A Novel Distance-Based Approach to Constrained Rank Aggregation

Farzad Farnoud (Hassanzadeh), Olgica Milenkovic, and Behrouz Touri

Department of Electrical and Computer Engineering, University of Illinois, Urbana-Champaign

E-mail: {hassanz1,milenkov,touri1}@illinois.edu

Abstract

We consider a classical problem in choice theory – vote aggregation – using novel distance measures between permutations that arise in several practical applications. The distance measures are derived through an axiomatic approach, taking into account various issues arising in voting with side constraints. The side constraints of interest include non-uniform relevance of the top and the bottom of rankings (or equivalently, eliminating negative outliers in votes) and similarities between candidates (or equivalently, introducing diversity in the voting process). The proposed distance functions may be seen as *weighted versions* of the Kendall τ distance and *weighted versions* of the Cayley distance. In addition to proposing the distance measures and providing the theoretical underpinnings for their applications, we also consider algorithmic aspects associated with distance-based aggregation processes. We focus on two methods. One method is based on approximating weighted distance measures by a generalized version of Spearman's footrule distance, and it has provable constant approximation guarantees. The second class of algorithms is based on a non-uniform Markov chain method inspired by PageRank, for which currently only heuristic guarantees are known. We illustrate the performance of the proposed algorithms for a number of distance measures for which the optimal solution may be easily computed.

I. Introduction

Rank aggregation, sometimes referred to as *ordinal data fusion*, is a classical problem frequently encountered in the social sciences, web search and Internet service studies, expert opinion analysis, and economics [2], [8], [17], [20], [24], [25]. Rank aggregation plays a special role in information retrieval based on different search models, in cases when users initiate several queries for the information of interest to them, or in situations when one has to combine various sources of evidence or use different document surrogates [11].

The problem can be succinctly described as follows: a set of "voters" or "experts" is presented with a set of candidates (objects, individuals, movies, etc.). Each voter's task is to produce a *ranking*, that is, an arrangement of the candidates in which the candidates are *ranked* from the most preferred to the least preferred. The voters' rankings are then passed to an aggregator. The aggregator outputs a single ranking, termed the aggregate ranking, to be used as a representative of all voters.

Rank aggregation for votes including two candidates reduces to a simple majority count. The situation becomes significantly more complex when three or more candidates are considered. Two of the most obvious extensions of vote aggregation for two candidates to the case of more than two candidates are the majority rule and the Condorcet method (pairwise majority count). In the first case, one reduces the problem to counting how many times each candidate ended up at the top of the list. This candidate is declared the winner, and removed from all rankings. The same process is then performed to identify the second, etc., candidate in the list. In the second case, one aims at identifying the majority winner of pairwise competitions. Unfortunately, both methods are plagued by a number of issues that have cast doubt on the plausibility of fair vote aggregation. Examples include the famous Condorcet paradox [5], where pairwise comparisons may lead to intransitive results (i.e., for example, a may be preferred to b, b to c, and c to a).

To mitigate such problems, two other important categories of rank aggregation methods were studied in the past. These include *score-based* methods and *distance-based* methods. In score-based methods, the first variant of which was proposed by Borda [4], each candidate is assigned a score based on its position in each of the votes (rankings). The candidates are then ranked based on their total score. One argument in support of using Borda's count method is that it ranks highly those candidates supported at least to a certain extent by almost all voters, rather than candidates who are ranked highly only by the simple majority of voters. In distance-based methods [17], the aggregate is the deemed to be the ranking "closest" to the set of votes, or at the smallest cumulative distance from the votes, where closeness of two rankings is measured via some adequately chosen distance function. This approach can be thought of as finding the center of mass of the rankings, or the median – centroid – of the rankings, with the rankings representing point masses in a metric space. Well-known distance measures for rank aggregation include the Kendall τ, the Cayley distance, and Spearman's Footrule [7].

Clearly, the most important aspect of distance-based rank aggregation is to choose an appropriate distance function. One may argue that almost all problems arising in connection with the majority method or score based approaches directly translate into problems concerning the chosen distance measures. To address this issue, Kemeny [17], [18] presented a set of intuitively

justifiable axioms that a distance measure must satisfy to be deemed suitable for aggregation purposes, and showed that only one distance measure satisfies the axioms – namely, the Kendall τ distance. The Kendall τ distance between two rankings is the *smallest* number of swaps of adjacent elements that transforms one ranking into the other. For example, the Kendall τ distance between the rankings $(1, 3, 4, 2)$ and $(1, 2, 3, 4)$ is two; we may first swap 2 and 4 and then 2 and 3 to transform $(1, 3, 4, 2)$ to $(1, 2, 3, 4)$. Besides its use in social choice and computer science theory, the Kendall τ distance has also received significant attention in the coding theory literature, due to its applications in modulation coding for flash memories [16], [19], [21].

Unfortunately, the Kendall τ is not a suitable distance measure for aggregation problems involving various electoral and Internet search engine constraints. Two such important constraints include differentiating the significance of the top versus the bottom of a ranking and differentiating candidates based on their "similarity".

In the first example, consider the following scenario. One may view the process of forming the aggregate ranking as one of "tweaking" a starting ranking so as to make it as close as possible to *all* given voters' rankings. The effect of changing the ordering of candidates at the top or at the bottom is in principle the same – i.e., if switching the top two elements in the aggregate reduces the total distance from the votes by the same amount as switching the bottom two elements, then both options are equally valid to be used. But in many applications, changes at, or near to, the top of rankings should not affect the distance between rankings to same extent as changes at, or near to, the bottom of rankings. In other words, one should penalize making changes at the top of the list more than making changes at the bottom of the list, given that low ranked items are usually not very relevant. So far, only a handful of results are known for rank aggregation distances that address the problem of positional relevance, i.e., the significance of the top versus the bottom of rankings. One approach was described in [20], where the proposed distances were based on *heuristic arguments* only. These approaches do not have axiomatic underpinnings, and efficient aggregation algorithms to accompany them are not known.

In the second example, consider a voting process were candidates should be ranked both based on merit and on a diversity criteria – for example, not having more than two of the top ten candidates working in information theory. One may argue that in this case, using the Kendall τ distance for aggregation and reshuffling some candidates in order to satisfy the constraints, suffices to solve the problem. For example, one may move all except the two highest-ranked information theorists below position ten and leave the ranking unchanged otherwise. It is clear that this procedure may not be viewed as fair, since ranks of *all* candidates were affected by the rankings of information theorists in the first place. Alternatively, one may reduce the search space only to rankings that satisfy the constraints, but this approach is computationally highly challenging.

Henceforth, we focus our attention on distance based aggregation methods catering to constraints of the form described above. The goal of our work is to provide an axiomatic underpinning for novel distance measures between rankings that take into account predetermined top-bottom and similarity/diversity constraints. In addition to their applications in computer science and social choice theory, these distance measures may be used in a variety of applications, ranging from bioinformatics to network analysis [12].

Motivation – Top vs. Bottom

Consider the ranking π of the "World's 10 best cities to live in", according to a report composed by the Economist Intelligence Unit [10]:

$$\pi = (\text{Melbourne, Vienna, Vancouver, Toronto, Calgary,}$$
$$\text{Adelaide, Sydney, Helsinki, Perth, Auckland})$$

Now consider two other rankings that both differ from π by one swap of adjacent entries:

$$\pi' = (\text{Melbourne, Vienna, Vancouver, Calgary, Toronto,}$$
$$\text{Adelaide, Sydney, Helsinki, Perth, Auckland}),$$
$$\pi'' = (\text{Vienna, Melbourne, Vancouver, Toronto, Calgary,}$$
$$\text{Adelaide, Sydney, Helsinki, Perth, Auckland}).$$

The astute reader probably immediately noticed that the top candidate was changed in π'', but otherwise took some time to realize where the adjacent swap appeared in π'. This is a consequence of the well-known fact that humans pay more attention to the top of the list rather than any other location in the ranking, and hence notice changes in higher positions easier[1]. Note that the Kendall τ distance between π and π' and between π and π'' is one, but it would appear reasonable to assume that the distance between π and π'' be larger than that between π and π', as the corresponding swap occurred in a *more significant* (higher ranked) position in the list.

---

[1]Note that one may argue that people are equally drawn to explore the highest and lowest ranked items in a list. For example, if about a hundred cities were ranked, it would be reasonable to assume that readers would be more interested in knowing the best and worst ten cities, rather than the cities occupying positions 41 to 60. These positional differences may also be addressed within the framework proposed in the paper.
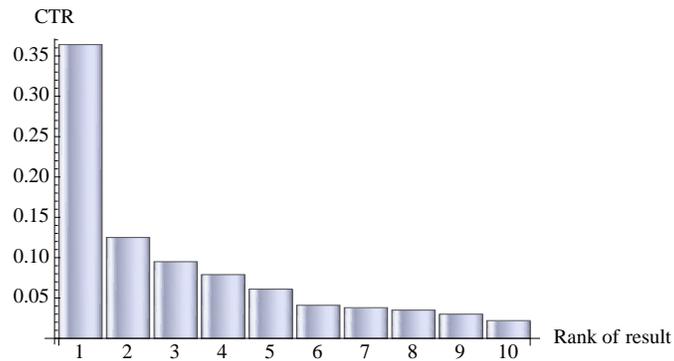
Figure 1: Clickthrough rates (CTRs) of webpages appearing on the first page of Google search.

The second example corresponds to the well-studied notion of *Clickthrough rates* (CTRs) of webpages in search engine results pages (SERPs). The CTR is used to assess the popularity of a webpage or the success rate of an online ad. It may be roughly defined as the number of times a link is clicked on divided by the total number of times that it appeared. A recent study by Optify Inc. [22] showed that the difference between the average CTR of the first (highest-ranked) result and the average CTR of the second (runner-up) result is very large, and much larger than the corresponding difference between the average CTRs of the lower ranked items (See Figure 1). Hence, in terms of directing search engine traffic, swapping higher-ranked adjacent pairs of search results has a larger effect on the performance of Internet services than swapping lower-ranked search results.

The aforementioned findings should be considered when forming an aggregate ranking of webpages. For example, in studies of CTRs, one is often faced with questions regarding traffic flow from search engines to webpages. One may think of a set of keywords, each producing a different ranking of possible webpages, with the aggregate representing the median ranking based on different sets of keywords. Based on Figure 1, if the ranking of a webpage is in the bottom half, its exact position is not as relevant as when it is ranked in the top half. Furthermore, a webpage appearing roughly half of the time at the top and roughly half of the time at the bottom will generate more incoming traffic than a webpage with persistent average ranking.

Throughout the paper, we refer to the above-mentioned problem as the "top-vs-bottom" problem. Besides the importance in emphasizing the relevance of the top of the list, distance measures that penalize perturbations at the top of the list more than perturbations at the bottom of the list have another important application in practice – to eliminate negative outliers. As will be shown in subsequent sections, top-vs-bottom distance measures allow candidates to be highly ranked in the aggregate even though they have a certain (small) number of highly negative ratings. The policy of eliminating outliers before rating items or individuals is a well-known one, but has not been considered in the social choice literature in the context of distance-based rank aggregation.

Motivation – Similarity of Candidates

In many vote aggregation problems, the identity of the candidates may not be known. On the other hand, many other applications require that the identity of the candidates be revealed. In this case, candidates are frequently partitioned in terms of some similarity criteria – for example, area of expertise, gender, working hour schedule etc. Hence, pairs of candidates may have different degrees of similarity and swapping candidates that are similar should be penalized less than swapping candidates that are not similar according to the given ranking criteria. For example, in a faculty search ranking one may want to have at least one but not more than two physicists ranked among the top 10 candidates, or at least two women among the top 5 candidates.

Pertaining to the Economist Intelligence Unit ranking, one may also consider the identity of the elements that are swapped, and not only their position. In this case, it may be observed that the swap in π'' involves cities on two different continents, which may shift the general opinion about the cities' countries of origin. On the other hand, the two cities swapped in π' are both in Canada, so that the swap is not likely to change the perception of quality of living in that country. This points to the need for distance measures that take into account similarities and dissimilarities among candidates.

Distance measures capable of integrating these criteria directly into the aggregation process are not know in the literature. A class of distance measures introduced by the authors in [12], termed weighted transposition distance, is suitable for this task as it can take into account similarities and dissimilarities of candidates. The weighted transposition distance can be viewed as a generalization of both the previously described Kendall τ and the so called *Cayley distance* between permutations. The Cayley distance between two rankings is the *smallest* number of (not necessarily adjacent) swaps required to transform one ranking into the other. For example, the Cayley distance between the permutations $(1, 2, 3, 4)$ and $(1, 4, 3, 2)$ is one, since the former can be transformed into the latter by swapping the elements 2 and 4. Note that while the Cayley distance allows for arbitrary

swaps, the Kendall τ distance allows for swaps of adjacent elements only. It is straightforward to see that every statements made about swapping elements i and j may be converted to statements made about swapping elements at positions i and j by using the inverse of the ranking/permutation. Similarity of items may be captured by assigning costs or weights to swaps, and choosing the transposition weights so that swapping dissimilar items induces a higher weight/distance compared to swapping similar items. This approach is the topic of the next two sections.

To address the top-vs-bottom and similarity issues, we axiomatically describe a class of distance functions by assigning different weights to different adjacent and non-adjacent swaps, termed the weighted Kendall and weighted transposition distance measures, respectively. Furthermore, we show that the proposed distance functions can be computed in polynomial time in some special cases and provide a polynomial-time 2-approximation algorithm for the general case. The results we present also pertain to algorithmic aspects of rank aggregation [2], [17], [24], [25].

In this setting, we describe the performance of an algorithm for rank aggregation based on a generalization of Spearman's footrule distance and solving a minimum weight matching problem (this algorithm is inspired by a procedure described in [8]) and a combination of the matching algorithm with local descent methods. Furthermore, we describe an algorithm reminiscent of PageRank [8], where the "hyperlink probabilities" are chosen according to swapping likelihoods (weights).

The remainder of the paper is organized as follows. An overview of relevant concepts, definitions, and terminology is presented in Section II. Weighted Kendall distance measures, as well their axiomatic definitions, are presented in Section III. Section IV is devoted to the weighted transposition distance and its computational aspects. Novel rank aggregation algorithms for the weighted Kendall and weighted transposition distances are presented in Section V.

II. Preliminaries

Formally, a ranking is a list of candidates arranged in order of preference, with the first candidate being the most preferred and the last candidate being the least preferred one.

Consider the set of all possible rankings of a set of n candidates. Via an arbitrary, but fixed, injective mapping from the set of candidates to $\{1, 2, \cdots, n\} = [n]$, each ranking may be represented as a permutation. The mapping is often implicit and we usually equate rankings of n candidates with permutations in $\mathbb{S}_n$, where $\mathbb{S}_n$ denotes the symmetric group of order n. This is equivalent to assuming that the set of candidates is the set $[n]$. For notational convenience, we use Greek lower-case letters for permutations, and explicitly write permutations as ordered sets $\sigma = (\sigma(1), \ldots, \sigma(n))$.

Let e denote the identity permutation $(1, 2, \cdots, n)$. For two permutations $\pi, \sigma \in \mathbb{S}_n$, the product $\mu = \pi\sigma$ is defined via the identity $\mu(i) = \pi(\sigma(i)), i = 1, 2, \cdots n$.

Definition 1. A *transposition* $\tau = (a\,b)$, for $a, b \in [n]$ and $a \neq b$, is a permutation that swaps a and b and keeps all other elements of e fixed. That is,

$$\tau(i) = \begin{cases} b, & i = a, \\ a, & i = b, \\ i, & \text{else} \end{cases}$$

If $|a - b| = 1$, the transposition is referred to as an *adjacent transposition*.

Note that for $\pi \in \mathbb{S}_n$, $\pi(a\,b)$ is obtained from π by swapping elements in positions a and b, and $(a\,b)\pi$ is obtained by swapping a and b in π. For example, $(3, 1, 4, 2)(2\,3) = (3, 4, 1, 2)$ and $(2\,3)(3, 1, 4, 2) = (2, 1, 4, 3)$.

For our future analysis, we define the set

$$A(\pi, \sigma) = \big\{\tau = (\tau_1, \cdots, \tau_{|\tau|}) :$$
$$\sigma = \pi\tau_1 \cdots \tau_{|\tau|}, \tau_i = (a_i\, a_i + 1), i \in [|\tau|]\big\}$$

i.e., the set of all ordered sequences of adjacent transpositions that *transform* π into σ. The fact that $A(\pi, \sigma)$ is non-empty, for any $\pi, \sigma \in \mathbb{S}_n$, is obvious. Using $A(\pi, \sigma)$, the Kendall τ distance between two permutations π and σ, denoted by $K(\pi, \sigma)$, may be written as

$$K(\pi, \sigma) = \min_{\tau \in A(\pi, \sigma)} |\tau|.$$

For a ranking $\pi \in \mathbb{S}_n$ and $a, b \in [n]$, π is said to *rank a before b* or *higher than b* if $\pi^{-1}(a) < \pi^{-1}(b)$. We denote this relationship as $a <_\pi b$. Two rankings π and σ *agree* on the relative order of a pair $\{a, b\}$ of elements if both rank a before b or both rank b before a. Furthermore, the two rankings π and σ *disagree* on the relative order of a pair $\{a, b\}$ if one ranks a before b and the other ranks b before a. For example, consider $\pi = (1, 2, 3, 4)$ and $\sigma = (4, 2, 1, 3)$. We have that $4 <_\sigma 1$ and that π and σ agree on $\{2, 3\}$ but disagree on $\{1, 2\}$.

Given a distance function d over the permutations in $\mathbb{S}_n$ and a set $\Sigma = \{\sigma_1, \cdots, \sigma_m\}$ of m votes (rankings), the distance-based aggregation Σ problem can be stated as follows: find the ranking π^* that minimizes the cumulative distance from Σ, i.e.,

$$\pi^* = \arg\min_{\pi \in \mathbb{S}_n} \sum_{i=1}^m \mathsf{d}(\pi, \sigma_i). \tag{1}$$

In words, the goal is to find a ranking π that represents the median of the set of permutations Σ. The choice of the distance function d is an important aspect of distance-based rank aggregation and the focus of the paper.

In [17], Kemeny presented a set of axioms that a distance function for rank aggregation should satisfy and proved that the only distance that satisfies the axioms is the Kendall τ. A critical concept in Kemeny's axioms is the idea of "betweenness," defined below.

Definition 2. A ranking ω is said to be *between* two rankings π and σ, denoted by π–ω–σ, if for each pair of elements $\{a, b\}$, ω either agrees with π or σ or both. The rankings $\pi_1, \pi_2, \cdots, \pi_s$ are said to be *on a line*, denoted by π_1–π_2–$\cdots$–π_s, if for every i, j, and k for which $1 \leq i < j < k \leq s$, we have π_i–π_j–π_k.

In Kemeny's work, rankings are allowed to have ties. The basis of our subsequent analysis is the same set of axioms, listed below. However, our focus is on ranking without ties, in other words, permutations.

Axioms I
1) d is a metric.
2) d is left-invariant.
3) For any π, σ, and ω, $\mathsf{d}(\pi, \sigma) = \mathsf{d}(\pi, \omega) + \mathsf{d}(\omega, \sigma)$ if and only if ω is between π and σ.
4) The smallest positive distance is one.

Axiom 2 states that relabeling of objects should not change the distance between permutations. In other words, $\mathsf{d}(\sigma\pi, \sigma\omega) = \mathsf{d}(\pi, \omega)$, for any $\pi, \sigma, \omega \in \mathbb{S}_n$. Axiom 3 may be viewed through a geometric lens: the triangle inequality has to be satisfied with equality for all points that lie on a line between π and σ. Axiom 4 is only used for normalization purposes.

Kemeny's original exposition included a fifth axiom which we state for completeness: If two rankings π and σ agree except for a segment of k elements, the position of the segment does not affect the distance between the rankings. Here, a segment represents a set of objects that are ranked consecutively – i.e., a substring of the permutation. As an example, this axiom implies that

$$\mathsf{d}((1, 2, 3, \underbrace{4, 5, 6}), (1, 2, 3, \underbrace{6, 5, 4})) =$$
$$\mathsf{d}((1, \underbrace{4, 5, 6}, 2, 3), (1, \underbrace{6, 5, 4}, 2, 3))$$

where the segment is underscored by braces. This axiom clearly enforces a property that *is not desirable* for metrics designed to address the top-vs-bottom issue: changing the position of the segment in two permutations does not alter their mutual distance. One may hence believe that removing this axiom (as was done in Axioms I) will lead to distance measures capable of handling the top-vs-bottom problem. But as we show below, for rankings without ties, omitting this axiom does not change the outcome of Kemeny's analysis. In other words, the axiom is redundant. This is a rather surprising fact, and we conjecture that the same is true of rankings with ties.

In the remainder of this section, we demonstrate the redundancy of Kemeny's fifth axiom and use our novel proof method to identify how to change the axioms in Axioms I in order to arrive at distance measures that cater to the need of top-vs-bottom and similarity problems. For reasons that will become clear in the next section, we refer to distance measures resulting from such axioms as *weighted distances*.

The main result of this section is Theorem 8, stating that the unique distance satisfying Axioms I is the Kendall τ distance. The theorem is proved with the help of Lemmas 3, 4, 5, 6 and 7.

Lemma 3. *For any distance measure* d *that satisfies Axioms I, and for any sequence of permutations* $\pi_1, \pi_2, \cdots, \pi_s$ *such that* π_1–π_2–$\cdots$–π_s, *one has*

$$\mathsf{d}(\pi_1, \pi_s) = \sum_{k=1}^{s-1} \mathsf{d}(\pi_k, \pi_{k+1}).$$

Proof: The lemma follows from Axiom I.3 by induction. ∎

Lemma 4. *For any* d *that satisfies Axioms I and for* $i \in [n-1]$, *we have*

$$\mathsf{d}\left((i\,i+1), e\right) = \mathsf{d}\left((1\,2), e\right).$$

Proof: We first show that $\mathsf{d}\left((2\,3), e\right) = \mathsf{d}\left((1\,2), e\right)$. Repeating the same argument used for proving this special case gives $\mathsf{d}\left((i\,i+1), e\right) = \mathsf{d}\left((i-1\,i), e\right) = \cdots = \mathsf{d}\left((1\,2), e\right)$.

To show that $\mathsf{d}\left((2\,3), e\right) = \mathsf{d}\left((1\,2), e\right)$, we evaluate $\mathsf{d}(\pi, e)$ in two ways, where we choose $\pi = (3, 2, 1, 4, 5, \cdots, n)$.

On the one hand, note that π–ω–η–e, where $\omega = \pi(1\,2) = (2, 3, 1, 4, 5, \cdots, n)$ and $\eta = \omega(2\,3) = (2, 1, 3, 4, 5, \cdots, n)$. As a result,

$$\begin{aligned}
\mathsf{d}(\pi, e) &= \mathsf{d}(\pi, \omega) + \mathsf{d}(\omega, \eta) + \mathsf{d}(\eta, e) \\
&= \mathsf{d}(\omega^{-1}\pi, e) + \mathsf{d}(\eta^{-1}\omega, e) + \mathsf{d}(\eta, e) \\
&= \mathsf{d}((1\,2), e) + \mathsf{d}((2\,3), e) + \mathsf{d}((1\,2), e)
\end{aligned} \tag{2}$$

where the first equality follows from Lemma 3, while the second is a consequence of the left-invariance property of the distance measure.

On the other hand, note that π–α–β–e, where $\alpha = \pi(2\,3) = (3,1,2,4,5,\cdots,n)$ and $\beta = \alpha(1\,2) = (1,3,2,4,5,\cdots,n)$. For this case,

$$
\begin{aligned}
\mathsf{d}(\pi, e) &= \mathsf{d}(\pi, \alpha) + \mathsf{d}(\alpha, \beta) + \mathsf{d}(\beta, e) \\
&= \mathsf{d}(\alpha^{-1}\pi, e) + \mathsf{d}(\beta^{-1}\alpha, e) + \mathsf{d}(\beta, e) \\
&= \mathsf{d}((2\,3), e) + \mathsf{d}((1\,2), e) + \mathsf{d}((2\,3), e).
\end{aligned}
\tag{3}
$$

Equations (2) and (3) imply that $\mathsf{d}\left((2\,3), e\right) = \mathsf{d}\left((1\,2), e\right)$. ∎

Lemma 5. *For any* d *that satisfies Axioms I,* $\mathsf{d}(\gamma, e)$ *equals the minimum number of adjacent transpositions required to transform γ into e.*

Proof: Let

$$
\begin{aligned}
L(\pi, \sigma) = \{\tau = \left(\tau_1, \cdots, \tau_{|\tau|}\right) &\in A(\pi, \sigma) : \\
&\pi - \pi\tau_1 - \pi\tau_1\tau_2 - \cdots - \sigma\}
\end{aligned}
$$

be the subset of $A(\pi, \sigma)$ consisting of sequences of transpositions that transform π into σ by passing through a line. Let s be the minimum number of adjacent transpositions that transform γ into e. Furthermore, let $(\tau_1, \tau_2, \cdots, \tau_s) \in A(\gamma, e)$ and define $\gamma_i = \gamma\tau_1 \cdots \tau_i, i = 0, \cdots, s$, with $\gamma_0 = \gamma$ and $\gamma_s = e$.

First, we show $\gamma_0 - \gamma_1 - \cdots - \gamma_s$, that is,

$$
(\tau_1, \tau_2, \cdots, \tau_s) \in L(\gamma, e).
\tag{4}
$$

Suppose this were not the case. Then, there exist $i < j < k$ such that $\gamma_i, \gamma_j,$ and γ_k are not on a line, and thus, there exists a pair $\{r, s\}$ for which γ_j disagrees with both γ_i and γ_k. Hence, there exist two transpositions, $\tau_{i'}$ and $\tau_{j'}$, with $i < i' \leq j$ and $j < j' \leq k$ that swap r and s. We can in this case remove $\tau_{i'}$ and $\tau_{j'}$ from $(\tau_1, \cdots, \tau_s)$ to obtain $(\tau_1, \cdots, \tau_{i'-1}, \tau_{i'+1}, \cdots, \tau_{j'-1}, \tau_{j'+1}, \tau_s) \in A(\gamma, e)$ with length $s - 2$. This contradicts the optimality of the choice of s. Hence, $(\tau_1, \tau_2, \cdots, \tau_s) \in L(\gamma, e)$. Then Lemma 3 implies that

$$
\mathsf{d}(\gamma, e) = \sum_{i=1}^{s} \mathsf{d}(\tau_i, e).
\tag{5}
$$

Lemma 4 states that all adjacent transpositions have the same distance from the identity. Since transpositions $\tau_i, 1 \leq i \leq s$, in (5) are adjacent transpositions, $\mathsf{d}(\tau_i, e) = a$ for some $a > 0$ and thus $\mathsf{d}(\gamma, e) = sa$.

In (5), the minimum positive distance is obtained when $s = 1$. That is, the minimum positive distance from identity equals a and is obtained when γ is an adjacent transposition. Axiom I.4 states that the minimum positive distance is 1. By left-invariance, this axiom implies that the minimum positive distance of any permutation from the identity is 1. Hence, $a = 1$ and for any $\gamma \in \mathbb{S}_n$,

$$
\mathsf{d}(\gamma, e) = \sum_{i=1}^{s} \mathsf{d}(\tau_i, e) = sa = s.
$$

∎

Lemma 6. *For any* d *that satisfies Axioms I, and for $\pi, \sigma \in \mathbb{S}_n$, we have*

$$
\mathsf{d}(\pi, \sigma) = \min\{s : (\tau_1, \cdots, \tau_s) \in A(\pi, \sigma)\}.
$$

Proof: We have $(\tau_1, \cdots, \tau_s) \in A(\pi, \sigma)$ if and only if

$$
(\tau_1, \cdots, \tau_s) \in A(\sigma^{-1}\pi, e).
$$

Left-invariance of d implies that $\mathsf{d}(\pi, \sigma) = \mathsf{d}(\sigma^{-1}\pi, e)$. Hence,

$$
\begin{aligned}
\mathsf{d}(\pi, \sigma) &= \mathsf{d}(\sigma^{-1}\pi, e) \\
&= \min\{s : (\tau_1, \cdots, \tau_s) \in A(\sigma^{-1}\pi, e)\} \\
&= \min\{s : (\tau_1, \cdots, \tau_s) \in A(\pi, \sigma)\}
\end{aligned}
$$

where the second equality follows from Lemma 5. ∎

For $\pi, \sigma \in \mathbb{S}_n$, let

$$
I(\pi, \sigma) = \{\{i, j\} : i <_\pi j, j <_\sigma i\}
$$

be the set of pairs $\{i, j\}$ on which π and σ disagree. The number $|I(\pi, \sigma)|$ is usually referred to as the number of *inversions* between the two permutations.

The following lemma show that the Kendall τ distance between a permutation π and e equals the number of inversions in π. The result of the lemma is known, but a sketch of a proof is provided for completeness.

Lemma 7. *For $\pi, \sigma \in \mathbb{S}_n$,*

$$K(\pi, \sigma) = |I(\pi, \sigma)|.$$

Proof: Consider a sequence $\tau_1, \cdots, \tau_k$ of adjacent transpositions that transforms π into σ, i.e., $\sigma = \pi \tau_1 \cdots \tau_k$, with $k = K(\pi, \sigma)$. Let $\pi_j = \pi \tau_1 \cdots \tau_j$. Each τ_i decreases the number of inversions by at most one. Hence,

$$|I(\pi_j, \sigma)| \geq |I(\pi_{j-1}, \sigma)| - 1$$

and thus

$$0 = |I(\pi_k, \sigma)| \geq |I(\pi, \sigma)| - k.$$

Since $k = K(\pi, \sigma)$, we obtain

$$K(\pi, \sigma) \geq |I(\pi, \sigma)|.$$

On the other hand, it is easy to see that one can find $\tau_i, i \in [k]$ in such a way that each τ_i decreases the number of inversions by one. For example, Bubble Sort [3] is one such well-known algorithm for accomplishing this task. Hence,

$$K(\pi, \sigma) = |I(\pi, \sigma)|.$$

$\blacksquare$

Theorem 8. *The unique distance* d *that satisfies Axioms I is*

$$K(\pi, \sigma) = \min \{s : (\tau_1, \cdots, \tau_s) \in A(\pi, \sigma)\}.$$

Proof: We show below that K satisfies Axiom I.3, as proving that K satisfies the other axioms is straightforward. Uniqueness follows from Lemma 6.

To show that K satisfies Axiom I.3, we use Lemma 7 stating that

$$K(\pi, \sigma) = |I(\pi, \sigma)|.$$

Fix $\pi, \sigma \in \mathbb{S}_n$. For any $\omega \in \mathbb{S}_n$, it is clear that

$$I(\pi, \sigma) \subseteq I(\pi, \omega) \cup I(\omega, \sigma). \tag{6}$$

Suppose first that ω is not between π and σ. Then there exists a pair $\{a, b\}$ with $a <_\pi b$ and $a <_\sigma b$ but with $a >_\omega b$. Since $\{a, b\} \notin I(\pi, \sigma)$ but $\{a, b\} \in I(\pi, \omega) \cup I(\omega, \sigma)$, we find that

$$|I(\pi, \sigma)| < |I(\pi, \omega) \cup I(\omega, \sigma)|,$$

and thus

$$\begin{aligned}
K(\pi, \sigma) &= |I(\pi, \sigma)| \\
&< |I(\pi, \omega) \cup I(\omega, \sigma)| \\
&\leq |I(\pi, \omega)| + |I(\omega, \sigma)| \\
&= K(\pi, \omega) + K(\omega, \sigma).
\end{aligned}$$

Hence, if ω is not between π and σ, then

$$K(\pi, \sigma) \neq K(\pi, \omega) + K(\omega, \sigma).$$

Next, suppose ω is between π and σ. This immediately implies that $I(\pi, \omega) \subseteq I(\pi, \sigma)$ and $I(\omega, \sigma) \subseteq I(\pi, \sigma)$. These relations, along with (6) imply that

$$I(\pi, \omega) \cup I(\omega, \sigma) = I(\pi, \sigma). \tag{7}$$

We claim that $I(\pi, \omega) \cap I(\omega, \sigma) = \emptyset$. To see this, observe that if $\{a, b\} \in I(\pi, \omega) \cap I(\omega, \sigma)$, then the relative rankings of a and b are the same for π and σ and so, $\{a, b\} \notin I(\pi, \sigma)$. The last statement contradicts (7) and thus

$$I(\pi, \omega) \cap I(\omega, \sigma) = \emptyset. \tag{8}$$

From (7) and (8), we may write

$$\begin{aligned}
K(\pi, \sigma) &= |I(\pi, \sigma)| \\
&= |I(\pi, \omega) \cup I(\omega, \sigma)| \\
&= |I(\pi, \omega)| + |I(\omega, \sigma)| \\
&= \mathsf{d}(\pi, \omega) + \mathsf{d}(\omega, \sigma),
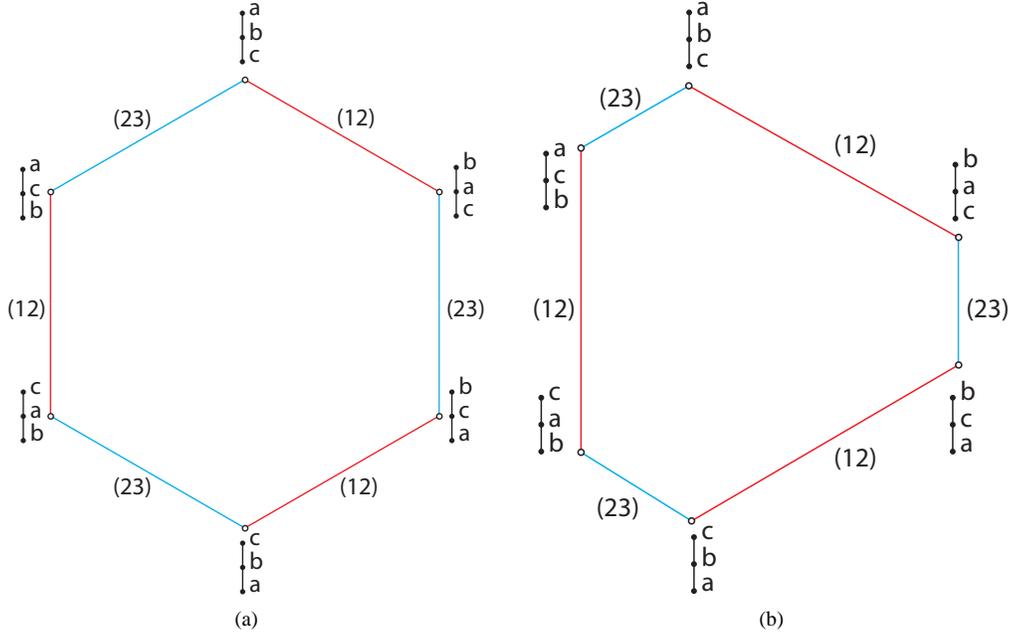\end{aligned}$$

Figure 2: The graphs for the Kendall τ distance (a), and weighted Kendall distance (b)

and this completes the proof of the fact that K satisfies Axiom I.3. $\blacksquare$

A distance d over $\mathbb{S}_n$ is called a *graphic distance* [6] if there exists a graph G with vertex set $\mathbb{S}_n$ such that for $\pi, \sigma \in \mathbb{S}_n$, $\mathsf{d}(\pi, \sigma)$ is equal to the length of the shortest path between π and σ in G. Note that this definition implies that the edge set of G is the set

$$\{(\alpha, \beta) : \alpha, \beta \in \mathbb{S}_n, \mathsf{d}(\alpha, \beta) = 1\}.$$

The Kendall τ distance is a graphic distance. To see the validity of this claim, take the corresponding graph to have vertices indexed by permutations, with an edge between each pair of permutations that differ by only one adjacent transposition.

In the next section, we introduce the weighted Kendall distance which may be viewed as the shortest path between permutations over a *weighted graph,* and show how this distance arises from modifying Kemeny's axioms.

III. THE WEIGHTED KENDALL DISTANCE

The proof of the uniqueness of the Kendall τ distance under Axioms I reveals an important insight: the Kendall τ distance arises due to the fact that adjacent transpositions have uniform costs, which is a consequence of the betweenness property described in one of the axioms. If one had a ranking problem in which weights of transpositions either depended on the identity of the elements involved in the transposition or their positions, the uniformity assumption would have to be changed. As we show below, a way to achieve this goal is to redefine the axioms in terms of the betweenness property.

Axioms II

1) d is a pseudo-metric, i.e. a generalized metric in which two distinct points may be at distance zero.
2) d is left-invariant.
3) For any π, σ disagreeing on more than one pair of elements, there exists *some* ω, distinct from π and σ and between them, such that $\mathsf{d}(\pi, \sigma) = \mathsf{d}(\pi, \omega) + \mathsf{d}(\omega, \sigma)$.

Axiom II.1 allows for the option that some transpositions are not penalized or counted, due to the side constraints of the voting process. Intuitively, Axiom II.3 states that there exists at least one point on some line between π and σ, for which the triangle inequality is an equality. In other words, there exists one "shortest line" between two permutations, and not all straight lines are required to be of the same length (see Figure 2 for an illustration).

Lemma 9. *For any distance* d *that satisfies Axioms II, and for distinct* π *and* σ, *we have*

$$\mathsf{d}(\pi, \sigma) = \min_{(\tau_1, \cdots, \tau_s) \in A(\pi, \sigma)} \sum_{i=1}^{s} \mathsf{d}(\tau_i, e).$$

Proof: The proof follows by induction on $K(\pi, \sigma)$, the Kendall τ distance between π and σ.

First, suppose that $K(\pi, \sigma) = 1$, i.e., π and σ disagree on one pair of adjacent elements. Then, we have $\sigma = \pi(a\, a+1)$ for some $a \in [n-1]$. For each $(\tau_1, \cdots, \tau_s) \in A(\pi, \sigma)$, there exists an index j such that $\tau_j = (a\, a+1)$ and thus

$$\sum_{i=1}^{s} \mathsf{d}(\tau_i, e) \geq \mathsf{d}(\tau_j, e) = \mathsf{d}((a\, a+1), e)$$

implying

$$\min_{(\tau_1, \cdots, \tau_s) \in A(\pi, \sigma)} \sum_{i=1}^{s} \mathsf{d}(\tau_i, e) \geq \mathsf{d}((a\, a+1), e). \tag{9}$$

On the other hand, since $((a\, a+1)) \in A(\pi, \sigma)$,

$$\min_{(\tau_1, \cdots, \tau_s) \in A(\pi, \sigma)} \sum_{i=1}^{s} \mathsf{d}(\tau_i, e) \leq \mathsf{d}((a\, a+1), e). \tag{10}$$

From (9) and (10),

$$\mathsf{d}(\pi, \sigma) = \min_{(\tau_1, \cdots, \tau_s) \in A(\pi, \sigma)} \sum_{i=1}^{s} \mathsf{d}(\tau_i, e) = \mathsf{d}((a\, a+1), e)$$

where the last equality follows from the left-invariance of d.

Next, suppose that $K(\pi, \sigma) > 1$, i.e., π and σ disagree on more than one pair of adjacent elements, and that for all $\mu, \eta \in \mathbb{S}_n$ with $K(\mu, \eta) < K(\pi, \sigma)$, the lemma holds. Then, there exists ω, distinct from π and σ and between them, such that

$$\mathsf{d}(\pi, \sigma) = \mathsf{d}(\pi, \omega) + \mathsf{d}(\omega, \sigma),$$
$$K(\pi, \omega) < K(\pi, \sigma),$$
$$K(\omega, \sigma) < K(\pi, \sigma).$$

By the induction hypothesis, there exist $(\nu_1, \cdots, \nu_k) \in A(\pi, \omega)$ and $(\nu_{k+1}, \cdots, \nu_s) \in A(\omega, \sigma)$, for some s and k, such that

$$\mathsf{d}(\pi, \omega) = \sum_{i=1}^{k} \mathsf{d}(\nu_i, e),$$

$$\mathsf{d}(\omega, \sigma) = \sum_{i=k+1}^{s} \mathsf{d}(\nu_i, e),$$

and thus

$$\mathsf{d}(\pi, \sigma) = \sum_{i=1}^{s} \mathsf{d}(\nu_i, e) \geq \min_{(\tau_1, \cdots, \tau_{s'}) \in A(\pi, \sigma)} \sum_{i=1}^{s'} \mathsf{d}(\tau_i, e),$$

where the inequality follows from the fact that $(\nu_1, \cdots, \nu_s) \in A(\pi, \sigma)$. To complete the proof, note that by the triangle inequality,

$$\mathsf{d}(\pi, \sigma) \leq \min_{(\tau_1, \cdots, \tau_{s'}) \in A(\pi, \sigma)} \sum_{i=1}^{s'} \mathsf{d}(\tau_i, e).$$

∎

Definition 10. A distance d_φ is termed a *weighted Kendall distance* if there exists a nonnegative *weight function* φ over the set of adjacent transpositions such that

$$\mathsf{d}_\varphi(\pi, \sigma) = \min_{(\tau_1, \cdots, \tau_s) \in A(\pi, \sigma)} \sum_{i=1}^{s} \varphi_{\tau_i},$$

where φ_{τ_i} is the weight assigned to transposition τ_i by φ.

The weight of a transform $\tau = (\tau_1, \cdots, \tau_s)$ is denoted by $\mathrm{wt}(\tau)$ and is defined as

$$\mathrm{wt}(\tau) = \sum_{i=1}^{s} \varphi_{\tau_i}.$$

Hence, $\mathsf{d}_\varphi(\pi, \sigma)$ may be written as

$$\mathsf{d}_\varphi(\pi, \sigma) = \min_{\tau \in A(\pi, \sigma)} \mathrm{wt}(\tau).$$

Note that a weighted Kendall distance is completely determined by its weight function φ.

Theorem 11. *A distance* d *satisfies Axioms II if and only if it is a weighted Kendall distance.*

Proof: It follows immediately from Lemma 9 that a distance d satisfying Axioms II is a weighted Kendall distance by letting

$$\varphi_\theta = \mathsf{d}(\theta, e)$$

for every transposition θ taken from the set of adjacent transpositions $\mathbb{A}_n$ in $\mathbb{S}_n$.

The proof of the converse is omitted since it is easy to verify that a weighted Kendall distance satisfies Axioms II. ■

The weighted Kendall distance provides a natural solution for the top-vs-bottom issue. For instance, recall the example of ranking cities to live in, with

$$\pi = (\text{Melbourne, Vienna, Vancouver, Toronto, Calgary,}$$
$$\text{Adelaide, Sydney, Helsinki, Perth, Auckland}),$$
$$\pi' = (\text{Melbourne, Vienna, Vancouver, Calgary, Toronto,}$$
$$\text{Adelaide, Sydney, Helsinki, Perth, Auckland}),$$
$$\pi'' = (\text{Vienna, Melbourne, Vancouver, Toronto, Calgary,}$$
$$\text{Adelaide, Sydney, Helsinki, Perth, Auckland}),$$

and choose the weight function $\varphi_{(i\,i+1)} = 0.9^{i-1}$ for $i = 1, 2, \cdots, 9$. Then, $\mathsf{d}_\varphi(\pi, \pi') = 0.9^4 = 0.66 < \mathsf{d}_\varphi(\pi, \pi'') = 1$ as expected. In this case, we have chosen the weight function to be exponentially decreasing – the choice of the weight function in general depends on the application.

Computing the Weighted Kendall Distance for Monotonic Weight Functions

Computing the weighted Kendall distance between two permutations for an arbitrary weight function is not as straightforward a task as computing the Kendall τ distance. However, in what follows, we show that for an important class of weight functions – termed "monotonic" weight functions – the weighted Kendall distance may be computed efficiently.

Definition 12. A weight function $\varphi : \mathbb{A}_n \to \mathbb{R}^+$, where $\mathbb{A}_n$ as before denotes the set of adjacent transpositions in $\mathbb{S}_n$, is decreasing if $i > j$ implies that $\varphi_{(i\,i+1)} \leq \varphi_{(j\,j+1)}$. Increasing weight functions are defined similarly. A weight function is monotonic if it is increasing or decreasing.

Monotonic weight functions are of importance in the top-vs-bottom model as they can be used to emphasize the significance of the top of the ranking by assigning higher weights to transpositions at the top of the list. An example of a decreasing weight function is the exponential weight described in the previous subsection.

Suppose that $\tau = (\tau_1, \cdots, \tau_{|\tau|})$ of length $|\tau|$ transforms π into σ. The transformation may be viewed as a sequence of moves of elements i, $i = 1, \ldots, n$, from position $\pi^{-1}(i)$ to position $\sigma^{-1}(i)$. Let the *walk* followed by element i while moved by the transform τ be denoted by $p^{i,\tau} = \left(p_1^{i,\tau}, \cdots, p_{|p^{i,\tau}|+1}^{i,\tau}\right)$, where $|p^{i,\tau}|$ is the length of the walk $p^{i,\tau}$.

For example, consider

$$\pi = (3, 2, 4, 1),$$
$$\sigma = (1, 2, 3, 4),$$
$$\tau = (\tau_1, \tau_2, \tau_3, \tau_4)$$
$$= ((3\,4), (2\,3), (1\,2), (2\,3))$$

and note that $\sigma = \pi\tau_1\tau_2\tau_3\tau_4$. We have

$$p^{1,\tau} = (4, 3, 2, 1),$$
$$p^{2,\tau} = (2, 3, 2),$$
$$p^{3,\tau} = (1, 2, 3),$$
$$p^{4,\tau} = (3, 4).$$

We first bound the lengths of the walks $p^{i,\tau}, i \in [n]$. Let $I_i(\pi, \sigma)$ be the set consisting of elements $j \in [n]$ such that π and σ disagree on the pair $\{i, j\}$. In the transform τ, all elements of $I_i(\pi, \sigma)$ must be swapped with i by some $\tau_k, k \in [|\tau|]$. Each such swap contributes length one to the total length of the walk $p^{i,\tau}$ and thus, $|p^{i,\tau}| \geq |I_i(\pi, \sigma)|$.

As before, let d_φ denote the weighted Kendall distance with weight function φ. Since for any $\tau \in A(\pi, \sigma)$,

$$\sum_{i=1}^{|\tau|} \varphi_{\tau_i} = \sum_{i=1}^{n} \frac{1}{2} \sum_{j=1}^{|p^{i,\tau}|} \varphi_{(p_j^{i,\tau}\, p_{j+1}^{i,\tau})},$$

Algorithm 1 FindTauMonotone

Input: $\pi, \sigma \in \mathbb{S}_n$
Output: $\tau^\star = \arg\min_{\tau \in A(\pi,\sigma)} \mathrm{wt}(\tau)$
1: $\pi_0 \leftarrow \pi$
2: $t \leftarrow 0$
3: **for** $r = \sigma(1), \sigma(2), \cdots, \sigma(n)$ **do**
4:    **while** $\pi_t^{-1}(r) > \sigma^{-1}(r)$ **do**
5:       $\tau_{t+1}^\star \leftarrow \left(\pi_t^{-1}(r) - 1 \quad \pi_t^{-1}(r)\right)$
6:       $\pi_{t+1} \leftarrow \pi_t \tau_{t+1}^\star$
7:       $t \leftarrow t + 1$
8:    **end while**
9: **end for**

---

we have

$$\mathsf{d}_\varphi(\pi, \sigma) = \min_{\tau \in A(\pi,\sigma)} \sum_{i=1}^n \frac{1}{2} \sum_{j=1}^{|p^{i,\tau}|} \varphi_{(p_j^{i,\tau} \, p_{j+1}^{i,\tau})}.$$

Thus,

$$\mathsf{d}_\varphi(\pi, \sigma) \geq \sum_{i=1}^n \frac{1}{2} \min_{p^i \in P_i(\pi,\sigma)} \sum_{j=1}^{|p^i|} \varphi_{(p_j^i \, p_{j+1}^i)}, \tag{11}$$

where for each i, $P_i(\pi, \sigma)$ denotes the set of walks of length $|I_i(\pi, \sigma)|$, starting from $\pi^{-1}(i)$ and ending in $\sigma^{-1}(i)$. For convenience, let

$$p^{i,\star}(\pi, \sigma) = \arg\min_{p^i \in P_i(\pi,\sigma)} \sum_{j=1}^{|p^i|} \varphi_{(p_j^i \, p_{j+1}^i)}$$

be the minimum weight walk from $\pi^{-1}(i)$ to $\sigma^{-1}(i)$ with length $|I_i(\pi, \sigma)|$.

If clear from the context, we write $p^{i,\star}(\pi, \sigma)$ as $p^{i,\star}$.

We show next that for decreasing weight functions, the bound given in (11) is achievable and thus the value on the right-hand-side gives the weighted Kendall distance for this class of weight functions.

Consider $\pi, \sigma \in \mathbb{S}_n$ and a decreasing weight function φ. For each i, it follows that $p^{i,\star}(\pi, \sigma)$ extends to positions with largest possible indices, i.e., $p^{i,\star} = (\pi^{-1}(i), \cdots, \ell_i - 1, \ell_i, \ell_i - 1, \cdots, \sigma^{-1}(i))$ where ℓ_i is the solution to the equation

$$\ell_i - \pi^{-1}(i) + \ell_i - \sigma^{-1}(i) = I_i(\pi, \sigma)$$

and thus $\ell_i = \left(\pi^{-1}(i) + \sigma^{-1}(i) + I_i(\pi, \sigma)\right)/2$.

We show next that there exists a transform $\tau^\star$ with $p^{i,\tau^\star} = p^{i,\star}$, and so equality in (11) can be achieved. The transform is described in Algorithm 1. The transform in question, $\tau^\star$, converts π into σ in n rounds. In Algorithm 1, the variable r takes values $\sigma(1), \sigma(2), \cdots, \sigma(n)$, in that given order. For each value of r, $\tau^\star$ moves r through a sequence of adjacent transpositions from its current position in π_t, $\pi_t^{-1}(r)$, to position $\sigma^{-1}(r)$.

Fix $i \in [n]$. For values of r, used in Algorithm 1, such that $\sigma^{-1}(r) < \sigma^{-1}(i)$, i is swapped with r via an adjacent transposition if $\pi^{-1}(r) > \pi^{-1}(i)$. For $r = i$, i is swapped with all elements k such that $\pi^{-1}(k) < \pi^{-1}(i)$ and $\sigma^{-1}(i) < \sigma^{-1}(k)$. For r such that $\sigma^{-1}(r) > \sigma^{-1}(i)$, i is not swapped with other elements. Hence, i is swapped precisely with elements of the set $I_i(\pi, \sigma)$ and thus, $|p^{i,\tau^\star}(\pi, \sigma)| = |I_i(\pi, \sigma)|$. Furthermore, it can be seen that, for each i, $p^{i,\tau^\star}(\pi, \sigma) = (\pi^{-1}(i), \cdots, \ell_i' - 1, \ell_i', \ell_i' - 1, \cdots, \sigma^{-1}(i))$, for some ℓ_i'. Since $|p^{i,\tau^\star}(\pi, \sigma)| = |I_i(\pi, \sigma)|$, ℓ_i' also satisfies the equation

$$\ell_i' - \pi^{-1}(i) + \ell_i' - \sigma^{-1}(i) = I_i(\pi, \sigma),$$

implying that $\ell_i' = \ell_i$ and thus $p^{i,\tau^\star} = p^{i,\star}$. Consequently, one has the following result.

Proposition 13. *For rankings $\pi, \sigma \in \mathbb{S}_n$, and a decreasing weighted Kendall weight function φ, we have*

$$\mathsf{d}_\varphi(\pi, \sigma) = \sum_{i=1}^n \frac{1}{2} \left(\sum_{j=\pi^{-1}(i)}^{\ell_i - 1} \varphi_{(j \, j+1)} + \sum_{j=\sigma^{-1}(i)}^{\ell_i - 1} \varphi_{(j \, j+1)} \right)$$

where $\ell_i = \left(\pi^{-1}(i) + \sigma^{-1}(i) + I_i(\pi, \sigma)\right)/2$.

Increasing weight functions may be analyzed similarly.

Example 14. Consider the rankings $\pi = 4312$ and $e = 1234$ and a decreasing weight function φ. We have $I_i(\pi, e) = 2$ for $i = 1, 2$ and $I_i(\pi, e) = 3$ for $i = 3, 4$. Furthermore,

$$\ell_1 = \frac{3 + 1 + 2}{2} = 3, \qquad\qquad p^{1,\star} = (3, 2, 1),$$

$$\ell_2 = \frac{4 + 2 + 2}{2} = 4, \qquad\qquad p^{2,\star} = (4, 3, 2),$$

$$\ell_3 = \frac{2 + 3 + 3}{2} = 4, \qquad\qquad p^{3,\star} = (2, 3, 4, 3),$$

$$\ell_4 = \frac{1 + 4 + 3}{2} = 4, \qquad\qquad p^{4,\star} = (1, 2, 3, 4).$$

The minimum weight transformation is

$$\tau^\star = \left(\underbrace{(3\,2), (2\,1)}_{1}, \underbrace{(4\,3), (3\,2)}_{2}, \underbrace{(4\,3)}_{3} \right),$$

where the numbers under the braces denote the value r corresponding to the indicated transpositions. The distance between π and e is

$$\mathsf{d}_\varphi(\pi, e) = \varphi_{(1\,2)} + 2\varphi_{(2\,3)} + 2\varphi_{(3\,4)}.$$

Example 15. The bound given in (11) is not tight for general weight functions as seen in this example. Consider $\pi = (4, 2, 3, 1)$, $\sigma = (1, 2, 3, 4)$, and a weight function φ with $\varphi_{(1\,2)} = 2, \varphi_{(2\,3)} = 1$, and $\varphi_{(3\,4)} = 2$. Note that the domain of φ is the set of adjacent transpositions. We have

$$p^{1,\star} = (4, 3, 2, 1),$$
$$p^{2,\star} = (2, 3, 2),$$
$$p^{3,\star} = (3, 2, 3),$$
$$p^{4,\star} = (1, 2, 3, 4).$$

Suppose that a transform τ exists such that $p^{i,\star} = p^{i,\tau}, i = 1, 2, 3, 4$. From $p^{i,\star}$, it follows that in τ, transpositions $(1\,2)$ and $(3\,4)$ each appear once and $(2\,3)$ appears twice. It can be shown, by considering all possible re-orderings of $\{(1\,2), (1\,2), (2\,3), (2\,3), (2\,3)\}$ or by an application of [12, Lemma 5] that τ does not transform π into σ. Hence, for this example, the lower bound (11) is not achievable.

Weight Functions with Two Identical Non-zero Weights

Another example of a weighted Kendall τ distance for which a closed form solution may be found is described below.

For a pair of integers $a, b, 1 \leq a < b < n$, define the weight function as:

$$\varphi_{(i\ i+1)} = \begin{cases} 1, & i \in \{a, b\} \\ 0, & \text{else,} \end{cases} \tag{12}$$

i.e., a function which only penalizes moves involving candidates in positions a and b.

Such weight functions may be used in voting problems where one only penalizes moving a link from one page (say, top-ten page) to another page (say, ten-to-twenty page). In other words, one only penalizes moving an item from a "high-ranked" set of positions to "average-rank" or "low-rank" positions.

An algorithm for computing the weighted Kendall distance for this case is given in the Appendix.

Approximating the Weighted Kendall Distance for General Weight Functions

The result of the previous subsection implies that at least for one class of weight functions that capture the importance of the top entries in a ranking, computing the weighted Kendall distance has time complexity $O(n^2)$. Hence, distance computation efficiency does not represent a bottleneck for the employment of this form of the weighted Kendall distance.

In what follows, we present a polynomial-time 2-approximation algorithm for computing the most general form of weighted Kendall distances, as well as two algorithms for computing this distance exactly. While the exact computation has super exponential time complexity, for a small number of candidates – say, less than 10 – the computation can be performed in reasonable time. A small number of candidates and a large number of voters are frequently encountered in social choice applications, but less frequently in computer science.

In order to approximate the weighted Kendall distance, $\mathsf{d}_\varphi(\pi, \sigma)$, we use the function $D_\varphi(\pi, \sigma)$, defined as

$$D_\varphi(\pi, \sigma) = \sum_{i=1}^{n} w(\pi^{-1}(i) : \sigma^{-1}(i)),$$

where

$$w(k:l) = \begin{cases} \sum_{h=k}^{l-1} \varphi_{(h\,h+1)}, & \text{if } k < l, \\ \sum_{h=l}^{k-1} \varphi_{(h\,h+1)}, & \text{if } k > l, \\ 0, & \text{if } k = l, \end{cases}$$

denotes the sum of the weights of adjacent transpositions $(k\,k+1), (k+1\,k+2), \cdots, (l-1\,l)$ if $k < l$, the sum of the weights of adjacent transpositions $(l\,l+1), (l+1\,l+2), \cdots, (k-1\,k)$ if $l < k$, and 0 if $k = l$.

The following proposition states lower and upper bounds for d_φ in terms of D_φ. The propositions is useful in practice, since D_φ can be computed in time $O(n^2)$, and provides the desired 2-approximation.

Proposition 16. *For a weighted Kendall weight function φ and for permutations π and σ,*

$$\frac{1}{2} D_\varphi(\pi, \sigma) \le \mathsf{d}_\varphi(\pi, \sigma) \le D_\varphi(\pi, \sigma).$$

We omit the proof of the proposition, since it follows from a more general result stated in the next section, and only remark that the lower-bound presented above proposition is weaker than the lower-bound given by (11).

Next, we discuss computing the *exact* weighted Kendall distance via algorithms for finding minimum weight paths in graphs. As already pointed out, the Kendall τ and the weighted Kendall distance are graphic distances. In the latter case, we define a graph G with vertex set indexed by $\mathbb{S}_n$ and an edge of weight $\varphi_{(i\,i+1)}, i \in [n-1]$, between each pair of vertices π and σ for which there exists an i such that $\pi = \sigma(i\,i+1)$. The numbers of vertices and edges of G are $|V| = n!$ and $|E| = n!(n-1)/2$, respectively. Dijkstra's algorithm with Fibonacci heaps [14] for finding the minimum weight path in a graph provides the distances of all $\pi \in \mathbb{S}_n$ to the identity in time $O(|E| + |V| \log |V|) = O(n!\, n \log n)$.

The complexity of the algorithm for finding the distance between $\pi \in \mathbb{S}_n$ and the identity may be actually shown to be $O(n(K(\pi, e))!)$, which is significantly smaller than $\Omega(n!)$ for permutations at small Kendall τ distance. The minimum weight path algorithm is based on the following observation. For π in $\mathbb{S}_n$, there exists a transform $\tau = (\tau_1, \tau_2, \cdots, \tau_m)$ of minimum weight that transforms π into e, such that $m = K(\pi, e)$. In other words, each transposition of τ eliminates one inversion when transforming π into e. Hence, $\pi\tau_1$ has one less inversion than π. As a result,

$$\mathsf{d}_\varphi(\pi, e) = \min_{i:\pi(i)>\pi(i+1)} \left(\varphi_{(i\,i+1)} + \mathsf{d}_\varphi(\pi(i\,i+1), e) \right) \tag{13}$$

Note that the minimum is taken over all positions i for which i and $i+1$ form an inversion, i.e., for which $\pi(i) > \pi(i+1)$. Suppose that computing the weighted Kendall distance between the identity and a permutation π, with $K(\pi, e) = d$, can be performed in time T_d. From (13), we have

$$T_d = a\,n + d\,T_{d-1}, \quad \text{for } d \ge 2,$$

and $T_1 = a\,n$, for some constant a. By letting $U_d = T_d/(a\,n\,d!)$, we obtain $U_d = U_{d-1} + \frac{1}{d!}, d \ge 2$, and $U_1 = 1$. Hence, $U_d = \sum_{i=1}^{d} \frac{1}{i!}$. It can then be shown that $d!\,U_d = \lfloor d!(e-1) \rfloor$, and thus $T_d = a\,n\lfloor d!\,(e-1) \rfloor = O(nd!)$.

The expression (13) can also be used to find the distances of all $\pi \in \mathbb{S}_n$ from the identity by first finding the distances of permutations $\pi \in \mathbb{S}_n$ with $K(\pi, e) = 1$, then finding the distances of permutations $\pi \in \mathbb{S}_n$ with $K(\pi, e) = 2$, and so on[2]. Unfortunately, the average Kendall τ distance between a randomly chosen permutation and the identity is $\binom{n}{2}/2$ (see the derivation of this known and a related novel result regarding the weighted Kendall distance in the Appendix), which limits the applicability of this algorithm to uniformly and at random chosen votes limited.

Aggregation with Weighted Kendall Distances: Examples

In order to explain the potential of the weighted Kendall distance in addressing the top-vs-bottom aggregation issue, in what follows, we present a number of examples that illustrate how the choice of the weight function influences the final form of the aggregate. We focus on *decreasing weight functions* and compare our results to those obtained using the classical Kendall τ distance.

Throughout the remainder of the paper, we refer to a solution of the aggregation problem using the Kendall τ as a *Kemeny aggregate*. All the aggregation results are obtained via exhaustive search since the examples are small and only used for illustrative purposes. Aggregation is, in general, a hard problem and we postpone the analysis of the complexity of computing aggregate rankings, and aggregate approximation algorithms, until the next section.

Example 17. Consider the set of rankings listed in Σ, where each row represents a ranking (vote),

$$\Sigma = \begin{pmatrix} \begin{array}{|ccccc|} \hline 4 & 1 & 2 & 5 & 3 \\ \hline 4 & 2 & 1 & 3 & 5 \\ \hline 1 & 4 & 5 & 2 & 3 \\ \hline 2 & 3 & 1 & 5 & 4 \\ \hline 5 & 3 & 1 & 2 & 4 \\ \hline \end{array} \end{pmatrix}.$$

---

[2]Note that such an algorithm requires that the set of permutations at a given Kendall τ distance from the identity be known.

The Kemeny optimal solution for this set of rankings is $(1, 4, 2, 5, 3)$. Note that despite the fact that candidate 4 was ranked twice at the top of the list – more than any other candidate – it is ranked only second in the aggregate. This may be attributed to the fact that 4 was ranked last by two voters.

Consider next the weight function $\varphi^{(2/3)}$ with $\varphi^{(2/3)}_{(i\,i+1)} = (2/3)^{i-1}, i \in [4]$. The optimum aggregate ranking for this weight equals $(4, 1, 2, 5, 3)$. The optimum aggregate based on $\varphi^{(2/3)}$ puts 4 before 1, similar to what a plurality vote would do[3]. The reason behind this swap is that $\varphi^{(2/3)}$ emphasizes strong showings of a candidate and downplays its weak showings, since weak showings have a smaller effect on the distance as the weight function is decreasing. In other words, higher ranks are more important than lower ranks when determining the position of a candidate.

Example 18. Consider the set of rankings listed in Σ,

$$\Sigma = \begin{pmatrix} \begin{array}{cccc} 1 & 4 & 2 & 3 \\ \hline 1 & 4 & 3 & 2 \\ \hline 2 & 3 & 1 & 4 \\ \hline 4 & 2 & 3 & 1 \\ \hline 3 & 2 & 4 & 1 \end{array} \end{pmatrix}.$$

The Kemeny optimal solution is $(4, 2, 3, 1)$. Note that although the majority of voters prefer 1 to 4, 1 is ranked last and 4 is ranked first. More precisely, we observe that according to the pairwise majority test, 1 beats 4 but loses to 2 and 3. On the other hand, 4 is preferred to both 2 and 3 but, as mentioned before, loses to 1. Problems like this do not arise due to a weakness of Kemeny's approach, but due to the inherent "rational intractability" of rank aggregation. As stated by Arrow [1], for *any* "reasonable" rank aggregation method, there exists a set of votes such that the aggregated ranking prefers one candidate to another while the majority of voters prefer the later to the former.

Let us now focus on a weighted Kendall distance with weight function $\varphi_{(i\,i+1)} = (2/3)^{i-1}, i = 1, 2, 3$. The optimal aggregate ranking for this distance equals $(1, 4, 2, 3)$. Again, we see a candidate with both strong showings and weak showings, candidate 1, beat a candidate with a rather average performance. Note that in this solution as well, there exist candidates for which the opinion of the majority is ignored: 1 is placed before 2 and 3, while according to the pairwise majority opinion it loses to both.

Example 19. Consider the set of rankings listed in Σ,

$$\Sigma = \begin{pmatrix} \begin{array}{ccccc} 5 & 4 & 1 & 3 & 2 \\ \hline 1 & 5 & 4 & 2 & 3 \\ \hline 4 & 3 & 5 & 1 & 2 \\ \hline 1 & 3 & 4 & 5 & 2 \\ \hline 4 & 2 & 5 & 3 & 1 \\ \hline 1 & 2 & 5 & 3 & 4 \\ \hline 2 & 4 & 3 & 5 & 1 \end{array} \end{pmatrix}.$$

With the weight function $\varphi_{(i\,i+1)} = (2/3)^{i-1}, i \in [4]$, the aggregate equals $(4, 1, 5, 2, 3)$. The winner is 4, while the plurality rule winner is 1 as it appears three times on the top. Next, we increase the rate of decay of the weight function and let $\varphi_{(i\,i+1)} = (1/3)^{i-1}, i \in [4]$. The solution now is $(1, 4, 2, 5, 3)$, and the winner is candidate 1, the same as the plurality rule winner. This result is a consequence of the fact that the plurality winner is the aggregate based on the weighted Kendall distance with weight function $\varphi^{(p)}$,

$$\varphi^{(p)}_{(i\,i+1)} = \begin{cases} 1, & i = 1, \\ 0, & \text{else.} \end{cases}$$

The Kemeny aggregate is $(4, 5, 1, 2, 3)$.

A shortcoming of distance-based rank aggregation is that sometimes the solution is not unique, and that the possible solutions differ widely. The following example describes one such scenario.

Example 20. Suppose that the votes are given by Σ,

$$\Sigma = \begin{pmatrix} \begin{array}{ccc} 1 & 2 & 3 \\ \hline 1 & 2 & 3 \\ \hline 3 & 2 & 1 \\ \hline 2 & 1 & 3 \end{array} \end{pmatrix}.$$

Here, the permutations $(1, 2, 3), (2, 1, 3)$ are the Kemeny optimal solutions, with cumulative distance 4 from Σ. When the Kemeny optimal solution is not unique, it may be possible to obtain a unique solution by using a non-uniform weight function.

---

[3]In *plurality votes*, the candidate with the most first-place rankings is declared the winner.

In this example, it can be shown that for any non-uniform weight function φ with $\varphi_{(1\,2)} > \varphi_{(2\,3)}$, the solution is unique, namely, $(1, 2, 3)$.

A similar situation occurs if the last vote is changed to $(2, 3, 1)$. In that case, the permutations $(1, 2, 3)$, $(2, 1, 3)$, and $(2, 3, 1)$ are the Kemeny optimal solutions with cumulative distance 5 from Σ. Again, for any non-uniform weight function φ with $\varphi_{(1\,2)} > \varphi_{(2\,3)}$ the solution is unique and equal to $(1, 2, 3)$.

To summarize, the above examples illustrate how a proper choice for the weighted Kendall distance insures that top ranks are emphasized and how one may over-rule a moderate number of low rankings using a specialized distance formula. One may argue that certain generalizations of Borda's method, involving non-uniform gaps between ranking scores, may achieve similar goals. This is not the case, as will be illustrated in what follows.

One major difference between generalized Borda and weighted Kendall distances is in the already mentioned *majority criteria* [15], which states that the candidate ranked first by the majority of voters has to be ranked first in the aggregate[4]. Borda's aggregate with an arbitrary score assignments does not have this property, while aggregates obtained via weighted Kendall distances with decreasing weights (not identically equal to zero) have this property.

We first show that the Borda method with a fixed, but otherwise arbitrary set of scores may not satisfy the majority criterion. We prove this claim for $n = 3$. A similar argument can be used to establish this claim for $n > 3$.

Suppose, for simplicity, that the number m of voters is odd and that, for each vote, a score s_i is assigned to a candidate with rank i, $i = 1, 2, 3$. Here, we assume that $s_1 > s_2 > s_3 \geq 0$. Suppose also that $(m + 1)/2$ of the votes equal (a, b, c) and that $(m - 1)/2$ of the votes equal (b, c, a). Let the total Borda scores for candidates a and b be denoted by S and S', respectively. We have

$$S = \frac{m+1}{2}s_1 + \frac{m-1}{2}s_3,$$
$$S' = \frac{m+1}{2}s_2 + \frac{m-1}{2}s_1,$$

and thus $S - S' = s_1 - m\left(\frac{s_2 - s_3}{2}\right) - \frac{s_2 + s_3}{2}$. If $m > \frac{2s_1 - (s_2 + s_3)}{s_2 - s_3}$, then $S - S' < 0$ and Borda's method ranks b higher than a. As a result, candidates a, ranked highest by more than half of the voters, is not ranked first according to Borda's rule. This is not the case with weighted Kendall distances, as shown below.

Proposition 21. *An aggregate ranking obtained using the weighted Kendall distance with a decreasing weight function not identically equal to zero satisfies the majority criterion.*

Proof: Suppose that the weight function is φ, and let $w_i = \varphi_{(i\ i+1)}$. Since w is decreasing and not identically equal to zero, we have $w_1 > 0$. Let a_1 be a candidate that is ranked first by a majority of voters. Partition the set of votes into two sets, C and D, where C is the set of votes that rank a_1 first and D is the set of votes that do not. Furthermore, denote the aggregate ranking by π.

Suppose that a_1 is not ranked first in π and that π is of the form

$$(a_2, \cdots, a_i, a_1, a_{i+1}, \cdots, a_n),$$

for some $i \geq 2$. Let $\pi' = (a_1, a_2, \cdots, a_n)$. We show that

$$\sum_{j=1}^{m} \mathsf{d}_\varphi(\pi, \sigma_j) > \sum_{j=1}^{m} \mathsf{d}_\varphi(\pi', \sigma_j)$$

which contradicts the optimality of π. Hence, a_1 must be ranked first in π.

For $\sigma \in C$, we have

$$\mathsf{d}_\varphi(\pi, \sigma) = \mathsf{d}_\varphi(\pi, \pi') + \mathsf{d}_\varphi(\pi', \sigma). \tag{14}$$

To see the validity of this claim, note that if π is to be transformed to σ via Algorithm 1, it is first transformed to π' by moving a_1 to the first position.

For $\sigma \in D$, we have

$$\mathsf{d}_\varphi(\pi', \sigma) \leq \mathsf{d}_\varphi(\pi', \pi) + \mathsf{d}_\varphi(\pi, \sigma), \tag{15}$$

which follows from the triangle inequality.

---

[4] Note that a candidate ranked first by the majority is a Condorcet candidate. It is desirable that an aggregation rule satisfy the majority criterion and indeed most do, including the Condorcet method, the plurality rule, the single transferable vote method, and the Coombs method.

To complete the proof, we write

$$
\begin{aligned}
\sum_{j=1}^{m} \mathsf{d}_\varphi(\pi, \sigma_j) &= \sum_{\sigma \in C} \mathsf{d}_\varphi(\pi, \sigma) + \sum_{\sigma \in D} \mathsf{d}_\varphi(\pi, \sigma) \\
&\geq \sum_{\sigma \in C} \mathsf{d}_\varphi(\pi', \sigma) + |C| \mathsf{d}_\varphi(\pi, \pi') \\
&\quad + \sum_{\sigma \in C} \mathsf{d}_\varphi(\pi', \sigma) - |D| \mathsf{d}_\varphi(\pi, \pi') \\
&= \sum_{j=1}^{m} \mathsf{d}_\varphi(\pi', \sigma) + (|C| - |D|) \, \mathsf{d}_\varphi(\pi, \pi') \\
&> \sum_{j=1}^{m} \mathsf{d}_\varphi(\pi', \sigma)
\end{aligned}
$$

where the first inequality follows from (14) and (15), and the second inequality follows from the facts that $|C| > |D|$ and that $\mathsf{d}_\varphi(\pi, \pi') \geq w_1 > 0$.

∎

IV. WEIGHTED TRANSPOSITION DISTANCE

The definition of the Kendall τ distance and the weighted Kendall distance is based on transforming one permutation into another using *adjacent* transpositions. If, instead, all transpositions are allowed – including non-adjacent transpositions – a more general distance measure, termed *weighted transposition distance* is obtained. This distance measure, as will be demonstrated below, represents a generalization of the weighted Kendall distance suitable for addressing similarity issues among candidates. It is worth pointing out that the weighted transposition distance is not based on the axiomatic approach described in the previous section.

Definition 22. Consider a function φ that assigns to each transpositions θ, a non-negative weight φ_θ. The weight of a sequence of transpositions is defined as the sum of the weights of its transpositions. That is, the weight of the sequence $\tau = (\tau_1, \cdots, \tau_{|\tau|})$ of transpositions equals

$$
\mathrm{wt}(\tau) = \sum_{i=1}^{|\tau|} \varphi_{\tau_i}.
$$

For simplicity, we also denote the weighted transposition distance between two permutations $\pi, \sigma \in \mathbb{S}_n$, with weight function φ, by d_φ. This distance equals the minimum weight of a sequence $\tau = (\tau_1, \cdots, \tau_{|\tau|})$ of transpositions such that $\sigma = \pi \tau_1 \cdots \tau_{|\tau|}$. As before, we refer to such a sequence of transpositions as a transform converting π into σ and let $A_T(\pi, \sigma)$ denote the set of transforms that convert π into σ.

With this notation at hand, the weighted transposition distance between π and σ may be written as

$$
\mathsf{d}_\varphi(\pi, \sigma) = \min_{\tau \in A_T(\pi, \sigma)} \mathrm{wt}(\tau).
$$

The Kendall τ distance and the weighted Kendall distance may be viewed as special cases of the weighted transposition distance: to obtain the Kendall τ distance, let

$$
\varphi_\theta = \begin{cases} 1, & \theta = (i\ i+1), i = 1, 2, \cdots, n-1 \\ \infty, & \text{else,} \end{cases}
$$

and to obtain the weighted Kendall distance, let

$$
\varphi_\theta = \begin{cases} w_i, & \theta = (i\ i+1), i = 1, 2, \cdots, n-1 \\ \infty, & \text{else,} \end{cases}
$$

for a non-negative weight function w.

When applied to the inverse of rankings, the weighted transposition distance can be successfully used to model similarities of objects in rankings. In such a setting, permutations that differ by a transposition of two similar items are at a smaller distance than permutations that differ by a transposition of two dissimilar items, as demonstrated in the next subsection.

A. Weighted Transposition Distance as Similarity Distance: Examples

We illustrate the concept of distance measures taking into account similarities via the following example, already mentioned in the Motivation section. Suppose that four cities: Melbourne, Sydney, Helsinki, and Vienna are ranked based on certain criteria as

$$\pi = (\text{Helsinki, Sydney, Vienna, Melbourne}),$$

and according to another set of criteria as

$$\sigma = (\text{Melbourne, Vienna, Helsinki, Sydney}).$$

The distance between π and σ is defined as follows. We assign weights to swapping the positions of cities in the rankings, e.g., suppose that the weight of swapping cities in the same country is 1, on the same continent 2, and 3 otherwise. The *similarity distance* between π and σ is the minimum weight of a sequence of transpositions that convert π into σ. By inspection, or by methods discussed in the next subsection, one can see that the similarity distance between π and σ equals 6. One of the sequences of transpositions of weight 6 is as follows: first swap Helsinki and Sydney with weight 3, then swap Melbourne and Sydney with weight 1, and finally swap Vienna and Helsinki with weight 2.

To express the similarity distance formally, we write the rankings as permutations, representing Melbourne by 1, Sydney by 2, Vienna by 3, and Helsinki by 4. This is equivalent to assuming that the identity ranking is

$$e = (\text{Melbourne, Sydney, Vienna, Helsinki}).$$

We then have $\pi = (4, 2, 3, 1)$ and $\sigma = (1, 4, 2, 3)$. The weight function φ equals

$$\varphi_{(1\,2)} = 1, \quad \varphi_{(1\,3)} = 3, \quad \varphi_{(1\,4)} = 3$$
$$\varphi_{(2\,3)} = 3, \quad \varphi_{(2\,4)} = 3, \quad \varphi_{(3\,4)} = 2.$$

It should be clear from the context that the indices in the weight function refer to the candidates, and not to their positions.

Example 23. Consider the votes listed in Σ below,

$$\Sigma = \left(\begin{array}{cccc} 1 & 2 & 3 & 4 \\ \hline 3 & 2 & 1 & 4 \\ 4 & 1 & 3 & 2 \end{array} \right).$$

Suppose that even numbers and odd numbers represent different types of candidates in a way that the following weight function is appropriate

$$\varphi_{(i\,j)} = \begin{cases} 1, & \text{if } i, j \text{ are both odd or both even,} \\ 2, & \text{else.} \end{cases}$$

Note that the votes are "diverse" in the sense that they alternate between odd and even numbers. On the other hand, the Kemeny aggregate is $(1, 3, 2, 4)$, which puts all odd numbers ahead of all even numbers. Aggregation using the similarity distance described above yields $(1, 2, 3, 4)$, a solution which may be considered "diverse" since the even and odd numbers alternate in the solution. The reason behind this result is that the Kemeny optimal solution is oblivious to the identity of the candidates and their (dis)similarities, while aggregation based on similarity distances take such information into account.

Example 24. Consider the votes listed in Σ below,

$$\Sigma = \left(\begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \\ \hline 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 6 & 5 & 2 & 1 & 4 \\ 3 & 6 & 5 & 2 & 1 & 4 \\ 5 & 4 & 1 & 6 & 3 & 2. \end{array} \right).$$

Suppose that the weight function is the same as the one used in the previous example. In this case, neither the Kemeny aggregates nor the weighted transposition distance aggregates are unique. More precisely, Kendall τ gives four solutions:

$$\left(\begin{array}{cccccc} 3 & 5 & 1 & 6 & 2 & 4 \\ \hline 3 & 5 & 1 & 2 & 4 & 6 \\ 1 & 3 & 5 & 2 & 4 & 6. \\ 1 & 3 & 5 & 6 & 2 & 4. \end{array} \right),$$

while there exist nine optimal aggregates under the weighted transposition distance of the previous example, of total distance 10:

$$
\begin{pmatrix}
\begin{array}{cccccc}
5 & 6 & 3 & 4 & 1 & 2 \\ \hline
5 & 4 & 3 & 2 & 1 & 6 \\ \hline
5 & 2 & 3 & 6 & 1 & 4 \\ \hline
3 & 4 & 1 & 2 & 5 & 6 \\ \hline
3 & 6 & 1 & 4 & 5 & 2 \\ \hline
3 & 2 & 1 & 6 & 5 & 4 \\ \hline
1 & 4 & 5 & 2 & 3 & 6 \\ \hline
1 & 2 & 5 & 6 & 3 & 4 \\ \hline
1 & 6 & 5 & 4 & 3 & 2.
\end{array}
\end{pmatrix}
$$

Note that *none of* the Kemeny optimal aggregates have good diversity properties: the top half of the rankings consists exclusively of odd numbers. On the other hand, the optimal weighted transposition rankings *all contain* exactly one even element among the top-three candidates. Such diversity properties are hard to prove theoretically.

B. Computing the Weighted Transposition Distance

In this subsection, we describe how to compute or approximate the weighted transposition distance d_φ, given the weight function φ. An in-depth analysis of a special class of weight functions and their corresponding transposition distance may be found in the authors' recent work [12].

We find the following definitions useful in our subsequent derivations. For a given weight function φ, we let $\mathcal{K}_\varphi$ denote a complete undirected weighted graph with vertex set $[n]$, where the weight of each edge (i, j) equals the weight of the transposition $(i\,j)$, $\varphi_{(i\,j)}$. For a subgraph H of $\mathcal{K}_\varphi$, with edge set E_H, we define the weight of H as

$$
\mathrm{wt}(H) = \sum_{(i,j) \in E_H} \varphi_{(i\,j)},
$$

that is, the sum of the weights of edges of H. For $\pi, \sigma \in \mathbb{S}_n$, we define $D_\varphi(\pi, \sigma)$ as[5]

$$
D_\varphi(\pi, \sigma) = \sum_{i=1}^{n} \mathrm{wt}\left(p_\varphi^*(\pi^{-1}(i), \sigma^{-1}(i))\right),
$$

where $p_\varphi^*(a, b)$ denotes the minimum weight path from a to b in $\mathcal{K}_\varphi$.

It is easy to verify that D_φ is a pseudo-metric and that it is left-invariant,

$$
D_\varphi(\eta\pi, \eta\sigma) = D_\varphi(\pi, \sigma), \qquad \pi, \sigma, \eta \in \mathbb{S}_n.
$$

A weight function φ is a *metric weight function* if it satisfies the triangle inequality in the sense that

$$
\varphi_{(a\,b)} \leq \varphi_{(a\,c)} + \varphi_{(b\,c)}, \qquad a, b, c \in [n]. \tag{16}
$$

Lemma 25. *For a non-negative weight function φ and a transposition $(a\,b) \in \mathbb{S}_n$,*

$$
\mathsf{d}_\varphi((a\,b), e) \leq 2\mathrm{wt}(p_\varphi^*(a, b)).
$$

If φ is a metric weight function, the bound may be improved to

$$
\mathsf{d}_\varphi((a\,b), e) \leq \mathrm{wt}(p_\varphi^*(a, b)).
$$

Proof: Consider a path $p = (v_0 = a, v_1, \cdots, v_{|p|} = b)$ from a to b in $\mathcal{K}_\varphi$. We have

$$
(a\,b) = (v_0\,v_1)(v_1\,v_2) \cdots (v_{|p|-2}\,v_{|p|-1})
$$
$$
(v_{|p|-1}\,v_{|p|})(v_{|p|-2}\,v_{|p|-1}) \cdots (v_1\,v_2)(v_0\,v_1).
$$

From the left-invariance of d_φ,

$$
\mathsf{d}_\varphi((a\,b), e) = 2\sum_{i=1}^{|p|-1} \varphi_{(v_{i-1}\,v_i)} - \varphi_{(v_{|p|-1}\,v_{|p|})}
$$
$$
= 2\mathrm{wt}(p) - \varphi_{(v_{|p|-1}\,v_{|p|})}
$$
$$
\leq 2\mathrm{wt}(p).
$$

---

[5]Note that this definition is consistent with the definition of a specialization of this function, given in Proposition 16.

Since p is an arbitrary path from a to b in $\mathcal{K}_\varphi$, we have

$$\mathsf{d}_\varphi((a\,b), e) \leq \mathrm{wt}(\tau) \leq 2\mathrm{wt}(p_\varphi^*(a, b)),$$

and this proves the first claim.

Now, assume that φ is a metric weight function and consider the path $p = (v_0, v_1, \cdots, v_{|p|})$ from $v_0 = a$ to $v_{|p|} = b$. From (16),

$$\begin{aligned}
\varphi_{(a\,b)} = \varphi_{(v_0\,v_{|p|})} &\leq \varphi_{(v_0\,v_1)} + \varphi_{(v_1\,v_{|p|})} \\
&\leq \varphi_{(v_0\,v_1)} + \varphi_{(v_1\,v_2)} + \varphi_{(v_2\,v_{|p|})} \\
&\leq \cdots \\
&\leq \sum_{i=1}^{|p|-1} \varphi_{(v_i\,v_{i+1})} \\
&= \mathrm{wt}(p).
\end{aligned}$$

Since p is arbitrary, we have

$$\mathsf{d}_\varphi((a\,b), e) \leq \varphi_{(a\,b)} \leq \mathrm{wt}(p_\varphi^*(a, b)).$$

This completes the proof of the lemma. ∎

While Lemma 25 suffices to prove all our subsequent results, we remark that one may prove a slightly stronger result, presented in our companion paper [12],

$$\mathsf{d}_\varphi((a\ b), e) = \min_{p=(v_0=a, v_1, \cdots, v_{|p|}=b)} \left(2\mathrm{wt}(p) - \max_{0 \leq i < |p|} \varphi_{(v_i\,v_{i+1})} \right).$$

The proof is based on significantly more involved techniques that are beyond the scope of this paper.

Lemma 26. *For a weight function φ and for $\pi, \sigma \in \mathbb{S}_n$,*

$$\mathsf{d}_\varphi(\pi, \sigma) \leq 2D_\varphi(\pi, \sigma).$$

If φ is a metric weight function, the bound may be improved to

$$\mathsf{d}_\varphi(\pi, \sigma) \leq D_\varphi(\pi, \sigma).$$

Proof: To prove the first claim, it suffices to show that $\mathsf{d}_\varphi(\pi, e) \leq 2D_\varphi(\pi, e)$ since both d_φ and D_φ are left-invariant.

Let $\{c_1, c_2, \cdots, c_k\}$ be the cycle decomposition of π. We have, from the triangle inequality and the left-invariance property of d_φ, that

$$\mathsf{d}_\varphi(\pi, e) \leq \sum_{i=1}^{k} \mathsf{d}_\varphi(c_i, e),$$

and, from the definition of D_φ, that

$$D_\varphi(\pi, e) = \sum_{i=1}^{k} D_\varphi(c_i, e).$$

Hence, we only need to prove that

$$\mathsf{d}_\varphi(c, e) \leq 2D_\varphi(c, e) \tag{17}$$

for a single cycle $c = (a_1\,a_2\,\cdots\,a_{|c|})$, where $|c|$ is the length of c.

Since c may be written as

$$c = (a_1\,a_2)(a_2\,a_3)\cdots(a_{|c|-1}\,a_{|c|}),$$

we have

$$\begin{aligned}
\mathsf{d}_\varphi(c, e) &\leq \sum_{i=1}^{|c|-1} \varphi_{(a_i\,a_{i+1})} \\
&\overset{(a)}{\leq} \sum_{i=1}^{|c|-1} 2\mathrm{wt}(p_\varphi^*(a_i, a_{i+1})) \\
&\leq \sum_{i=1}^{|c|} 2\mathrm{wt}(p_\varphi^*(a_i, c(a_i))) \\
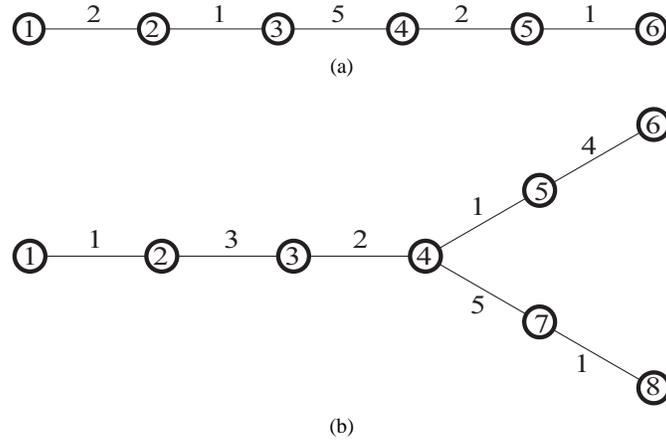&\leq 2D_\varphi(c, e)
\end{aligned}$$

Figure 3: A defining path (a), which may correspond to a metric-path weight function or an extended-path weight function, and a defining tree (b), which may correspond to a metric-tree weight function or an extended-tree weight function.

where (a) follows from Lemma 25.

The proof of the second claim is similar. ∎

The next lemma provides a lower bound for d_φ in terms of D_φ.

Lemma 27. *For $\pi, \sigma \in \mathbb{S}_n$,*

$$d_\varphi(\pi, \sigma) \geq \frac{1}{2} D_\varphi(\pi, \sigma).$$

Proof: Since d_φ and D_φ are both left-invariant, it suffices to show that

$$d_\varphi(\pi, e) \geq \frac{1}{2} D_\varphi(\pi, e).$$

Let $(\tau_1, \cdots, \tau_l)$, with $\tau_j = (a_j\, b_j)$, be a minimum weight transform of π into e, so that $d_\varphi(\pi, e) = \sum_{i=1}^{l} \varphi_{(a_j\, b_j)}$. Furthermore, define $\pi_j = \pi \tau_1 \cdots \tau_j$, $0 \leq j \leq l$. Then,

$$D_\varphi\left(\pi_{j-1}, e\right) - D_\varphi\left(\pi_j, e\right) \leq 2\mathrm{wt}\left(p_\varphi^*\left(a_j, b_j\right)\right)$$
$$\leq 2\varphi_{(a_j\, b_j)}, \tag{18}$$

where the first inequality follows from considering the maximum possible decrease of the value of D_φ induced by one transposition, while the second inequality follows from the definition of p_φ^*. By summing up the terms in (18) over $0 \leq j \leq l$, and thus obtaining a telescoping inequality of the form $D_\varphi(\pi, e) \leq 2\sum_{i=1}^{l} \varphi_{(a_j\, b_j)} = 2d_\varphi(\pi, e)$, we arrive at the desired result. ∎

From the previous two lemmas, we have the following theorem.

Theorem 28. *For $\pi, \sigma \in \mathbb{S}_n$ and an arbitrary non-negative weight function φ, we have*

$$\frac{1}{2} D_\varphi(\pi, \sigma) \leq d_\varphi(\pi, \sigma) \leq 2 D_\varphi(\pi, \sigma).$$

In addition, if φ is a metric weight function, then

$$\frac{1}{2} D_\varphi(\pi, \sigma) \leq d_\varphi(\pi, \sigma) \leq D_\varphi(\pi, \sigma).$$

For special classes of the weight function φ, the bounds in Theorem 28 may be improved further, as described in the next subsection.

C. Computing the Transposition Distance for Metric-Tree Weights

We start with the following definitions.

Definition 29. A weight function φ is a *metric-tree weight function* if there exists a weighted tree Θ over the vertex set $[n]$ such that for distinct $a, b \in [n]$, $\varphi_{(a\, b)}$ is the sum of the weights of the edges on the unique path from a to b in Θ. If Θ is a path, i.e., if Θ is a linear graph, then φ is called a *metric-path weight function*.
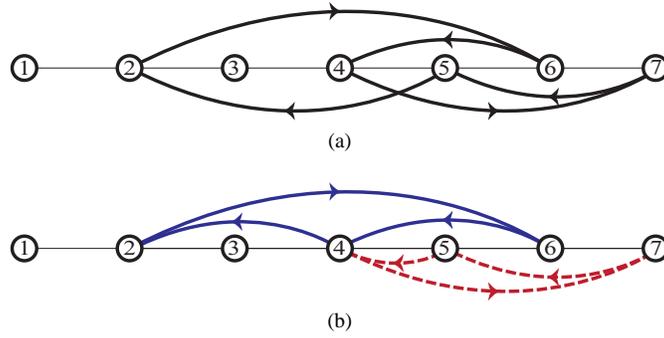
Figure 4: The cycle $(2\ 6\ 4\ 7\ 5)$ in Figure (a) is decomposed into two cycles, $(2\ 6\ 4)$ and $(4\ 7\ 5)$, depicted in Figure (b). Note that $(2\ 6\ 4\ 7\ 5) = (2\ 6\ 4)(4\ 7\ 5)$.

Furthermore, a weight function φ' is an *extended-tree weight function* if there exists a weighted tree Θ over the vertex set $[n]$ such that for distinct $a, b \in [n]$, $\varphi'_{(a\,b)}$ equals the the weight of the edge (a, b) whenever a and b are adjacent, and $\varphi'_{(a\,b)} = \infty$ otherwise. If Θ is a path, then φ' is called an *extended-path* weight function.

Note that the Kendall weight function, defined in the previous section, is an extended path weight function.

The tree or path corresponding to a weight function in the above definitions is termed the *defining tree or path* of the weight function. An example is given in Figure 3, where the numbers indexing the edges denote their weights.

For a metric-tree weight function φ with defining tree Θ, and for $a, b \in [n]$, the weight of the path $p^*_\varphi(a, b)$ equals the weight of the *unique* path from a to b in Θ. This weight, in turn, equals $\varphi_{(a\,b)}$. As a result, for metric-tree weights, $p^*_\varphi(a, b)$ equals the weight of the path from a to b in Θ.

Furthermore, from Lemma 27, we have $\mathsf{d}_\varphi((a\ b), e) \geq \frac{1}{2}D_\varphi((a\ b), e) = \mathrm{wt}(p^*_\varphi(a, b)) = \varphi_{(a\,b)}$. Since we also have $\mathsf{d}_\varphi((a\ b), e) \leq \varphi_{(a\,b)}$, it follows that

$$\mathsf{d}_\varphi((a\ b), e) = \varphi_{(a\,b)}. \tag{19}$$

The next lemma shows that the *exact distance* for metric-path weight functions can be computed in polynomial time.

Lemma 30. *For a metric-path weight function φ and for $\pi, \sigma \in \mathbb{S}_n$,*

$$\mathsf{d}_\varphi(\pi, \sigma) = \frac{1}{2}D_\varphi(\pi, \sigma).$$

Proof: From Lemma 27, we have that $\mathsf{d}_\varphi(\pi, \sigma) \geq \frac{1}{2}D_\varphi(\pi, \sigma)$. It remains to show that $\mathsf{d}_\varphi(\pi, \sigma) \leq \frac{1}{2}D_\varphi(\pi, \sigma)$. Since d_φ and D_φ are both left-invariant, it suffices to prove that $\mathsf{d}_\varphi(\pi, e) \leq \frac{1}{2}D_\varphi(\pi, e)$.

Let $\{c_1, c_2, \cdots, c_k\}$ be the cycle decomposition of π. Similar to the proof of Lemma 26, it suffices to show that

$$\mathsf{d}_\varphi(c, e) \leq \frac{1}{2}D_\varphi(c, e) \tag{20}$$

for any cycle $c = (a_1\, a_2\, \cdots\, a_{|c|})$.

The proof is by induction. For $|c| = 2$, (20) holds since, from (19), we have

$$\mathsf{d}_\varphi((a_1\, a_2), e) = \varphi_{(a\,b)} = \mathrm{wt}\left(p^*_\varphi(a_1, a_2)\right) = \frac{1}{2}D_\varphi((a_1\, a_2), e).$$

Assume that (20) holds for $2 \leq |c| < l$. We show that it also holds for $|c| = l$. We use Figure 4 for illustrative purposes. In all figures in this section, undirected edges describe the defining tree, while directed edges describe the cycle at hand.

Without loss of generality, assume that the defining path of φ, Θ, equals $(1, 2, \cdots, n)$. Furthermore, assume that $a_1 = \min\{i : i \in c\}$; if this were not the case, we could rewrite c by cyclically shifting its elements. Let $a_t = \min\{i : i \in c, i \neq a_1\}$ be the "closest" element to a_1 in Θ (that is, the closest element to a_1 in the cycle c). For example, in Figure 4, one has $c = (2\,6\,4\,7\,5)$, $a_1 = 2$ and $a_t = 4$. We have

$$c = (a_1\, a_2\, \cdots\, a_t\, \cdots\, a_l)$$
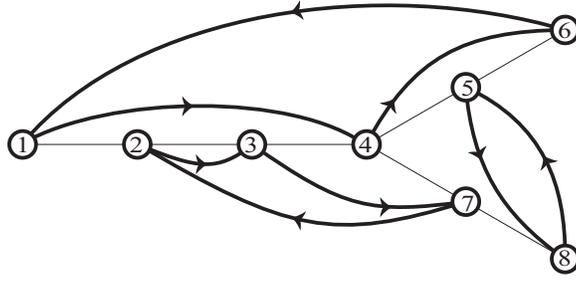$$= (a_1\, a_2\, \cdots\, a_t)(a_t\, a_{t+1}\, \cdots\, a_l)$$

Figure 5: If each of the cycles of a permutation lie on a path, the method of Lemma 30 can be used to find the weighted transposition distance.

and thus

$$\mathsf{d}_\varphi(c, e) \leq \mathsf{d}_\varphi((a_1 \ a_2 \ \cdots \ a_t), e) + \mathsf{d}_\varphi((a_t \ a_{t+1} \ \cdots \ a_l), e)$$

$$\leq \frac{1}{2} \sum_{i=1}^{t-1} \mathrm{wt}\left(p_\varphi^*(a_i, a_{i+1})\right) + \mathrm{wt}\left(p_\varphi^*(a_t, a_1)\right)$$

$$+ \frac{1}{2} \sum_{i=t}^{l-1} \mathrm{wt}\left(p_\varphi^*(a_i, a_{i+1})\right) + \mathrm{wt}\left(p_\varphi^*(a_l, a_t)\right)$$

$$= \frac{1}{2} \sum_{i=1}^{l} \mathrm{wt}\left(p_\varphi^*(a_i, c(a_i))\right)$$

$$= \frac{1}{2} D_\varphi(c, e).$$

where the second inequality follows from the induction hypothesis, while the first equality follows from the fact that $\mathrm{wt}\left(p_\varphi^*(a_t, a_1)\right) + \mathrm{wt}\left(p_\varphi^*(a_l, a_t)\right) = \mathrm{wt}\left(p_\varphi^*(a_l, a_1)\right)$. ∎

The approach described in the proof of Lemma 30 can also be applied to the problem of finding the weighted transposition distance when the weight function is a metric-tree weight function and *each of the cycles of the permutation* consist of elements that lie on some path in the defining tree. An example of such a permutation and such a weight function is shown in Figure 5. Note that in this example, a cycle consisting of elements $3, 5, 7$ would not correspond to a path.

In such a case, for each cycle c of π we can use the path in the defining tree that contains the elements of c to show that

$$\mathsf{d}_\varphi(c, e) = \frac{1}{2} D_\varphi(c, e). \tag{21}$$

For example the cycle $(1 \ 4 \ 6)$ lies on the path $(1, 2, 3, 4, 5, 6)$ and the cycle $(5 \ 8)$ lies on the path $(5, 4, 7, 8)$. Since (21) holds for each cycle c of π, we have

$$\mathsf{d}_\varphi(\pi, e) = \frac{1}{2} D_\varphi(\pi, e).$$

A similar scenario in which essentially the same argument as that of the proof of Lemma 30 can be used is as follows: the defining tree has one vertex with degree three and no vertices with degree larger than three (i.e., a tree with a Y shape), and for each cycle of π, there are two branches of the tree that do not contain two consecutive elements of c. It can then be shown that each such cycle can be decomposed into cycles that lie on paths in the defining tree, reducing the problem to the previously described one. An example is shown in Figure 6.

One may argue that the results of Lemma 30 and its extension to metric-trees have limited application, as they require that both the defining tree and the permutations/rankings used in the computation be of special form. In particular, one may require that a given ranking π is such that there are *no edges* between two different branches of Θ in the cycle graph of π. We show next that under certain conditions the probability of such permutations goes to zero as $n \to \infty$, by lower bounding the number P_n of permutations with the given constraint.

Let the set of vertices in the ith branch of a Y shaped defining tree Θ, $i = 1, 2, 3$, be denoted by B_i and let b_i denote the number of vertices in B_i. Clearly, $b_1 + b_2 + b_3 + 1 = n$.

Assume, without loss of generality, that the numbering of the branches is such that $b_1 \geq b_2 \geq b_3$. As an illustration, in
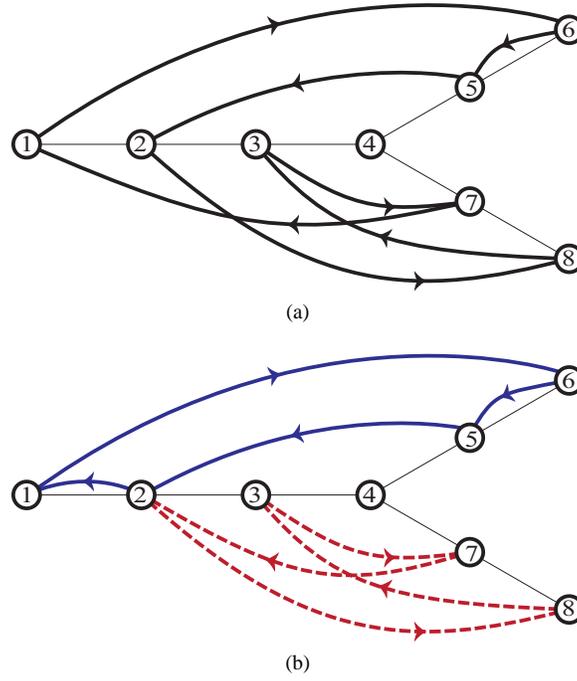
Figure 6: The cycle $(1\ 6\ 5\ 2\ 8\ 3\ 7)$ in Figure (a) is decomposed into two cycles, $(1\ 6\ 5\ 2)$ and $(2\ 8\ 3\ 7)$, as shown in Figure (b). Note that $(1\ 6\ 5\ 2\ 8\ 3\ 7) = (1\ 6\ 5\ 2)(2\ 8\ 3\ 7)$.

Figure 3b we have

$$
\begin{aligned}
B_1 &= \{1,2,3\}, & b_1 &= 3, \\
B_2 &= \{5,6\}, & b_2 &= 2, \\
B_3 &= \{7,8\}, & b_3 &= 2.
\end{aligned}
$$

The quantity P_n is greater than or equal to the number of permutations π whose cycle decomposition does not contain an edge between B_2 and B_3, and this quantity is, in turn, greater than or equal to the number of permutations π such that $\pi(j) \notin B_2 \cup B_3$ for $j \in B_2 \cup B_3$. The number of permutation with the latter property equals $\binom{b_1+1}{b_2+b_3}(b_2+b_3)!(b_1+1)!$. Hence,

$$
P_n \geq \frac{((b_1+1)!)^2}{(b_1+1-b_2-b_3)!},
$$

and thus

$$
\frac{P_n}{n!} \geq \frac{\prod_{j=n+1-2b_2-2b_3}^{n-b_2-b_3} j}{\prod_{j=n+1-b_2-b_3}^{n} j}.
$$

In particular, if $b_2 = b_3 = 1$, we have

$$
\frac{P_n}{n!} \geq \frac{(n-3)(n-2)}{(n-1)n} = 1 - \frac{4}{n} + O(n^{-2})
$$

and more generally, if $b_2 + b_3 = o(n)$, then

$$
\frac{P_n}{n!} \geq \frac{(n+o(n))^{b_2+b_3}}{(n+o(n))^{b_2+b_3}} \sim 1,
$$

or equivalently, $P_n \sim n!$.

Hence, if $b_2 + b_3 = o(n)$, the distance $\mathsf{d}_\varphi(\pi, e)$ of a randomly chosen permutation π from the identity equals $D_\varphi(\pi, e)/2$ with probability approaching 1 as $n \to \infty$.

It is worth noting that for metric-tree weight functions, the equality of Lemma 30 is not, in general, satisfied. To prove this claim, consider the metric-tree weight function φ in Figure 7, where, for $a, b \in [n], a < b$,

$$
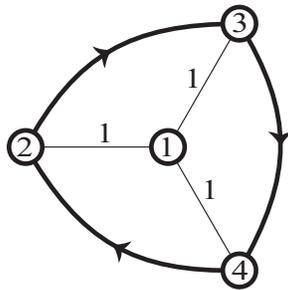\varphi_{(a\,b)} = \begin{cases} 1, & \text{if } a = 1, \\ \infty, & \text{if } a \neq 1. \end{cases}
$$

Figure 7: For the above metric-tree weight function and $\pi = (2\,3\,4)$, the equality of Lemma 30 does not hold.

It can be shown that for the permutation $\pi = (2\,3\,4)$, $\mathsf{d}_\varphi(\pi, e) = 4$, while $\frac{1}{2}D_\varphi(\pi, e) = 3$.

The following lemma provides a two approximation for transposition distances based on extended-path weight functions. As the weighted Kendall distance is a special case of the weighted transposition distance with extended-path weight functions, the lemma also implies Prop. 16.

Lemma 31. *For an extended-path weight function φ and for $\pi, \sigma \in \mathbb{S}_n$,*

$$\frac{1}{2}D_\varphi(\pi, \sigma) \le \mathsf{d}_\varphi(\pi, \sigma) \le D_\varphi(\pi, \sigma).$$

Proof: The lower bound follows from Lemma 27. To prove the upper bound, consider a metric-path weight function φ', with the same defining path Θ as φ, such that

$$\varphi'_{(a\ b)} = 2\varphi_{(a\ b)}$$

for any pair a, b adjacent in Θ. From Lemma 25, it follows that for distinct $c, d \in [n]$,

$$\mathsf{d}_\varphi((c\ d), e) \le 2p_\varphi^*(c, d) = p_{\varphi'}^*(c, d) = \mathsf{d}_{\varphi'}((c\ d), e).$$

Hence,

$$\mathsf{d}_\varphi(\pi, \sigma) \le \mathsf{d}_{\varphi'}(\pi, \sigma) = \frac{1}{2}D_{\varphi'}(\pi, \sigma) = D_\varphi(\pi, \sigma),$$

which proves the claimed result. ∎

V. AGGREGATION ALGORITHMS

Despite the importance of the rank aggregation problem in many areas of information retrieval, only a handful of results regarding the complexity of the problem are known. Among them, the most important results are the fact that finding a Kemeny optimal solution is NP-hard (see [8], [23] and references therein). Since the Kendall τ distance is a special case of the weighted Kendall distance, finding the aggregate ranking for the latter is also NP-hard. In particular, exhaustive search approaches – akin to the one we used in the previous sections – are not computationally feasible for large problems.

However, assuming that π^* is the solution to (1), the ranking σ_i closest to π^* provides a 2-approximation for the aggregate ranking. This easily follows from the fact that the Kendall τ distance satisfies the triangle inequality. As a result, one only has to evaluate the pairwise distances of the votes Σ in order to identify a 2-approximation aggregate for the problem. Assuming the weighted Kendall distance can be computed efficiently (for example, if the weight function is monotonic), the same is true of the weighted Kendall distance as it is also a metric and thus satisfies the triangle inequality.

A second method for obtaining a 2-approximation is an extension of a bipartite matching algorithm. For any distance function that may be written as

$$\mathsf{d}(\pi, \sigma) = \sum_{k=1}^{n} f(\pi^{-1}(k), \sigma^{-1}(k)), \tag{22}$$

where f denotes an arbitrary non-negative function, one can find an *exact solution* to (1) as described in the next section. The matching algorithm approach for classical Kendall τ aggregation was first proposed in [8].

A. Vote Aggregation Using Matching Algorithms

Consider a complete weighted bipartite graph $\mathcal{G} = (X, Y)$, with $X = \{1, 2, \cdots, n\}$ corresponding to the n ranks to be filled in, and $Y = \{1, 2, \cdots, n\}$ corresponding to the elements of $[n]$, i.e., the candidates. Let (i, j) denote an edge between $i \in X$

and $j \in Y$. We say that a perfect bipartite matching P corresponds to a permutation π whenever $(i, j) \in P$ if and only if $\pi(i) = j$. If the weight of (i, j) equals

$$\sum_{l=1}^{m} f(i, \sigma_l^{-1}(j)),$$

i.e., the weight incurred by $\pi(i) = j$, the minimum weight perfect matching corresponds to a solution of (1). The distance of (22) is a generalized version of Spearman's footrule since Spearman's footrule [7] can be obtained by choosing $f(x, y) = |x - y|$. Below, we explain how to use the matching approach for aggregation based on a general weighted Kendall distance. More details about this approach may be found in our companion conference paper [13].

Recall that for a weighted Kendall distance with weight function φ,

$$D_\varphi(\pi, \sigma) = \sum_{i=1}^{n} w(\pi^{-1}(i) : \sigma^{-1}(i)),$$

where

$$w(k : l) = \begin{cases} \sum_{h=k}^{l-1} \varphi_{(h\,h+1)}, & \text{if } k < l, \\ \sum_{h=l}^{k-1} \varphi_{(h\,h+1)}, & \text{if } k > l, \\ 0, & \text{if } k = l. \end{cases}$$

Note that D_φ is a distance measure of the form of (22), and thus a solution to problem (1) for $\mathsf{d} = D_\varphi$ can be found exactly in polynomial time.

Suppose that the set of votes is given by $\Sigma = \{\sigma_1, \cdots, \sigma_m\}$.

Proposition 32. *Let $\pi' = \arg\min_\pi \sum_{l=1}^{m} D_\varphi(\pi, \sigma_i)$ and $\pi^* = \arg\min_\pi \sum_{l=1}^{m} \mathsf{d}_\varphi(\pi, \sigma_i)$. The permutation π' is a 2-approximation to the optimal rank aggregate π^* if φ corresponds to a weighted Kendall distance.*

Proof: From Prop. 16, for a weighted Kendall weight function φ and for permutations π and σ,

$$\frac{1}{2} D_\varphi(\pi, \sigma) \leq \mathsf{d}_\varphi(\pi, \sigma) \leq D_\varphi(\pi, \sigma).$$

Thus we have

$$\sum_{l=1}^{m} \mathsf{d}_\varphi(\pi', \sigma_i) \leq \sum_{l=1}^{m} D_\varphi(\pi', \sigma_i).$$

and

$$\frac{1}{2} \sum_{l=1}^{m} D_\varphi(\pi^*, \sigma_i) \leq \sum_{l=1}^{m} \mathsf{d}_\varphi(\pi^*, \sigma_i)$$

From the optimality of π' with respect to D, we find

$$\sum_{l=1}^{m} D_\varphi(\pi', \sigma_i) \leq \sum_{l=1}^{m} D_\varphi(\pi^*, \sigma_i).$$

Hence

$$\sum_{l=1}^{m} \mathsf{d}_\varphi(\pi', \sigma_i) \leq 2 \sum_{l=1}^{m} \mathsf{d}_\varphi(\pi^*, \sigma_i).$$

∎

In fact, the above proposition applies to the larger class of weighted transposition distances with extended-path weights. It can similarly be shown that for a weighted transposition distance with general weights (resp. metric weights), π' is a 4-approximation (resp. a 2-approximation). Finally, for a weighted transposition distance with metric-path weights, π' represents the exact solution.

A simple approach for improving the performance of the matching based algorithm is to couple it with a local descent method. Assume that an estimate of the aggregate at step ℓ equals $\pi^{(\ell)}$. As before, let $\mathbb{A}_n$ be the set of adjacent transpositions in $\mathbb{S}_n$. Then

$$\pi^{(\ell+1)} = \pi^{(\ell)} \arg\min_{\tau \in \mathbb{A}_n} \sum_{i=1}^{m} \mathsf{d}(\pi^{(\ell)} \tau, \sigma_i).$$

The search terminates when the cumulative distance of the aggregate from the set of votes Σ cannot be decreased further. We choose the starting point $\pi^{(0)}$ to be the ranking π' of Prop. 32 obtained by the minimum weight bipartite matching algorithm. This method will henceforth be referred to as Bipartite Matching with Local Search (BMLS).

An important question at this point is how does the approximate nature of the BMLS aggregation process change the aggregate, especially with respect to the top-vs-bottom or similarity property? This question is hard, and we currently have

no mathematical results pertaining to this problem. Instead, we describe a number of simulation results that may guide future analysis of this issue.

In order to see the effect of the BMLS on vote aggregation, we revisit Examples 17-20. In *all except for one case* the solution provided by BMLS is the same as the exact solution, both for the Kendall τ and weighted Kendall distances.

The exception is Example 18. In this case, for the weight function $\varphi_{(i\,i+1)} = (2/3)^{i-1}, i \in [3]$, the exact solution equals $(1, 4, 2, 3)$ but the solution obtained via BMLS equals $(4, 2, 3, 1)$. Note that these two solutions differ significantly in terms of their placement of candidate 1, ranked first in the exact ranking and last in the approximate ranking. The distances between the two solutions, $\mathsf{d}_\varphi((1, 4, 2, 3), (4, 2, 3, 1))$, equals 2.11 and is rather large. Nevertheless, the cumulative distances to the votes are very close in value:

$$\sum_i \mathsf{d}_\varphi((1, 4, 2, 3), \sigma_i) = 9,$$

$$\sum_i \mathsf{d}_\varphi((4, 2, 3, 1), \sigma_i) = 9.11.$$

Hence, as with any other distance based approach, the approximation result may sometimes diverge significantly from the optimum solution while the closeness of the approximate solution to the set of votes is nearly the same as that of the optimum solution. One way to avoid such approximation errors is to use weight functions with sufficiently large "spreads" of weights for which the difference between solutions has to be smaller than a given threshold. This topic will be discussed elsewhere.

B. Vote Aggregation Using PageRank

An algorithm for data fusion based on the PageRank and HITS algorithms for ranking web pages was proposed in [8], [23]. PageRank is one of the most important algorithms developed for search engines used by Google, with the aim of scoring web-pages based on their relevance. Each webpage that has hyperlinks to other webpages is considered a voter, while the voter's preferences for candidates is expressed via the hyperlinks. When a hyperlink to a webpage is not present, it is assumed that the voter does not support the given candidate's webpage. Although the exact implementation details of PageRank are not known, it is widely assumed that the graph of webpages is endowed with near-uniform transition probabilities. The ranking of the webpages is obtained by computing the stationary probabilities of the chain, and ordering the pages according to the values of the stationary probabilities. The connectivity of the Markov chain provides information about pairwise candidate preferences, and states with high input probability correspond to candidates ranked highly in a large number of lists.

This idea can be easily adapted to the rank aggregation problem with weighted distances in several different settings. In such an adaptation, the states of a Markov chain correspond to the candidates and the transition probabilities are functions of the votes. Dwork et al. [8], [9] proposed four different ways for computing the transition probabilities from the votes. Below, we describe the method that is most suitable for our problem and provide a generalization of the algorithm for weighted distance aggregation.

Consider a Markov chain with states indexed by the candidates. Let P denote the transition probability matrix of the Markov chain, with P_{ij} denoting the probability of going from state (candidate) i to state j. In [8], the transition probabilities are evaluated as

$$P_{ij} = \frac{1}{m} \sum_{\sigma \in \Sigma} P_{ij}(\sigma),$$

where

$$P_{ij}(\sigma) = \begin{cases} \frac{1}{n}, & \text{if } \sigma^{-1}(j) < \sigma^{-1}(i), \\ 1 - \frac{\sigma^{-1}(i)-1}{n}, & \text{if } i = j, \\ 0, & \text{if } \sigma^{-1}(j) > \sigma^{-1}(i). \end{cases}$$

Our Markov chain model for weighted Kendall distance is similar, with a modification that includes incorporating transposition weights into the transition probabilities. To accomplish this task, we proceed as follows.

Let $w_k = \varphi_{(k\,k+1)}$, and let $i_\sigma = \sigma^{-1}(i)$ for candidate $i \in [n]$. We set

$$\beta_{ij}(\sigma) = \max_{l:j_\sigma \leq l < i_\sigma} \frac{\sum_{h=l}^{i_\sigma - 1} w_h}{i_\sigma - l} \tag{23}$$

if $j_\sigma < i_\sigma$, $\beta_{ij}(\sigma) = 0$ if $j_\sigma > i_\sigma$, and

$$\beta_{ii}(\sigma) = \sum_{k:k_\sigma > i_\sigma} \beta_{ki}(\sigma).$$

The transition probabilities equal

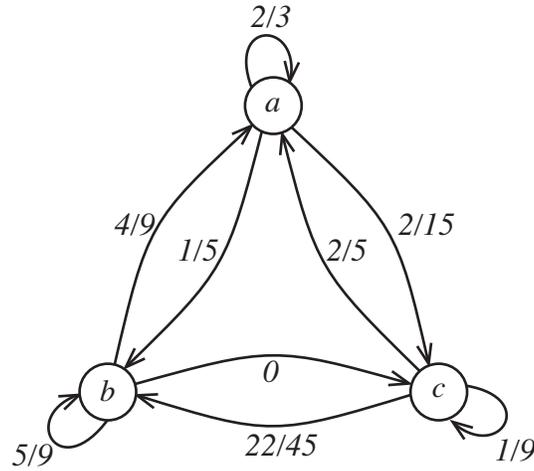$$P_{ij} = \frac{1}{m} \sum_{k=1}^m P_{ij}(\sigma_k),$$

Figure 8: The Markov chain for Example 33.

with

$$P_{ij}(\sigma) = \frac{\beta_{ij}(\sigma)}{\sum_k \beta_{ik}(\sigma)}.$$

Intuitively, the transition probabilities described above may be interpreted as follows. The transition probabilities are obtained by averaging the transition probabilities corresponding to individual votes $\sigma \in \Sigma$. For each vote σ, consider candidates j and k with $j_\sigma = i_\sigma - 1$ and $k_\sigma = i_\sigma - 2$. The probability of going from candidate i to candidate j is proportional to $w_{j_\sigma} = \varphi_{(j_\sigma\, i_\sigma)}$. This implies that if $w_{j_\sigma} > 0$, one moves from candidate i to candidate j with positive probability. Furthermore, larger values for w_{j_σ} result in higher probabilities for moving from i to j.

In the case of candidate k, it seems reasonable to let the probability of transitioning from candidate i to candidate k be proportional to $\frac{w_{j_\sigma} + w_{k_\sigma}}{2}$. However, since k is ranked before j by vote σ, it is natural to require that the probability of moving to candidate k from candidate i be at least as high as the probability of moving to candidate j from candidate i. This reasoning leads to $\beta_{ik} = \max\{w_{j_\sigma}, \frac{w_{j_\sigma} + w_{k_\sigma}}{2}\}$ and motivates using the maximum in (23). Finally, the probability of staying with candidate i is proportional to the sum of the β's from candidates placed below candidate i.

Example 33. Let the votes in Σ consist of $\sigma_1 = (a, b, c)$, $\sigma_2 = (a, b, c)$, and $\sigma_3 = (b, c, a)$, and let $w = (w_1, w_2) = (2, 1)$.

Consider the vote $\sigma_1 = (a, b, c)$. We have $\beta_{ba}(\sigma_1) = \frac{w_1}{1} = 2$. Note that if w_1 is large, then β_{ba} is large as well.

In addition, $\beta_{cb}(\sigma_1) = \frac{w_2}{1} = 1$ and

$$\beta_{ca} = \max\left\{\frac{w_1 + w_2}{2}, \beta_{cb}\right\} = \frac{3}{2}.$$

The purpose of the $\max$ function is to ensure that $\beta_{ca} \geq \beta_{cb}$, which is a natural requirement given that a is ranked before b according to σ_1.

Finally, $\beta_{aa}(\sigma_1) = \beta_{ca}(\sigma_1) + \beta_{ba}(\sigma_1) = 2 + \frac{3}{2} = \frac{7}{2}$ and $\beta_{bb}(\sigma_1) = \beta_{cb}(\sigma_1) = 1$. Note that according to the transition probability model, one also has $\beta_{aa} \geq \beta_{bb}$. This may again be justified by the fact that σ_1 places a higher than b.

Since $\sigma_1 = \sigma_2$, we have

$$P(\sigma_1) = P(\sigma_2) = \begin{pmatrix} 1 & 0 & 0 \\ 2/3 & 1/3 & 0 \\ 3/5 & 2/5 & 0 \end{pmatrix}.$$

Similar computations yield

$$\beta_{cb}(\sigma_3) = 2, \qquad \beta_{ac}(\sigma_3) = 1, \qquad \beta_{ab}(\sigma_3) = \frac{3}{2}$$
$$\beta_{aa}(\sigma_3) = 0, \quad \beta_{bb}(\sigma_3) = 2 + \frac{3}{2} = \frac{7}{2}, \quad \beta_{cc}(\sigma_3) = 1,$$

and thus

$$P(\sigma_3) = \begin{pmatrix} 0 & 3/5 & 2/5 \\ 0 & 1 & 0 \\ 0 & 2/3 & 1/3 \end{pmatrix}.$$

From the $P(\sigma_1), P(\sigma_2),$ and $P(\sigma_3)$, we obtain

Method	Aggregate ranking and average distance			
	$w = (1,0,0,0)$	$w = (1,1,1,1)$	$w = (1,1,0,0)$	$w = (0,1,0,0)$
OPT	$(\underline{1},4,3,2,5)$, 0.7273	$(2,3,4,5,1)$, 2.3636	$\left(\underline{2,3},4,5,1\right)$, 1.455	$\left(\underline{3,2},5,4,1\right)$, 0.636
BMLS	$(\underline{1},2,3,4,5)$, 0.7273	$(2,3,4,5,1)$, 2.3636	$\left(\underline{2,3},1,5,4\right)$, 1.455	$\left(\underline{2,3},1,5,4\right)$, 0.636
MC	$(\underline{1},2,5,4,3)$, 0.7273	$(2,3,4,5,1)$, 2.3636	$\left(\underline{2,1},3,4,5\right)$, 1.546	$\left(\underline{2,3},1,4,5\right)$, 0.636

Table I: The aggregate rankings and the average distance of the aggregate ranking from the votes for different weight functions w.

$$P = \frac{P(\sigma_1) + P(\sigma_2) + P(\sigma_3)}{3} = \begin{pmatrix} 2/3 & 1/5 & 2/15 \\ 4/9 & 5/9 & 0 \\ 2/5 & 22/45 & 1/9 \end{pmatrix}.$$

The Markov chain corresponding to P is given in Figure 8. The stationary distribution of this Markov chain is $(0.56657, 0.34844, 0.084986)$ which corresponds to the ranking (a, b, c).

Example 34. The performance of the Markov chain approach described above cannot be easily evaluated analytically, as is the case with any related aggregation algorithm proposed so far.

We hence test the performance of the scheme on examples for which the optimal solutions are easy to evaluate numerically. For this purpose, in what follows, we consider a simple test example, with $m = 11$. The set of votes (rankings) is listed below

$$\Sigma^T = \begin{pmatrix} 1 & 1 & 1 & 2 & 2 & 3 & 3 & 4 & 4 & 5 & 5 \\ 2 & 2 & 2 & 3 & 3 & 2 & 2 & 2 & 2 & 2 & 2 \\ 3 & 3 & 3 & 4 & 4 & 4 & 4 & 5 & 5 & 3 & 3 \\ 4 & 4 & 4 & 5 & 5 & 5 & 5 & 3 & 3 & 4 & 4 \\ 5 & 5 & 5 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

Note that due to the transpose operator, each column corresponds to a vote, e.g., $\sigma_1 = (1, 2, 3, 4, 5)$.

Let us consider candidates 1 and 2. Using the majority rule, one would arrive at the conclusion that candidate 1 should be the winner, given that 1 appears most often at the top of the list. Under a number of other aggregation rules, including Kemeny's rule and Borda's method, candidate 2 would be the winner.

Our immediate goal is to see how different weighted distance based rank aggregation algorithms would position candidates 1 and 2. The numerical results regarding this example are presented in Table I. In the table, OPT refers to an optimal solution which was found by exhaustive search, and MC refers to the Markov chain method.

If the weight function is $w = (w_1, \cdots, w_4) = (1, 0, 0, 0)$, where $w_i = \varphi_{(i\ i+1)}$, the optimal aggregate vote clearly corresponds to the plurality winner. That is, the winner is the candidate with most voters ranking him/her as the top candidate. A quick check of Table I reveals that all three methods identify the winner correctly. Note that the ranks of candidates other than candidate 1 obtained by the different methods are different, however this does not affect the distance between the aggregate ranking and the votes.

The next weight function that we consider is the uniform weight function, $w = (1, 1, 1, 1)$. This weight function corresponds to the conventional Kendall τ distance. As shown in Table I, all three methods produce $(2, 3, 4, 5, 1)$, and the aggregates returned by BMLS and MC are optimum.

The weight function $w = (1, 1, 0, 0)$ corresponds to *ranking of the top 2* candidates. OPT and BMLS return $2, 3$ as the top two candidates, both preferring 2 to 3. The MC method, however, returns $2, 1$ as the top two candidates, with a preference for 2 over 1, and a suboptimal cumulative distance. It should be noted that this may be attributed to the fact the the MC method is not designed to only minimize the average distance: another important factor in determining the winners via the MC method is that winning against strong candidates "makes one strong". In this example, candidate 1 beats the strongest candidate, candidate 2, three times, while candidate 3 beats candidate 2 only twice and this fact seems to be the reason for the MC algorithm to prefer candidate 1 to candidate 3. Nevertheless, the stationary probabilities of candidates 1 and 3 obtained by the MC method are very close to each other, as the vector of probabilities is $(\underline{0.137}, 0.555, \underline{0.132}, 0.0883, 0.0877)$.

The weight function $w = (0, 1, 0, 0)$ corresponds to *identifying the top 2* candidates – i.e., it is not important which candidate is the first and which is the second. The OPT and BMLS identify $\{2, 3\}$ as the top two candidates.

The MC method returns the stationary probabilities $(0, 1, 0, 0, 0)$ which means that candidate 2 is an absorbing state in the Markov chain. This occurs because candidate 2 is ranked first or second by all voters. The existence of absorbing states is a drawback of the Markov chain methods. One solution is to remove 2 from the votes and re-apply MC. The MC method in this case results in the stationary distribution $(p(1), p(3), p(4), p(5)) = (0.273, 0.364, 0.182, 0.182)$, which gives us the ranking $(3, 1, 4, 5)$. Together with the fact that candidate 2 is the strongest candidate, we obtain the ranking $(2, 3, 1, 4, 5)$.

VI. Appendix

A. Computing the Weight Functions with Two Identical Non-zero Weights

The goal is to find the weighted Kendall distance $d_\varphi(\pi, e)$ with the weight function of (12), for an arbitrary $\pi \in \mathbb{S}_n$. For this purpose, let $R_1 = \{1, \cdots, a\}$, $R_2 = \{a+1, \cdots, b\}$, and $R_3 = \{b+1, \cdots, n\}$, and define

$$N_{ij}^\pi = |\{k \in R_j : \pi^{-1}(k) \in R_i\}|, \quad i, j \in \{1, 2, 3\}.$$

That is, N_{ij}^π is the number of elements whose ranks in π belong to the set R_i and whose ranks in e belong to the set R_j. A sequence of transpositions that transforms π into e moves the N_{ij}^π elements of $\{k \in R_j : \pi^{-1}(k) \in R_i\}$ from R_i to R_j. Furthermore, note that any transposition that swaps two elements with ranks in the same region $R_i, i \in [3]$, has weight zero, while for any transposition τ_l that swaps an element ranked in R_1 with an element ranked in R_2 or swaps an element ranked in R_2 with an element ranked in R_3, we have $d_\varphi(\tau_l, e) = 1$.

It is straightforward to see that $\sum_j N_{ij}^\pi = \sum_j N_{ji}^\pi$. In particular, $N_{12}^\pi + N_{13}^\pi = N_{21}^\pi + N_{31}^\pi$ and $N_{31}^\pi + N_{32}^\pi = N_{13}^\pi + N_{23}^\pi$. We show next that

$$d_\varphi(\pi, e) = \begin{cases} 2N_{13}^\pi + N_{12}^\pi + N_{23}^\pi, & \text{if } N_{21}^\pi \geq 1 \text{ or } N_{23}^\pi \geq 1, \\ 2N_{13}^\pi + 1, & \text{if } N_{21}^\pi = N_{23}^\pi = 0. \end{cases}$$

Note that, from Prop. 16, we have

$$d_\varphi(\pi, e) \geq \frac{1}{2} D_\varphi(\pi, e) = 2N_{13}^\pi + N_{12}^\pi + N_{23}^\pi. \tag{24}$$

Suppose that $N_{21}^\pi \geq 1$ or $N_{23}^\pi \geq 1$. We find a transposition τ_l, with $d_\varphi(\tau_l, e) = 1$, such that $\pi' = \pi\tau_l$ satisfies $D_\varphi(\pi', e) = D_\varphi(\pi, e) - 2$, and at least one of the following conditions:

$$\begin{cases} N_{21}^{\pi'} \geq 1, \\ \text{or} \\ N_{23}^{\pi'} \geq 1, \\ \text{or} \\ \pi' = e. \end{cases} \tag{25}$$

Applying the same argument repeatedly, and using the triangle inequality, proves that $d_\varphi(\pi, e) \leq \frac{1}{2} D_\varphi(\pi, e)$ if $N_{21}^\pi \geq 1$ or $N_{23}^\pi \geq 1$. This, along with (24), shows that $d_\varphi(\pi, e) = \frac{1}{2} D_\varphi(\pi, e)$ if $N_{21}^\pi \geq 1$ or $N_{23}^\pi \geq 1$.

First, suppose that $N_{21}^\pi \geq 1$ and $N_{23}^\pi \geq 1$. It then follows that $N_{12}^\pi \geq 1$ or $N_{32}^\pi \geq 1$. Without loss of generality, assume that $N_{12}^\pi \geq 1$. Then τ_l can be chosen such that $N_{12}^{\pi'} = N_{12}^\pi - 1$ and $N_{21}^{\pi'} = N_{21}^\pi - 1$. We have $D_\varphi(\pi', e) = D_\varphi(\pi, e) - 2$, and since $N_{23}^\pi \geq 1$, condition (25) holds.

Next, suppose $N_{21}^\pi \geq 1$ and $N_{23}^\pi = 0$. If $N_{13}^\pi \geq 1$, choose τ_l such that

$$N_{21}^{\pi'} = N_{21}^\pi - 1,$$
$$N_{23}^{\pi'} = 1,$$
$$N_{13}^{\pi'} = N_{13}^\pi - 1,$$

where $\pi' = \pi\tau_l$. Since $N_{23}^{\pi'} = 1$, condition (25) is satisfied. If $N_{13}^\pi = 0$, then $N_{31}^\pi = N_{32}^\pi = 0$, and thus $N_{12}^\pi = N_{21}^\pi \geq 1$. In this case, we choose τ_l such that $N_{21}^{\pi'} = N_{12}^{\pi'} = N_{12}^\pi - 1$. As a result, we have either $N_{21}^{\pi'} \geq 1$ or $\pi' = e$. Hence, condition (25) is satisfied once again. Note that in both cases, for $N_{13}^\pi = 0$ as well as for $N_{13}^\pi \geq 1$, we have $D_\varphi(\pi', e) = D_\varphi(\pi, e) - 2$.

The proof for the case $N_{23}^\pi \geq 1$ and $N_{21}^\pi = 0$ follows along similar lines.

If $N_{21}^\pi = N_{23}^\pi = 0$, it can be verified by inspection that for every transposition τ_l with $d_\varphi(\tau_l, e) = 1$, we have $D_\varphi(\pi\tau_l, e) \geq D_\varphi(\pi, e)$. Hence, the inequality in (24) cannot be satisfied with equality, which implies that $d_\varphi(\pi, e) \geq 2N_{13}^\pi + 1$. Choose a transposition τ_l with $d_\varphi(\tau_l, e) = 1$ such that

$$N_{13}^{\pi'} = N_{13}^\pi - 1,$$
$$N_{12}^{\pi'} = 1,$$
$$N_{23}^{\pi'} = 1.$$

where $\pi' = \pi\tau_l$. We have

$$d_\varphi(\pi, e) \leq d_\varphi(\tau_l, e) + d_\varphi(\pi', e) = 1 + 2N_{13}^\pi.$$

This, along with $d_\varphi(\pi, e) \geq 2N_{13}^\pi + 1$, completes the proof.

B. The Average Kendall and Weighted Kendall Distance

The Kendall τ distance between two rankings may be viewed in the following way: each pair of candidates on which the two rankings disagree contribute one unit to the distance between the rankings. Owing to Algorithm 1, the weighted Kendall distance with a decreasing weight function can be regarded in a similar manner: each pair of candidates on which the two rankings disagree contributes $\varphi_{(s\ s+1)}$, for some s, to the distance between the rankings.

Consider a pair a and b such that $\pi^{-1}(b) < \pi^{-1}(a)$ and $\sigma^{-1}(a) < \sigma^{-1}(b)$. In Algorithm 1, there exists a transposition $\tau_t^\star = (s\ s+1)$ that swaps a and b where

$$s = \pi^{-1}(b) + \left|\left\{k : \sigma^{-1}(k) < \sigma^{-1}(a), \pi^{-1}(k) > \pi^{-1}(b)\right\}\right|,$$

that is, s equals $\pi^{-1}(b)$ plus the number of elements that appear before a in σ and after b in π. It is not hard to see that s can also be written in a way that is symmetric with respect to π and σ, as

$$s = \pi^{-1}(b) + \sigma^{-1}(a) - \left|\left\{k : \pi^{-1}(k) < \pi^{-1}(b), \sigma^{-1}(k) < \sigma^{-1}(a)\right\}\right| - 1$$
$$= n - 1 - \left|\left\{k : \pi^{-1}(k) > \pi^{-1}(b), \sigma^{-1}(k) > \sigma^{-1}(a)\right\}\right|.$$

As an example, consider $\varphi_{(i\ i+1)} = n - i$. Then,

$$\mathsf{d}_\varphi(\pi, \sigma) = \sum_{(b,a)\in\mathscr{I}(\pi,\sigma)} \left(1 + \left|\left\{k : \pi^{-1}(k) > \pi^{-1}(b), \sigma^{-1}(k) > \sigma^{-1}(a)\right\}\right|\right)$$

$$= K(\pi, \sigma) + \sum_{(b,a)\in\mathscr{I}(\pi,\sigma)} \left|\left\{k : \pi^{-1}(k) > \pi^{-1}(b), \sigma^{-1}(k) > \sigma^{-1}(a)\right\}\right|$$

where $\mathscr{I}(\pi,\sigma)$ is the set of ordered pairs (b,a) such that $\pi^{-1}(b) < \pi^{-1}(a)$ and $\sigma^{-1}(a) < \sigma^{-1}(b)$. Note that the weighted Kendall distance d_φ equals the Kendall τ distance plus a sum that captures the influence of assigning higher importance to the top positions of the rankings.

These observations allow us to easily compute the expected value of the distance between the identity permutation and a randomly and uniformly chosen permutation $\pi \in \mathbb{S}_n$. For $1 \le a < b \le n$ and $s \in [n-1]$, let X_{ab}^s be an indicator variable that equals one if and only if $\pi^{-1}(a) > \pi^{-1}(b)$ and

$$\left|\left\{k > a : \pi^{-1}(k) > \pi^{-1}(b)\right\}\right| = n - 1 - s.$$

The expected distance between the two permutations equals

$$E[\mathsf{d}_\varphi(\pi, e)] = \sum_{s=1}^{n-1} \varphi_{(s\ s+1)} \sum_{a=1}^{n-1} \sum_{b=a+1}^{n} E\left[X_{ab}^s\right]. \tag{26}$$

By the definition of X_{ab}^s, $E\left[X_{ab}^s\right]$ equals the probability of the event that $n - 1 - s$ elements of $\{a+1,\cdots,n\}\setminus\{b\}$ and a appear after b in π. There are $\binom{n-a-1}{n-s-1}$ ways to choose $n - s - 1$ elements from $\{a+1,\cdots,n\}\setminus\{b\}$, $\binom{n}{a-1}(a-1)!$ ways to assign positions to the elements of $\{1,2,\cdots,a-1\}$, $(s-a)!$ ways to arrange the $s - a$ elements of $\{a+1,\cdots,n\}\setminus\{b\}$ that appear before b, and $(n-s)!$ ways to arrange a and the $n - 1 - s$ elements of $\{a+1,\cdots,n\}\setminus\{b\}$ that appear after b. Hence,

$$E\left[X_{ab}^s\right] = \frac{1}{n!}\binom{n-a-1}{n-s-1}\binom{n}{a-1}(a-1)!(s-a)!(n-s)!$$
$$= \frac{n-s}{(n-a+1)(n-a)},$$

for $1 \le a \le s$, and $E\left[X_{ab}^s\right] = 0$ for $a > s$. Using this expression in (26), we obtain

$$E[\mathsf{d}_\varphi(e, \pi)] = \sum_{s=1}^{n-1} \varphi_{(s\ s+1)} \sum_{a=1}^{s} \frac{n-s}{n-a+1}$$
$$= \sum_{s=1}^{n-1} \varphi_{(s\ s+1)}(n-s)(H_n - H_{n-s}),$$

where $H_i = \sum_{l=1}^{i} \frac{1}{l}$. Indeed, for $\varphi_{(s\ s+1)} = 1, s \in [n-1]$, we recover the well known result that

$$E[\mathsf{d}_\varphi(e, \pi)] = \sum_{s=1}^{n-1}(n-s)(H_n - H_{n-s})$$
$$= \sum_{k=1}^{n-1} k(H_n - H_k)$$
$$= \frac{1}{2}\binom{n}{2}.$$

For $\varphi_{(s\ s+1)} = n - s$, the average distance equals

$$E[\mathsf{d}_\varphi(e, \pi)] = \sum_{s=1}^{n-1} (n-s)^2 (H_n - H_{n-s})$$
$$= \sum_{k=1}^{n-1} k^2 (H_n - H_k)$$
$$= \frac{1}{2}\binom{n}{2} + \frac{2}{3}\binom{n}{3}.$$

REFERENCES

[1] K. J. Arrow. *Social choice and individual values*. Yale Univ Pr, 1963.

[2] W. D. Cook and M. Kress. Ordinal ranking with intensity of preference. *Management Science*, 31(1):26–32, 01 1985.

[3] T. Cormen, C. Leiserson, R. Rivest, and C. Stein. *Introduction to algorithms*. MIT, Cambridge, MA, 2001.

[4] J.-C. de Borda. Mémoire sur les élections au scrutin. *Histoire de l'Académie royale des sciences*, 1784.

[5] M. de Condorcet. *Paris: Imprimerie royale*, (1):14–37, Feb.

[6] M. Deza and T. Huang. Metrics on permutations, a survey. *J. Comb. Inf. Sys. Sci*, 23:173–185, 1998.

[7] P. Diaconis. Group representations in probability and statistics. *Lecture Notes-Monograph Series*, 11, 1988.

[8] C. Dwork, R. Kumar, M. Naor, and D. Sivakumar. Rank aggregation methods for the web. In *Proceedings of the 10th international conference on World Wide Web*, pages 613–622. ACM, 2001.

[9] C. Dwork, R. Kumar, M. Naor, and D. Sivakumar. Rank aggregation revisited. Manuscript, Available: www.eecs.harvard.edu/~michaelm/CS222/rank2.pdf, 2001.

[10] Economist Intelligence Unit. A Summary of the Liveability Ranking and Overview [white paper]. http://www.eiu.com/site_info.asp?info_name=The_Global_Liveability_Report, Aug. 2012.

[11] M. Farah and D. Vanderpoten. An outranking approach for rank aggregation in information retrieval. *SIGIR̃07*, pages 14–37, July 2007.

[12] F. Farnoud and O. Milenkovic. Sorting of permutations by cost-constrained transpositions. *IEEE Trans. Information Theory*, 58(1):3 –23, Jan. 2012.

[13] F. Farnoud, B. Touri, and O. Milenkovic. Nonuniform vote aggregation algorithms. In *Int. Conf. Signal Processing and Communications*, India, July 2012.

[14] M. L. Fredman and R. E. Tarjan. Fibonacci heaps and their uses in improved network optimization algorithms. *J. ACM*, 34(3):596–615, July 1987.

[15] J. Hodge and R. Klima. *The mathematics of voting and elections: a hands-on approach*, volume 22 of *Mathematical World*. American Mathematical Society, 2005.

[16] A. Jiang, R. Mateescu, M. Schwartz, and J. Bruck. Rank modulation for flash memories. *Information Theory, IEEE Transactions on*, 55(6):2659 –2673, June 2009.

[17] J. G. Kemeny. Mathematics without numbers. *Daedalus*, 88(4):pp. 577–591, 1959.

[18] J. G. Kemeny and J. Snell. *Mathematical models in the social sciences*. Ginn, Boston, 1962.

[19] M. Kendall. *Rank correlation methods*. London: Griffin, 4th edition, 1970.

[20] R. Kumar and S. Vassilvitskii. Generalized distances between rankings. In *Proceedings of the 19th international conference on World wide web*, WWW '10, pages 571–580, New York, NY, USA, 2010. ACM.

[21] A. Mazumdar, A. Barg, and G. Zemor. Constructions of rank modulation codes. In *IEEE International Symposium on Information Theory*, pages 869 –873, July/Aug. 2011.

[22] Optify Inc. The Changing Face of SERPs: Organic Click Through Rate [white paper]. http://www.optify.net/inbound-marketing-resources/new-study-how-the-new-face-of 2012.

[23] V. Popov. Multiple genome rearrangement by swaps and by element duplications. *Theoretical Computer Science*, 385(1-3):115–126, 2007.

[24] F. Schalekamp and A. van Zuylen. Rank aggregation: Together we're strong. *Proc. of 11th ALENEX*, pages 38–51, 2009.

[25] D. Sculley. Rank aggregation for similar items. In *Proceedings of the Seventh SIAM International Conference on Data Mining*, 2007.