distance the problem is **NP**-complete even for two voters, whereas the Hausdorff Kendall tau distance problem is in $\mathbf{\Sigma_2^p}$, but not in **NP** or **coNP** unless **NP** = **coNP**, even for four voters. However, both problems are known to be **NP**-complete for any even number of at least four total or bucket 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 and spam reduction [2, 14], and also biological databases, similarity search, and classification [11, 18, 21, 23, 25, 30]. It was mathematically investigated by Borda [7] and Condorcet [12] (18th century) and even by Lullus [19, 20] (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 (coequal) or incomparable (like apples and oranges). Voters considering all unrelated pairs of candidates or items as tied are represented by *bucket orders*, such that ties define an equivalence relation on candidates within a bucket. Bucket orders are also known as partial rankings, weak orders or preference rankings [1, 15, 17]. As incomparable pairs of candidates come into play, more general orders are needed and the voters describe their preferences by *partial orders*. In this case unrelatedness (ties and incomparabilities) is not transitive. All voters accept any order, which does not contradict their preferences without any 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 ($\not\gtrless$, meaning tied or incomparable) candidates in partial orders.

One of the common distance measures for two total orders σ and τ is the Kendall tau distance, $K(\sigma, \tau)$, which counts the number of disagreements between σ and τ regarding pairs of candidates. There are several other measures for orders or permutations, such as Spearman's footrule, Spearman's rho and the Hamming distance [13]. Investigations on ranking problems focused on total orders. Its generalization to bucket orders was considered recently by Ailon [1] and Fagin et al. [15]. The focus and main result in [15] is the equivalence of several distance measures, amongst others the Hausdorff Kendall tau distance, introduced by Critchlow [13]. Ailon [1] studied the nearest neighbor Kendall tau distance for bucket orders. The rank aggregation problem for total orders under the Kendall tau distance is **NP**-complete [3] even for an even number of at least four voters [6, 14]. The problem for two voters is efficiently solvable, while the complexity for three voters is an open problem. The **NP**-hardness also holds for related problems, such as computing top-k-lists [1] or determining winners [3, 4, 29]. Some determining winners problems are even known to be $\mathbf{\Theta_2^p}$-complete [17], where $\mathbf{\Theta_2^p}$ is the class of problems solvable via truth-table reducibility or parallel access to an **NP** oracle [9, 26, 28]. 2-approximations are known for the rank aggregation problem for total orders under the Kendall tau distance [14] and for bucket orders under the nearest neighbor Kendall tau distance [1]. Betzler et al. [5] provide results on the fixed-parameter tractability of several rank aggregation problems. Caragiannis et al. [10] establish the (in)approximability of some determining winners problems.

In this work we extend the nearest neighbor and the Hausdorff Kendall tau distances K_{NN} and K_H to partial orders. We establish a sharp separation between efficient algorithms and **NP**- and **coNP**-completeness. In particular, we show that the nearest neighbor Kendall tau distance can be computed in $\mathcal{O}(n \log n)$ for two bucket orders on sets of n candidates. The same result is known for the Hausdorff Kendall tau distance [15]. The $\mathcal{O}(n \log n)$ computation of the Kendall tau distance for two total orders has been shown in [6]. In contrast, the computation of the nearest neighbor resp. Hausdorff Kendall tau distance is **NP**-complete resp. **coNP**-complete for a total and a partial order. These proofs are based on a reduction from the Oscm-4-star problem [22]. Furthermore, we show that the computation of the nearest neighbor Kendall tau distance of a partial and a total order is a special case of the constrained feedback arc set problem on tournaments, and establish a 2-approximation. Our results on distance problems are summarized in Table 1.

Table 1. Distance problems

distance of	K_{NN}	K_H
a total and a total order	$\mathcal{O}(n \log n)$ ([6])	$\mathcal{O}(n \log n)$ ([6])
a bucket and a bucket order	$\mathcal{O}(n \log n)$ (Th. 1)	$\mathcal{O}(n \log n)$ (Th. 1, [15])
a total and a partial order	**NP**-complete (Th. 2) 2-approximable (Th. 4)	**coNP**-complete (Th. 3)

We then turn to rank aggregation problems. The problem is **NP**-complete for many total orders under the Kendall tau distance [3, 6, 14]. We study rank aggregation problems for voters represented by partial orders. Under the nearest neighbor Kendall tau distance, this problem is **NP**-complete even for two voters. In contrast, under the Hausdorff Kendall tau distance it is **NP**-hard and **coNP**-hard for at least four voters, and thus unlikely to be in **NP** or **coNP** unless **NP** = **coNP**. In fact, the problem is in $\mathbf{\Sigma_2^p}$, which is the class of problems solvable by an **NP** machine, which has access to an **NP** oracle. Our results on rank aggregation problems are summarized in Table 2.

Table 2. Rank aggregation problems

	voters represented by		
number of voters	total orders	partial orders under K_{NN}	partial orders under K_H
1	$\mathcal{O}(n)$ (trivial)	$\mathcal{O}(n)$ (trivial)	open
2	$\mathcal{O}(n)$ (trivial)	**NP**-complete (Th. 5)	**coNP**-hard (Th. 6)
3	open	**NP**-complete (Th. 5)	**coNP**-hard (Th. 6)
4	**NP**-complete ([3, 6, 14])	**NP**-complete (Th. 5)	**NP**- and **coNP**-hard (Th. 6)

This work is organized as follows. In Sect. 2 we introduce orders and distances. In Sect. 3 we consider the complexity of computing the nearest neighbor and the Hausdorff Kendall tau distances and establish the 2-approximability of the computation of the nearest neighbor Kendall tau distance of a total and a partial order. We address the complexity of rank aggregation problems in Sect. 4 and conclude with some open problems in Sect. 5.

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 by* κ if $x \not\prec_\kappa y \wedge y \not\prec_\kappa x$, which we denote by $x \not\gtrless_\kappa y$. The intuition of $x \prec_\kappa y$ is that κ *ranks* x *before* y, which means a preference for x. For a partial order κ on a domain $\mathcal{D}$ and 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 $x \not\gtrless_\kappa x'$ for all $x, x' \in \mathcal{X}$. 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 $\mathcal{D}$. Clearly, total $\subset$ bucket $\subset$ 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, $K(\sigma, \tau) = |\{\{x, y\} \subseteq \mathcal{D} : x \prec_\sigma y \wedge y \prec_\tau x\}|$.

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., $x \prec_\kappa y$ implies $x \prec_\tau y$ for all $x, y \in \mathcal{D}$. We denote the set of total extensions of a partial order κ with $\mathrm{Ext}(\kappa)$.

Definition 1. *For partial orders κ and μ on a domain $\mathcal{D}$ define the* nearest neighbor *and the* Hausdorff (Kendall tau) distances *via their extensions*

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

$$K_H(\kappa, \mu) = \max\{\max_{\tau \in \mathrm{Ext}(\kappa)} \min_{\sigma \in \mathrm{Ext}(\mu)} K(\tau, \sigma), \max_{\sigma \in \mathrm{Ext}(\mu)} \min_{\tau \in \mathrm{Ext}(\kappa)} K(\sigma, \tau)\}\,.$$

A Hausdorff distance of k says that there is an item in one set such that all items in the other set have a distance of at most k. In contrast, a nearest neighbor distance of k says that the closest pair of items has a distance of at most k. The Hausdorff distance is known to be a metric, but it takes the negative view and focuses on disagreements. The nearest neighbor distance takes the positive view and favors agreements. However, it fails the axioms of a metric. It does neither satisfy the identity of indiscernible $d(x, y) = 0 \Leftrightarrow x = y$ nor does the triangle inequality hold.

The following facts are obtained immediately from Definition 1.

Lemma 1. *For total orders τ and σ the nearest neighbor and the Hausdorff distances coincide with the common Kendall tau distance, i.e., $K_{NN}(\tau, \sigma) = K_H(\tau, \sigma) = K(\tau, \sigma)$.*

For a partial order κ and a total order σ, the nearest neighbor distance is the distance to a closest neighbor, i. e., $K_{NN}(\kappa, \sigma) = \min_{\tau \in \mathrm{Ext}(\kappa)} K(\tau, \sigma)$, and the Hausdorff distance is the distance to a farthest neighbor, i. e., $K_H(\kappa, \sigma) = \max_{\tau \in \mathrm{Ext}(\kappa)} K(\tau, \sigma)$.

Next we state our distance and rank aggregation problems.

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

Accordingly, the rank aggregation problem *under d asks whether or not for orders $\kappa_1, \ldots, \kappa_r$ on $\mathcal{D}$ 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.

3 Distance problems

3.1 Nearest Neighbor and Hausdorff Kendall Tau Distances of Two Bucket Orders

Fagin et al. [15] have characterized the Hausdorff Kendall tau distance of two bucket orders in terms of refinements. 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, x and y are tied in $\pi * \gamma$ iff they are tied in γ and in π. Clearly, if π is a total order then $\pi * \gamma$ breaks all ties and is a total order, too. $*$ is an associative operation, so for a third bucket order η on $\mathcal{D}$, $\eta * \pi * \gamma$ makes sense. Note that the refinement is only defined for bucket orders, but not for partial orders.

From the definition of the refinement operation Fagin et al. [15] obtain the following characterization of the Hausdorff distance.

Lemma 2. *[15] Let γ and π be bucket orders on the domain $\mathcal{D}$, and let ρ be any total order on $\mathcal{D}$. Let γ^R resp. π^R be the reversal of γ resp. π, which is obtained by reversing the order of the buckets, while the buckets are preserved. Hence, $x \prec_\gamma y$ iff $y \prec_{\gamma^R} x$. Then*

$$K_H(\gamma, \pi) = \max\{K(\rho * \pi^R * \gamma, \rho * \gamma * \pi), K(\rho * \pi * \gamma, \rho * \gamma^R * \pi)\}\,.$$

Adapting the proof from [15] we obtain the corresponding characterization for the nearest neighbor Kendall tau distance. We directly reuse Lemma 3 and Lemma 4, which we state here without proof, and rephrase Lemma 5 to serve our purposes.

Lemma 3. *[15] 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 4. *[15] Let τ be a total order and let γ be a bucket order on the domain $\mathcal{D}$. Then the quantity $K(\tau, \sigma)$ taken over all $\sigma \in \mathrm{Ext}(\gamma)$ is minimized for $\sigma = \tau * \gamma$.*

Lemma 5. *Let π and γ be bucket orders and let ρ be a total order on the domain $\mathcal{D}$. Then the quantity $K(\sigma, \sigma * \gamma)$, taken over all $\sigma \in \operatorname{Ext}(\pi)$, is minimized if $\sigma = \rho * \gamma * \pi$.*

Proof. Note that for any $\sigma \in \operatorname{Ext}(\pi)$ 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 $K(\sigma, \sigma * \gamma)$. That means for all total orders τ,

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

from which the lemma follows.

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

Choose $\tau \in S$ minimizing $K(\tau, \rho * \gamma)$. Since $\rho * \gamma \notin S$, $\tau \neq \rho * \gamma$. Therefore, Lemma 3 guarantees that we can find a pair $x, y \in \mathcal{D}$ such that $\tau(y) = \tau(x) + 1$, but $\rho * \gamma(y) < \rho * \gamma(x)$. Construct 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 $\tau' \in S$ holds, which is a contradiction as τ is supposed to be the total order in S 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 $K(\tau' * \pi, \tau' * \pi * \gamma) = K(\tau * \pi, \tau * \pi * \gamma)$ and $\tau' \in S$.

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 $K(\tau' * \pi, \tau' * \pi * \gamma) = K(\tau * \pi, \tau * \pi * \gamma)$, and we again conclude that $\tau' \in S$.

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 again 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 $y \prec_{\tau * \rho * \gamma} x$. Thus there is exactly one more inversion between $\tau * \pi$ and $\tau * \pi * \gamma$ than between $\tau' * \pi$ and $\tau' * \pi * \gamma$. So we immediately obtain $K(\tau' * \pi, \tau' * \pi * \gamma) \leq K(\tau * \pi, \tau * \pi * \gamma)$ from which we conclude that $\tau' \in S$. ☐

To obtain a characterization of the nearest neighbor Kendall tau distance between two bucket orders in terms of refinements, we combine the results of Lemmas 4 and 5. Let $\sigma \in \operatorname{Ext}(\gamma)$ be fixed. Then by Lemma 4 the quantity $K(\sigma, \tau)$ for every $\tau \in \operatorname{Ext}(\pi)$ is minimized for $\tau = \sigma * \pi$.

By Lemma 5 for every $\sigma \in \operatorname{Ext}(\gamma)$, $K(\sigma, \sigma * \pi)$ is minimized for $\sigma = \rho * \pi * \gamma$. Therefore

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

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

Corollary 1. $K_{NN}(\gamma, \pi) = K(\rho * \pi * \gamma, \rho * \gamma * \pi)\,.$

Theorem 1. *For two bucket orders, the distance problems under the nearest neighbor and the Hausdorff Kendall tau distances can be solved in* $\mathcal{O}(n \log n)$ *time.*

Proof. Refinements can obviously be computed in linear time. As established by Biedl et al. [6], for two total orders σ and τ on a domain $\mathcal{D}$ of size n, $K(\sigma, \tau)$ is the number of crossings of the permutation graph for σ and τ, which can be counted in $\mathcal{O}(n \log n)$. □

3.2 Nearest Neighbor and Hausdorff Kendall Tau Distances of a Total and a Partial Order

We now turn to the general case including a partial order. This makes the computation of the Kendall tau distances intractable.

In detail, the computation of the nearest neighbor Kendall tau distance of a partial and a total order is **NP**-complete. This is proved by a reduction from OSCM-4-Star, the one-sided two-level crossing minimization problem for forests of 4-stars. From this result, we derive the **coNP**-completeness of the computation of the Hausdorff Kendall tau distance of a partial and a total order. The **NP**-completeness of OSCM-4-Star has been proven by Muñoz et al. [22] by a reduction from feedback arc set.

Theorem 2. *For a total and a partial order, the distance problem under the nearest neighbor Kendall tau distance is* **NP***-complete.*

Proof. We start with the definition of OSCM-4-Star. An instance of OSCM-4-Star (see Fig. 1) consists of a positive integer k, and an undirected forest of n 4-stars with its vertices placed on distinct positions on two levels. Each star i has a set $\mathcal{A}(i) = \{a_1(i), a_2(i), a_3(i), a_4(i)\}$ of vertices of degree one placed on the upper level and one vertex $a_*(i)$ of degree four placed on the lower level. Let $\mathcal{A} = \bigcup_{i \in \{1,\ldots,n\}} \mathcal{A}(i)$ and $\mathcal{A}_* = \bigcup_{i \in \{1,\ldots,n\}} a_*(i)$. The order of $\mathcal{A}$ on the upper level is fixed by a permutation σ, while the vertices of $\mathcal{A}_*$ can be permuted freely on the lower level. OSCM-4-Star asks if there is such a permutation τ of $\mathcal{A}_*$ causing at most k edge crossings.

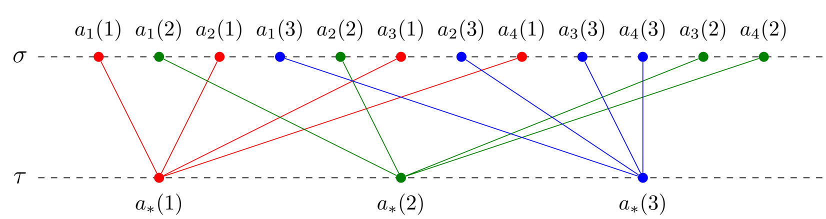

Fig. 1. Two-level drawing of 4-Stars.

We now reduce OSCM-4-Star to an instance of the distance problem, consisting of a partial order κ, a total order σ' on $\mathcal{D}$ and a positive integer k'.

Assume we are given an instance of OSCM-4-Star. We split each vertex $a_*(i)$ into four vertices of degree one as shown in Figure 2 and identify each of them with its adjacent vertex on the upper level such that each vertex from $\mathcal{A}$ now appears once on the upper and once on the lower level. We regard the resulting two-level drawing as a permutation graph, where two permutations σ and $\hat{\tau}$ are drawn as a two-level bipartite graph with the vertices (candidates) $\mathcal{A}$ on each level in the order given by σ and $\hat{\tau}$ and a straight-line edge between the two occurrences of each candidate $v \in \mathcal{A}$ on the two levels [6].

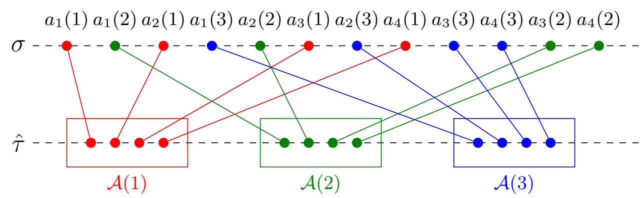

Fig. 2. Splitting the a_* star centers yields a permutation graph.

Note that the number of edge crossings in the permutation graph is equal to the Kendall tau distance of σ and $\hat{\tau}$.

Our objective is that the four candidates $a_1(i)$, $a_2(i)$, $a_3(i)$, $a_4(i)$ appear consecutively in $\hat{\tau}$ for each $i \in \{1, \ldots, n\}$ as then there is a direct correspondence between the number of crossings in the original OSCM-4-Star problem and $K(\sigma, \hat{\tau})$. This is equivalent to

$$\text{SCC} \quad \forall_{\substack{i,j \in \{1,\ldots,n\} \\ i \neq j}} \mathcal{A}(i) \prec_{\hat{\tau}} \mathcal{A}(j) \vee \mathcal{A}(j) \prec_{\hat{\tau}} \mathcal{A}(i)\,,$$

which we call the *separated candidate condition* (SCC).

We set $k' = 32n^2 + 24n + k$. To enforce SCC, we extend σ to σ' (and, correspondingly, solutions $\hat{\tau}$ to τ') by adding *blockers*. The nearest neighbor distance problem we reduce to then asks if there is a total order $\tau' \in \text{Ext}(\kappa)$ with $K(\tau', \sigma') \leq k'$. The blockers enforce that violating SCC causes the cost $K(\sigma', \tau')$ to exceed the upper bound k'. All solutions separating the gadgets, i. e., each $\mathcal{A}(i)$ together with the corresponding blockers, satisfy SCC. The blockers then incur a cost of $32n^2 + 24n$, such that only the crossing number k of the OSCM-4-Star instance determines whether or not the total cost maintains the upper bound k'. For each $i \in \{1, \ldots, n\}$ and $j \in \{1, 2\}$ we introduce the candidates $l_j(i)$, $l'_j(i)$, $r_j(i)$ and $r'_j(i)$. Let $\mathcal{L}_1(i) = \{l_1(i), l'_1(i)\}$, $\mathcal{L}_2(i) = \{l_2(i), l'_2(i)\}$, $\mathcal{R}_1(i) = \{r_1(i), r'_1(i)\}$, $\mathcal{R}_2(i) = \{r_2(i), r'_2(i)\}$, $\mathcal{L}(i) = \mathcal{L}_1(i) \cup \mathcal{L}_2(i)$, $\mathcal{R}(i) = \mathcal{R}_1(i) \cup \mathcal{R}_2(i)$, $\mathcal{L}_1 = \bigcup_{i \in \{1,\ldots,n\}} \mathcal{L}_1(i)$, $\mathcal{L}_2 = \bigcup_{i \in \{1,\ldots,n\}} \mathcal{L}_2(i)$, $\mathcal{R}_1 = \bigcup_{i \in \{1,\ldots,n\}} \mathcal{R}_1(i)$, and $\mathcal{R}_2 = \bigcup_{i \in \{1,\ldots,n\}} \mathcal{R}_2(i)$. We call $\mathcal{B}(i) = \mathcal{L}(i) \cup \mathcal{R}(i)$, $\mathcal{A}(i)$ and $\mathcal{G}(i) = \mathcal{L}(i) \cup \mathcal{A}(i) \cup \mathcal{R}(i)$ the *blockers of* i, the *inner elements of* i and *gadget* i respectively. Let $\mathcal{D} = \bigcup_{i \in \{1,\ldots,n\}} \mathcal{G}(i)$ be the set of candidates. Given n 4-stars, $\mathcal{D}$ has $12n$ candidates.

Now define the partial order κ on $\mathcal{D}$ as follows:

$$\underset{i\in\{1,\dots,n\}}{\forall} a_1(i) \prec_\kappa a_2(i) \prec_\kappa a_3(i) \prec_\kappa a_4(i)$$

$$\underset{x\in\{l,r\}}{\forall} \underset{i\in\{1,\dots,n\}}{\forall} x_1(i) \prec_\kappa x_1'(i) \prec_\kappa x_2(i) \prec_\kappa x_2'(i)$$

$$\underset{i\in\{1,\dots,n\}}{\forall} \mathcal{L}(i) \prec_\kappa \mathcal{A}(i) \prec_\kappa \mathcal{R}(i)$$

$$\underset{i,j\in\{1,\dots,n\}}{\forall} \underset{g_i\in\mathcal{G}(i)}{\forall} \underset{g_j\in\mathcal{G}(j)}{\forall} i \neq j \Rightarrow g_i \not\prec_\kappa g_j$$

Thus each gadget $\mathcal{G}(i)$ is totally ordered by κ, while elements of different gadgets are unrelated. Let the total order σ' on $\mathcal{D}$ be defined by

$$\underset{x\in\{l_1,l_2,r_1,r_2\}}{\forall} \underset{i\in\{1,\dots,n\}}{\forall} x(i) \prec_{\sigma'} x'(i)$$

$$\underset{i,j\in\{1,\dots,n\}}{\forall} i < j \Rightarrow \mathcal{L}_1(i) \prec_{\sigma'} \mathcal{L}_1(j) \wedge \mathcal{R}_2(i) \prec_{\sigma'} \mathcal{R}_2(j)$$

$$\underset{i,j\in\{1,\dots,n\}}{\forall} i > j \Rightarrow \mathcal{L}_2(i) \prec_{\sigma'} \mathcal{L}_2(j) \wedge \mathcal{R}_1(i) \prec_{\sigma'} \mathcal{R}_1(j)$$

$$\underset{a,a'\in\mathcal{A}}{\forall} a \prec_\sigma a' \Rightarrow a \prec_{\sigma'} a' \text{, and}$$

$$\mathcal{R}_2 \prec_{\sigma'} \mathcal{R}_1 \prec_{\sigma'} \mathcal{A} \prec_{\sigma'} \mathcal{L}_2 \prec_{\sigma'} \mathcal{L}_1 .$$

Before formally proving the correctness of our reduction, we analyse the solutions of the distance problem with respect to the number of crossings caused by blockers and inner elements. We use the following notation. For subsets $X, Y \subseteq \mathcal{D}$, let

$$\chi_{\tau'}^{\sigma'}(X,Y) = |\{(x,y) \in X \times Y : (\sigma'(x) - \sigma'(y))(\tau'(x) - \tau'(y)) < 0\}|$$

and let $\chi_{\tau'}^{\sigma'}(X) = \frac{1}{2}\chi_{\tau'}^{\sigma'}(X,X)$. $\chi_{\tau'}^{\sigma'}(X,Y)$ counts the number of pairs $(x,y) \in X \times Y$, where τ' and σ' disagree, i. e., where τ' ranks x before y and σ' ranks y before x or vice versa. Note that $K(\sigma',\tau') = \chi_{\tau'}^{\sigma'}(\mathcal{D})$.

The total number of crossings in the permutation graph consists of the crossings

- within each gadget, $\chi_{\tau'}^{\sigma'}(\mathcal{G}(i))$,
- between the blockers of one and the blockers of another gadget, $\chi_{\tau'}^{\sigma'}(\mathcal{B}(i),\mathcal{B}(j))$,
- between the inner elements of one and the blockers of another gadget, $\chi_{\tau'}^{\sigma'}(\mathcal{A}(i),\mathcal{B}(j))$, and
- between the inner elements of one and the inner elements of another gadget, $\chi_{\tau'}^{\sigma'}(\mathcal{A}(i),\mathcal{A}(j))$.

Claim 1. $\chi_{\tau'}^{\sigma'}(\mathcal{G}(i))$, the number of crossings within each gadget i, is fixed to 56.

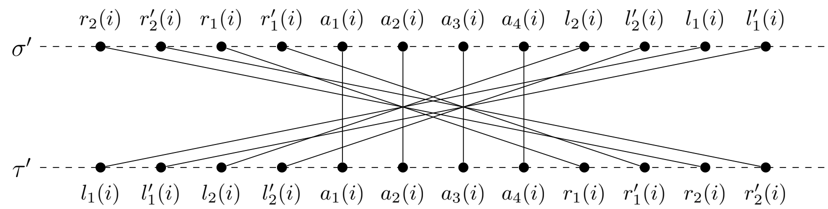

Fig. 3. $\chi^{\sigma'}_{\tau'}(\mathcal{G}(i)) = 56$ for each $i \in \{1, \dots, n\}$.

As each gadget i is totally ordered by κ, $\chi^{\sigma'}_{\tau'}(\mathcal{G}(i)) = 56$, as can be seen in Figure 3.

Claim 2. $\chi^{\sigma'}_{\tau'}(\mathcal{B}(i), \mathcal{B}(j))$, the number of crossings between the blockers of one and the blockers of another gadget, is 32 if τ' separates $\mathcal{G}(i)$ and $\mathcal{G}(j)$, i.e., $\mathcal{G}(i) \prec_{\tau'} \mathcal{G}(j) \vee \mathcal{G}(j) \prec_{\tau'} \mathcal{G}(i)$. Otherwise $\chi^{\sigma'}_{\tau'}(\mathcal{B}(i), \mathcal{B}(j)) > 32$.

Assume $i < j$. As each gadget is totally ordered, $\chi^{\sigma'}_{\tau'}(\mathcal{B}(i), \mathcal{B}(j))$ solely depends on the "interleaving" of the two gadgets, i.e., which blocker of gadget i is placed in which *sector* (between which blockers) of gadget j (see Fig. 4). The sectors of gadget j are defined as follows.

sector	description
S_0	left of $l_1(j)$
S_1	between $l_1(j)$ and $l'_1(j)$
S_2	between $l'_1(j)$ and $l_2(j)$
S_3	between $l_2(j)$ and $l'_2(j)$
S_4	between $l'_2(j)$ and $r_1(j)$
S_5	between $r_1(j)$ and $r'_1(j)$
S_6	between $r'_1(j)$ and $r_2(j)$
S_7	between $r_2(j)$ and $r'_2(j)$
S_8	right of $r'_2(j)$

The number of crossings can be computed independently for each pair of blocker and sector.

We model the interleavings as a path problem (see Fig. 5). Taking an edge (S_m, b) in a selected path means placing b in sector S_m. The number of additional crossings is $w(S_m, b)$ if b is placed in sector S_m instead of sector S_0. If, for example, $r'_1(i)$ is placed in sector S_2, i.e., $l'_1(j) \prec_{\tau'} r'_1(i) \prec_{\tau'} l_2(j)$, it has $\chi^{\sigma'}_{\tau'}(\{r'_1(i)\}, \mathcal{B}(j)) = 6$ crossings, while it has $\chi^{\sigma'}_{\tau'}(\{r'_1(i)\}, \mathcal{B}(j)) = 4$ crossings if it is placed in sector S_0, i.e., $r'_1(i) \prec_{\tau'} l_1(j)$. Thus we have $w(S_2, r'_1(i)) = 6 - 4 = 2$. Each path in the graph corresponds to an interleaving of the gadgets i and j. If all blockers of gadget i are placed to the left of gadget j, we have $\chi^{\sigma'}_{\tau'}(\mathcal{B}(i), \mathcal{B}(j)) = 32$. There are two shortest paths in the interleaving graph, which correspond to the interleavings where $\mathcal{G}(i)$ and $\mathcal{G}(j)$ are separated by τ', completing the proof of the claim.

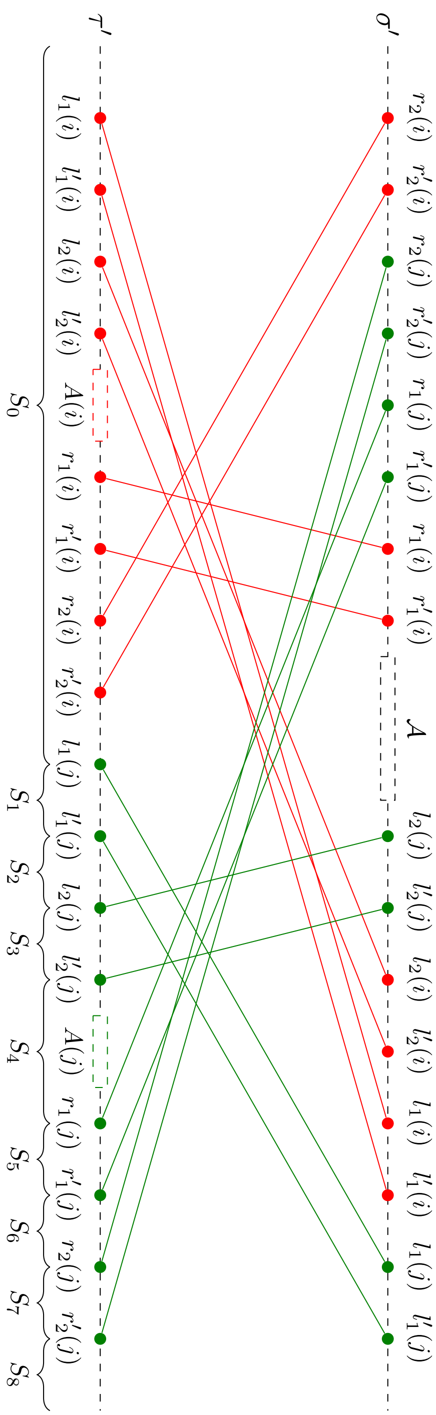

Fig. 4. Interleaving of two gadgets i and j. Each element of gadget i is placed by τ' in some of the sectors S_m. For clearness, in this example drawing all elements of gadget i are placed in sector S_0, i.e., left of $l_1(j)$.

Claim 3. $\chi^{\sigma'}_{\tau'}(\mathcal{A}(i), \mathcal{B}(j))$, the number of crossings between the inner elements of one and the blockers of another gadget, is 16 if τ' separates $\mathcal{A}(i)$ and $\mathcal{B}(j)$. Otherwise $\chi^{\sigma'}_{\tau'}(\mathcal{A}(i), \mathcal{B}(j)) > 16$.

Each inner element of $\mathcal{A}(i)$ crosses all elements of either $\mathcal{L}(j)$ or $\mathcal{R}(j)$, which implies $\chi^{\sigma'}_{\tau'}(\mathcal{A}(i), \mathcal{B}(j)) = 16$. Each inner element $a \in \mathcal{A}(i)$ which is not separated from $\mathcal{A}(j)$, implies $\mathcal{L}(j) \prec_{\tau'} a \prec_{\tau'} \mathcal{R}(j)$, and causes an additional cost of 4, since it intersects with the elements of both $\mathcal{L}(j)$ and $\mathcal{R}(j)$.

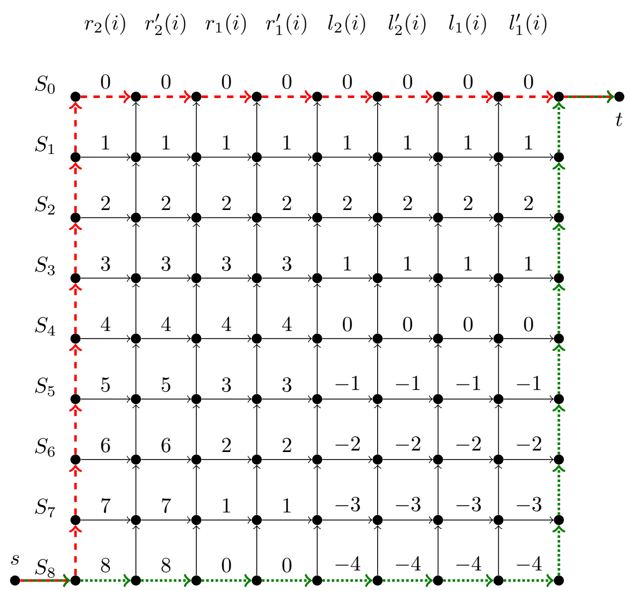

Fig. 5. Graph modeling the cost of interleaving two gadgets. The red (dashed) path corresponds to the interleaving shown in Figure 4. The red (dashed) and the green (dotted) paths are the only paths of minimal length and correspond to $\mathcal{G}(i) \prec_{\tau'} \mathcal{G}(j)$ and $\mathcal{G}(j) \prec_{\tau'} \mathcal{G}(i)$ respectively.

With the help of the above claims, we now establish the correctness of the reduction.

First, suppose there exists a solution τ for OSCM-4-Star with at most k crossings. From τ we derive a total order $\tau' \in \mathrm{Ext}(\kappa)$ on $\mathcal{D}$ with $K(\tau', \sigma') \leq k'$. In detail, let τ' be the total order that separates the gadgets and orders them according to the corresponding star centers in τ. The total number of crossings between τ' and σ' is exactly

$$\chi_{\tau'}^{\sigma'}(\mathcal{D}) = \sum_{i \in \{1,\ldots,n\}} \chi_{\tau'}^{\sigma'}(\mathcal{B}(i)) + \sum_{\substack{i,j \in \{1,\ldots,n\} \\ i<j}} \chi_{\tau'}^{\sigma'}(\mathcal{B}(i), \mathcal{B}(j)) + \\ + \sum_{\substack{i,j \in \{1,\ldots,n\} \\ i \neq j}} \chi_{\tau'}^{\sigma'}(\mathcal{A}(i), \mathcal{B}(j)) + \sum_{\substack{i,j \in \{1,\ldots,n\} \\ i<j}} \chi_{\tau'}^{\sigma'}(\mathcal{A}(i), \mathcal{A}(j)) \,.$$

We now make use of our claims, taking into account that for each $i, j \in \{1, \ldots, n\}, i \neq j$, τ' separates $\mathcal{G}(i)$ and $\mathcal{G}(j)$. Thus

$$\sum_{i \in \{1,\ldots,n\}} \chi_{\tau'}^{\sigma'}(\mathcal{B}(i)) = 56n \,,$$

$$\sum_{\substack{i,j \in \{1,\ldots,n\} \\ i<j}} \chi_{\tau'}^{\sigma'}(\mathcal{B}(i), \mathcal{B}(j)) = 32 \frac{n(n-1)}{2} \,, \text{ and}$$

$$\sum_{\substack{i,j \in \{1,\ldots,n\} \\ i \neq j}} \chi_{\tau'}^{\sigma'}(\mathcal{A}(i), \mathcal{B}(j)) = 16n(n-1) \,.$$

Additionally, SCC holds, and thus, according to our assumption that τ causes at most k crossings in OSCM-4-Star,

$$\sum_{\substack{i,j \in \{1,\ldots,n\} \\ i<j}} \chi_{\tau'}^{\sigma'}(\mathcal{A}(i), \mathcal{A}(j)) \leq k.$$

Summing the above yields

$$K_{NN}(\kappa, \sigma') \leq K(\tau', \sigma') = \chi_{\tau'}^{\sigma'}(\mathcal{D}) \leq 56n + 32 \frac{n(n-1)}{2} + 16n(n-1) + k = \\ = 32n^2 + 24n + k = k' \,.$$

Secondly, suppose that $K_{NN}(\kappa, \sigma') \leq k'$. In detail, let $\rho \in \mathrm{Ext}(\kappa)$ be a total order on $\mathcal{D}$ with $K(\rho, \sigma') \leq k'$. If ρ satisfies SCC, we are done, as then the solution of the original instance of OSCM-4-Star corresponding to τ had

$$\sum_{\substack{i,j \in \{1,\ldots,n\} \\ i<j}} \chi_{\rho}^{\sigma'}(\mathcal{A}(i), \mathcal{A}(j)) \leq k' - 32n^2 - 24n = k$$

crossings by applying our claims. However, if ρ does not satisfy SCC, we derive a total order $\tau' \in \mathrm{Ext}(\mathcal{D})$ from ρ as follows. τ' separates the gadgets and orders them as ρ ordered the respective a_1 candidates, i. e.,

$$\forall_{\substack{i,j \in \{1,\ldots,n\} \\ i \neq j}} \mathcal{G}(i) \prec_{\tau'} \mathcal{G}(j) \Leftrightarrow a_1(i) \prec_{\rho} a_1(j) \,.$$

τ' clearly satisfies SCC. If we show that $K(\tau', \sigma') \leq K(\rho, \sigma') \leq k'$ then the solution of the original OSCM-4-Star instance corresponding to τ' had at most k crossings. ρ has more crossings between pairs of blockers and at least as many crossings between blockers and inner elements as τ'. Therefore ρ must have fewer crossings between inner elements than τ'. Let a be any inner element and j be any gadget with $a \notin \mathcal{A}(j)$. In comparison to τ', ρ can save at most $\chi_{\tau'}^{\sigma'}(\{a\}, \mathcal{A}(j)) \leq 4$ crossings by placing a between the elements of $\mathcal{A}(j)$, but then we have $\chi_{\rho}^{\sigma'}(\{a\}, \mathcal{B}(j)) = \chi_{\tau'}^{\sigma'}(\{a\}, \mathcal{B}(j)) + 4$ as shown above. Thus ρ cannot achieve a total of fewer crossings than τ' even by the best possible placement of inner elements, completing our proof. ☐

Summarizing the above, we have proven that, given an instance of the OSCM-4-Star problem, we can construct an instance of the distance problem, such that there is a solution to OSCM-4-Star with at most k crossings iff there is a permutation $\tau' \in \mathrm{Ext}(\kappa)$ with $K(\sigma', \tau') \leq 32n^2 + 24n + k$. Thus the distance problem is **NP**-hard. As one can guess a permutation τ' and check if $\tau' \in \mathrm{Ext}(\kappa)$ and $K(\sigma', \tau') \leq k$ in polynomial time, the problem also is in **NP**. ☐

Theorem 3. *For a total and a partial order, the distance problem under the Hausdorff Kendall tau distance is* **coNP***-complete.*

Proof. Let a total order σ, a partial order κ on a domain $\mathcal{D}$, and a positive integer k be an instance of the distance problem under the nearest neighbor distance, which we have shown to be **NP**-complete (Theorem 2). From that we immediately obtain the **NP**-completeness of a *modified distance problem*. Here we are also given a total order σ', a partial order κ' on a domain $\mathcal{D}$, and a positive integer k', but now ask if there exists a total order τ' in $\mathrm{Ext}(\kappa')$ with $K(\tau', \sigma') \geq k'$. The modified distance problem is **NP**-complete as we set $\kappa' = \kappa$, $\sigma' = \sigma^R$ and $k' = \binom{|\mathcal{D}|}{2} - k$, where σ^R means the reverse of σ. Observe that for each $\tau' \in \mathrm{Ext}(\kappa)$, each of the $\binom{|\mathcal{D}|}{2}$ pairs of candidates will contribute exactly one either to $K(\tau', \sigma)$ or to $K(\tau', \sigma^R)$. Thus there exists a total order $\tau \in \mathrm{Ext}(\kappa)$ with $K(\tau, \sigma) \leq k$ iff there exists a total order $\tau' \in \mathrm{Ext}(\kappa')$ with $K(\tau', \sigma') = K(\tau', \sigma^R) \geq \binom{|\mathcal{D}|}{2} - k$, as we can simply set $\tau = \tau'$.

From the **NP**-completeness of the modified distance problem, the **coNP**-completeness of the distance problem under the Hausdorff distance of a total and a partial order follows. Let a total order σ'', a partial order κ'' on a domain $\mathcal{D}$, and a positive integer k'' be an instance of the distance problem under the Hausdorff distance. Since $K_H(\kappa'', \sigma'') = \max_{\tau'' \in \mathrm{Ext}(\kappa'')} K(\tau'', \sigma'')$ (Lemma 1), we ask if $K(\tau'', \sigma'') \leq k''$ holds for all $\tau'' \in \mathrm{Ext}(\kappa'')$. Therefore the distance problem under the Hausdorff distance of a total and a partial order is the complementary problem of the modified distance problem. The **NP**-completeness of a problem implies the **coNP**-completeness of the complementary problem [16]. ☐

Theorem 4. *For a total and a partial order, the distance problem under the nearest neighbor Kendall tau distance is 2-approximable.*

Proof. Considering the problem of computing the nearest neighbor distance of a partial order κ and a total order σ, we intuitively ask for the total extension τ of

κ where as many pairs $i \not\gtrless_\kappa j$ as possible are ordered according to σ. Thus, we transform κ and σ into a tournament graph $G = (V, E)$, i. e., a directed graph with either $(i, j) \in E$ or $(j, i) \in E$ for each $i, j \in V$: We introduce a vertex for each candidate, and for each pair of vertices $i, j \in V$ we introduce an edge $(i, j) \in E$ if $i \prec_\kappa j$ (κ-*edges*), or if $i \not\gtrless_\kappa j \wedge i \prec_\sigma j$ (σ-*edges*). Observe that each σ-edge $(i, j) \in E$ corresponds to two candidates i and j that are unrelated by κ, while the direction of the edge indicates in which way to break the unrelatedness according to σ. Then a total order $\tau \in \mathrm{Ext}(\kappa)$ minimizing $K(\tau, \sigma)$ corresponds to a permutation of the vertices, such that there is no κ-edge (i, j) with $j \prec_\tau i$ and such that the number of σ-edges (i, j) with $j \prec_\tau i$ is minimized. Clearly, determining a $\tau \in \mathrm{Ext}(\kappa)$ which minimizes $K(\tau, \sigma)$ corresponds to finding the smallest subset of σ-edges whose removal makes G acyclic. This is a special case of the constrained feedback arc set problem on tournaments, where we are given a tournament graph containing a partial order κ and ask for the smallest subset of edges not belonging to κ whose removal makes the graph acyclic. This problem is 3-approximable according to van Zuylen et al. [27]. Observe, that [27] also presents 2-approximable cases of the problem, but these demand different preconditions. The general idea of [27] is to derive a second tournament from an optimal solution of a linear program, which is a lower bound for the constrained feedback arc set problem. This tournament is then turned into the total order τ by the pivoting algorithm described below.

We now refine the analysis of the proof of [27], taking into account that in our special case the σ-edges are acyclic, since they are obtained from a total order, and obtain a 2-approximation. The remainder of this section thus is mostly due to [27], we just make some own refinements at the end of the proof.

We define edge weights for each distinct $i, j \in V$. If $(i, j) \in E$, let $w_{ij} = 1$ and $w_{ji} = 0$. If, in contrast, $(j, i) \in E$, let $w_{ij} = 0$ and $w_{ji} = 1$. The input of the pivoting algorithm is another tournament $G' = (V, E')$, which we obtain from the following linear program (LP). LP contains variables x_{ij} and x_{ji} for each distinct $i, j \in V$, which we use to determine E'.

LP: $\min \sum_{i<j} (x_{ij} w_{ji} + x_{ji} w_{ij})$

st.	(1)	$x_{ij} + x_{jk} + x_{ki} \geq 1$	for all distinct i, j, k
	(2)	$x_{ij} + x_{ji} = 1$	for all $i \neq j$
	(3)	$x_{ij} = 1$	for all $i \prec_\kappa j$
	(4)	$x_{ij} \geq 0$	for all $i \neq j$

Given an optimal solution x of LP we construct G', where $(i, j) \in E'$ only if $x_{ij} \geq \frac{1}{2}$. We will break ties in such a way as to ensure that there are no (directed) triangles containing a κ-edge, i. e., there is no $\{(i, j), (j, k), (k, i)\} \subseteq E'$ such that (i, j) is a κ-edge. Therefore use an arbitrary $\rho \in \mathrm{Ext}(\kappa)$ and in case that $x_{ij} = \frac{1}{2} = x_{ji}$, let $(i, j) \in E'$ iff $i \prec_\rho j$. Now suppose for contradiction, there is a triangle $\{(i, j), (j, k), (k, i)\} \subseteq E'$ such that (i, j) is a κ-edge. As (i, j) is a κ-edge, $x_{ij} = 1$ and $x_{ji} = 0$ due to conditions (2) and (3) of LP. Furthermore, $x_{jk}, x_{ki} \geq \frac{1}{2}$, because $(j, k), (k, i) \in E'$. As condition (1) must also hold for the reverse triangle $(i, k), (k, j), (j, i)$ and $x_{ji} = 0$, $x_{ik} + x_{kj} = 1$. Thus both x_{jk} and x_{ki} exactly take the value $\frac{1}{2}$. As the ties are broken according to ρ,

$\{(i,j),(j,k),(k,i)\} \subseteq E'$ implies $j \prec_\rho k$ and $k \prec_\rho i$. But this would imply $j \prec_\rho i$, which contradicts $i \prec_\rho j$ as $\rho \in \mathrm{Ext}(\kappa)$. Thus we have ensured that no κ-edge is contained in a triangle of E'.

The pivoting algorithm now recursively computes τ using G' to repeatedly find a pivot vertex. Later we will refer to the chosen vertex. Given a pivot vertex k, the algorithm puts vertex j to the left or right of k depending on $(j,k) \in E'$ or $(k,j) \in E'$. It then recurses on the set of vertices to the left and right of k. Note that G' is kept during the whole execution of the algorithm and LP is not solved again in the recursive calls.

For a pair of vertices i,j with $(i,j) \in E'$ the only way to have $j \prec_\tau i$ is being in the same recursive call, and a pivot k must be chosen such that $(j,k) \in E'$ and $(k,i) \in E'$. In other words, (i,j), (j,k) and (k,i) form a triangle in E'. Hence, $\tau \in \mathrm{Ext}(\kappa)$ as we have ensured that there are no triangles in E' containing a κ-edge.

Now we analyse the cost of our solution τ compared to the optimal solution of LP, which is a lower bound. Observe that for the sake of clarity, we consider the cost that occurs in the first call of the recursive algorithm. However, the analysis of the cost is analogous in the later recursive calls. The main idea of the analysis is to consider pairs $\{i,j\} \in V$ and compare the cost of $\{i,j\}$ in τ, i. e., w_{ji} if $i \prec_\tau j$ or w_{ij} if $j \prec_\tau i$, with the cost $c_{ij} = x_{ij}w_{ji} + x_{ji}w_{ij}$ of $\{i,j\}$ in the optimal solution of LP.

Let k be the pivot vertex. In this call, we consider the cost for pairs $\{j,k\}$ and for pairs $\{i,j\}$ such that i and j do not both end up on the same side of k. Note that if a cost is incurred for a pair of vertices, then no other cost is incurred for this pair in later iterations. Clearly, the cost we incur for a pair $\{j,k\}$ when k is the pivot, is at most $2c_{jk} = 2(x_{jk}w_{kj} + x_{kj}w_{jk})$. Similarly, if $\{(i,k),(k,j),(i,j)\} \subseteq E'$, then the cost for the pair $\{i,j\}$ is at most $2c_{ij} = 2(x_{ij}w_{ji} + x_{ji}w_{ij})$. Hence, the only problematic pairs are the *critical edges* (j,i) that are in a triangle with k in E', i. e., pairs such that $\{(i,k),(k,j),(j,i)\} \subseteq E'$ and the algorithm orders i before j, even though $(j,i) \in E'$. For the pivot k, let $T_k(E')$ denote the set of its critical edges, so $T_k(E') = \{(j,i) : (i,k),(k,j),(j,i) \in E'\}$. We now show that in each iteration it is possible to choose a pivot vertex k such that

$$\frac{\sum_{(j,i)\in T_k(E')} w_{ji}}{\sum_{(j,i)\in T_k(E')} c_{ji}} \leq 2\,,$$

which implies that the cost of τ is at most twice the cost of the optimal solution of LP. We consider if $\sum_{k\in V}\sum_{(j,i)\in T_k(E')} w_{ji} \leq 2\cdot\sum_{k\in V}\sum_{(j,i)\in T_k(E')} c_{ji}$ holds. If the desired ratio between the cost of τ and the cost of the optimal solution holds for the sum over all pivot elements, then, in particular, there is a pivot with the desired ratio. We observe that each edge contained in a triangle in E' is a critical edge for exactly one possible pivot k. Thus, instead of summing over all pivot elements and over all their critical edges, we may sum over all edges in all triangles in E'. Formally, let T be the set of triangles $\{(i,k),(k,j),(j,i)\} \subseteq E'$, and for a triangle $t \in T$, let $w(t) = \sum_{e \in t} w_e$ and let $c(t) = \sum_{e \in t} c_e$. Then

$$\sum_{k \in V} \sum_{(j,i) \in T_k(E')} w_{ji} = \sum_{t \in T} \sum_{(j,i) \in t} w_{ji} = \sum_{t \in T} w(t)\,,$$
$$\sum_{k \in V} \sum_{(j,i) \in T_k(E')} c_{ji} = \sum_{t \in T} \sum_{(j,i) \in t} c_{ji} = \sum_{t \in T} c(t)\,.$$

We will show that for any $t \in T$, $w(t) \le 2c(t)$, which ensures the existence of an eligible pivot element and completes the proof.

Let $t = \{e_1, e_2, e_3\}$. For $e = (i,j)$, let $w_e = w_{ij}$, $x_e = x_{ij}$, $\overline{w_e} = w_{ji}$, and $\overline{x_e} = x_{ji}$. Thus we have to show

$$w(t) = w_{e_1} + w_{e_2} + w_{e_3} \le$$
$$\le 2\left(\underbrace{x_{e_1}\overline{w_{e_1}} + \overline{x_{e_1}}w_{e_1}}_{c_{e_1}} + \underbrace{x_{e_2}\overline{w_{e_2}} + \overline{x_{e_2}}w_{e_2}}_{c_{e_2}} + \underbrace{x_{e_3}\overline{w_{e_3}} + \overline{x_{e_3}}w_{e_3}}_{c_{e_3}}\right) = 2c(t)\,.$$

t consists solely of σ-edges, which do not form a cycle. Thus the initial tournament $G = (V, E)$ contains at least one and at most two of e_1, e_2, and e_3. W. l. o. g. suppose $e_1 \notin E$ and $e_3 \in E$ and thus $w_{e_1} = 0$ and $w_{e_3} = 1$. We derive the simplified inequality

$$w(t) = w_{e_2} + 1 \le 2\left(x_{e_1} + x_{e_2}\overline{w_{e_2}} + \overline{x_{e_2}}w_{e_2} + \overline{x_{e_3}}\right) = 2c(t)\,.$$

Now we distinguish the two cases whether or not $e_2 \in E$.
Case 1: First $e_2 \in E$, and therefore $w_{e_2} = 1$, yields

$$w(t) = 2 \le 2\left(x_{e_1} + \overline{x_{e_2}} + \overline{x_{e_3}}\right) = 2c(t)\,.$$

We observe that $x_{e_1} \ge \frac{1}{2}$ as $e_1 \in E'$. Furthermore $\overline{x_{e_2}} + \overline{x_{e_3}} \ge \frac{1}{2}$, due to condition (1) of LP and to $\overline{x_{e_1}} \le \frac{1}{2}$. Thus the inequality holds.
Case 2: Secondly $e_2 \notin E$, and therefore $w_{e_2} = 0$, yields

$$w(t) = 1 \le 2\left(x_{e_1} + x_{e_2} + \overline{x_{e_3}}\right) = 2c(t)\,.$$

Observe that $x_{e_1} \ge \frac{1}{2}$. Thus the inequality holds. □

4 Rank Aggregation Problems

In this section we address the rank aggregation problems for partial orders under the nearest neighbor and the Hausdorff distances.

Theorem 5. *The rank aggregation problem for a partial and a total order under the nearest neighbor Kendall tau distance is* **NP***-complete.*

Proof. We show the **NP**-completeness by a reduction from the distance problem under the nearest neighbor distance of a partial order and a total order (Theorem 2), in which we ask if $K_{NN}(\kappa, \sigma) \leq k$ for a partial order κ and a total order σ on a domain $\mathcal{D}$, and an integer k. We reduce to an instance of the rank aggregation problem by simply taking κ and σ as voters, keep k unchanged, and ask if there exists a total order τ^* on $\mathcal{D}$ with $K_{NN}(\kappa, \tau^*) + K_{NN}(\sigma, \tau^*) \leq k$.

First suppose $K_{NN}(\kappa, \sigma) \leq k$. We now set $\tau^* = \sigma$ and have $K_{NN}(\kappa, \tau^*) + K_{NN}(\sigma, \tau^*) = K_{NN}(\kappa, \sigma) + K_{NN}(\sigma, \sigma) \leq k$.

Conversely, suppose there exists a total order τ with $K_{NN}(\kappa, \tau) + K_{NN}(\sigma, \tau) \leq k$. Then let τ' be a total order having $K_{NN}(\kappa, \tau') + K_{NN}(\sigma, \tau') \leq k$ and minimizing $K(\tau', \sigma)$. If $\tau' = \sigma$, we are clearly done, so suppose for contradiction $\tau' \neq \sigma$. Then there are $x, y \in \mathcal{D}$ with $\tau'(y) = \tau'(x) + 1$ and $\sigma(y) < \sigma(x)$, as shown, e. g., in [15]. Now derive τ^* from τ' by switching x and y. Then $K_{NN}(\tau^*, \sigma) = K(\tau^*, \sigma) = K(\tau', \sigma) - 1$. Additionally, $K(\kappa', \tau^*) \leq K(\kappa', \tau') + 1$ for each $\kappa' \in \mathrm{Ext}(\kappa)$, thus $K_{NN}(\kappa, \tau^*) \leq K_{NN}(\kappa, \tau') + 1$ and $K_{NN}(\kappa, \tau^*) + K_{NN}(\sigma, \tau^*) \leq k$. As $K(\sigma, \tau^*) < K(\sigma, \tau')$, we obtain a contradiction.

The above reduction clearly runs in polynomial time. The problem is solvable by an **NP** machine even for an unbounded number of voters $\kappa_1, \ldots, \kappa_r$ as we can guess a solution τ^* and total orders $\kappa'_1, \ldots, \kappa'_r$ and then verify that $\kappa'_1 \in \mathrm{Ext}(\kappa_1), \ldots, \kappa'_r \in \mathrm{Ext}(\kappa_r)$ and $\sum_{i=1}^{r} K(\kappa'_i, \tau^*) \leq k$, which implies $\sum_{i=1}^{r} K_{NN}(\kappa_i, \tau^*) \leq k$.

From the rank aggregation problem with two voters we can immediately reduce to the corresponding problem with any finite number of voters by adding additional voters π represented by a bucket order, which consists of one bucket containing all candidates. These voters then have $K_{NN}(\pi, \tau^*) = 0$ for any τ^* and thus do not affect the solution. $\square$

Theorem 6. *The rank aggregation problem for a partial and a total order under the Hausdorff Kendall tau distance is* **coNP***-hard. The problem is in* $\mathbf{\Sigma_2^P}$*, i. e., solvable by an* **NP** *machine which has access to an* **NP** *oracle [26, 28].*

Proof. We show the **coNP**-hardness by a reduction from the distance problem under the Hausdorff distance of a partial order and a total order (Theorem 3). The reduction is completely analogous to the one in the proof of Theorem 5.

The problem is solvable by an **NP** machine which has access to an **NP** oracle, even for an unbounded number of voters $\kappa_1, \ldots, \kappa_r$ as we can guess a solution τ^* and integers $k_1, \ldots, k_r$ with $\sum_{i=1}^{r} k_i \leq k$ and then use the oracle to verify that $K_H(\kappa_1, \tau^*) \leq k_1, \ldots$, and $K_H(\kappa_r, \tau^*) \leq k_r$.

From the rank aggregation problem with two voters we again reduce to the corresponding problem with any finite number of voters by adding additional voters π represented by a bucket order, which consists of one bucket containing all candidates. These voters have $K_H(\pi, \tau^*) = \binom{|\mathcal{D}|}{2}$ for any τ^* and thus do not affect the solution. $\square$

As the rank aggregation problem for partial orders under the Hausdorff distance is **coNP**-hard for two or more voters (Theorem 6) and **NP**-hard for four or more voters ([6, 14]), we obtain:

Corollary 2. *The rank aggregation problem for partial orders under the Hausdorff Kendall tau distance is not in* **NP** *or* **coNP** *unless* **NP** = **coNP**.

5 Conclusion and Open Problems

In this work we have studied the nearest neighbor and Hausdorff versions of the Kendall tau distance for bucket and partial orders and established efficient computations and hardness results. The approximability of the distance problem for two partial orders and of the rank aggregation problem under the nearest neighbor Kendall tau distance are on our agenda of open problems.

In a companion paper [8] we have investigated related problems for the Spearman footrule distance, which takes the L_1-norm on total orders. The Kendall tau and Spearman footrule distance show differences on related problems. Also interval orders with intervals for each candidate and irrational voters, which may contain cycles of strict preferences, shall be considered in our future work.

An interesting problem is the rank aggregation problem with only one voter whose preferences are specified by a partial order. On the one hand it is solvable in linear time for the nearest neighbor Kendall tau distance since any total extension of the given partial order is a consensus ranking and an extension can be computed by topological sorting. On the other hand it remains open for the Hausdorff Kendall tau distance. Here we must compute the center of all extensions of the partial order.

A challenging problem is the precise placement of the rank aggregation problem of many partial orders in the polynomial hierarchy. The problem seems harder than winner and election problems [17], which are complete for the class $\mathbf{\Theta_2^P}$. It has the flavor of a $\mathbf{\Sigma_2^P}$-complete problem as its structure ("Does there exist a consensus ranking, such that for all extensions ...?") resembles typical $\mathbf{\Sigma_2^P}$-complete problems (see [24] for an overview). We therefore conjecture:

Conjecture 1. The rank aggregation problem for partial orders under the Hausdorff distance is $\mathbf{\Sigma_2^P}$-complete.

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. N. Betzler, M. R. Fellows, J. Guo, R. Niedermeier, and F. A. Rosamond. Fixed-parameter algorithms for Kemeny rankings. *Theoretical Computer Science*, 410:4554–4570, 2009.
6. T. Biedl, F. J. Brandenburg, and X. Deng. On the complexity of crossings in permutations. *Discrete Mathematics*, 309:1813–1823, 2009.
7. J. C. Borda. Mémoire aux les élections au scrutin., 1781.
8. F. J. Brandenburg, A. Gleißner, and A. Hofmeier. The nearest neighbor Spearman footrule distance for bucket, interval, and partial orders. Technical Report MIP-1101, Fakultät für Informatik und Mathematik, Universität Passau, 2011. http://www.fim.uni-passau.de/wissenschaftler/forschungsberichte/mip-1101.html.
9. S. R. Buss and L. Hay. On truth-table reducibility to SAT. *Information and Computation*, 91:86–102, 1991.
10. I. Caragiannis, J. A. Covey, M. Feldman, C. M. Homan, C. Kaklamanis, N. Karanikolas, A. D. Procaccia, and J. S. Rosenschein. On the approximability of Dodgson and Young elections. In *Proceedings of the 20th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1058–1067. SIAM, 2009.
11. W. W. Cohen, R. E. Schapire, and Y. Singer. Learning to order things. *Journal of Artificial Intelligence Research (JAIR)*, 10:243–270, 1999.
12. M.-J. Condorcet. Éssai sur l'application de l'analyse à la probalité des décisions rendues à la pluralité des voix, 1785.
13. D. E. Critchlow. *Metric methods for analyzing partially ranked data*. Number 34 in Lecture notes in statistics. Springer, Berlin, 1985.
14. 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.
15. R. Fagin, R. Kumar, M. Mahdian, D. Sivakumar, and E. Vee. Comparing partial rankings. *SIAM Journal on Discrete Mathematics*, 20:628–648, 2006.
16. 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.
17. E. Hemaspaandra, H. Spakowski, and J. Vogel. The complexity of Kemeny elections. *Theoretical Computer Science*, 349:382–391, 2005.
18. G. Lebanon and J. D. Lafferty. Cranking: Combining rankings using conditional probability models on permutations. In *Proceedings of the 19th International Conference on Machine Learning (ICML)*, pages 363–370. Morgan Kaufmann, 2002.
19. R. Lullus. Artifitium electionis personarum, 1283.
20. R. Lullus. De arte eleccionis, 1299.
21. 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.
22. X. Muñoz, W. Unger, and I. Vrto. One sided crossing minimization is NP-hard for sparse graphs. In *Proceedings of the 9th International Symposium on Graph Drawing*, volume 2265 of *Lecture Notes in Computer Science*, pages 115–123. Springer, 2002.
23. 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.
24. M. Schaefer and C. Umans. Completeness in the polynomial-time hierarchy: A compendium. *SIGACT News*, 33:32–49, September 2002.
25. J. Sese and S. Morishita. Rank aggregation method for biological databases. *Genome Informatics*, 12:506–507, 2001.

26. L. J. Stockmeyer. The polynomial-time hierarchy. *Theoretical Computer Science*, 3:1–22, 1976.
27. A. van Zuylen and D. P. Williamson. Deterministic pivoting algorithms for constrained ranking and clustering problems. *Mathematics of Operations Research*, 34:594–620, 2009.
28. K. W. Wagner. Bounded query classes. *SIAM Journal on Computing*, 19(5):833–846, 1990.
29. 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.
30. 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.