SIAM J. DISCRETE MATH.
Vol. 20, No. 3, pp. 628–648
© 2006 Society for Industrial and Applied Mathematics

COMPARING PARTIAL RANKINGS*

RONALD FAGIN†, RAVI KUMAR‡, MOHAMMAD MAHDIAN§, D. SIVAKUMAR¶, AND ERIK VEE†

Abstract. We provide a comprehensive picture of how to compare *partial rankings*, that is, rankings that allow ties. We propose several metrics to compare partial rankings and prove that they are within constant multiples of each other.

Key words. partial ranking, bucket order, permutation, metric

AMS subject classifications. 06A06, 68R99

DOI. 10.1137/05063088X

1. Introduction.

The study of metrics on permutations (i.e., full rankings) is classical and several well-studied metrics are known [10, 22], including the Kendall tau distance and the Spearman footrule distance. The rankings encountered in practice, however, often have ties (hence the name *partial rankings*), and metrics on such rankings are much less studied.

Aside from its purely mathematical interest, the problem of defining metrics on partial rankings is valuable in a number of applications. For example the *rank aggregation* problem for partial rankings arises naturally in multiple settings, including in online commerce, where users state their preferences for products according to various criteria, and the system ranks the products in a single, cohesive way that incorporates all the stated preferences, and returns the top few items to the user. Specific instances include the following: selecting a restaurant from a database of restaurants (where the ranking criteria include culinary preference, driving distance, star rating, etc.), selecting an air-travel plan (where the ranking criteria include price, airline preference, number of hops, etc.), and searching for articles in a scientific bibliography (where the articles may be ranked by relevance of subject, year, number of citations, etc.). In all of these scenarios, it is easy to see that many of the ranking criteria lead to ties among the underlying set of items. To formulate a mathematically sound aggregation problem for such partially ranked lists (as has been done successfully for fully ranked lists [12] and "top k lists" [16]), it is sometimes necessary to have a well-defined distance measure (preferably a metric) between partial rankings.

In this paper we focus on four metrics between partial rankings. These are obtained by suitably generalizing the Kendall tau distance and the Spearman footrule distance on permutations in two different ways. In the first approach, we associate

*Received by the editors May 6, 2005; accepted for publication (in revised form) February 7, 2006; published electronically September 5, 2006. This paper is an expansion of a portion of the paper [14].

http://www.siam.org/journals/sidma/20-3/63088.html

†IBM Almaden Research Center, 650 Harry Road, San Jose, CA 95120 (fagin@almaden.ibm.com, vee@almaden.ibm.com).

‡Yahoo! Research, 701 First Ave., Sunnyvale, CA 94089 (ravikumar@yahoo-inc.com). This author's work was done at the IBM Almaden Research Center.

§Microsoft Research, Redmond, WA 98052 (mahdian@microsoft.com). Part of this author's work was supported by NSF grant CCR-0098066. Part of this work was done while the author was visiting the IBM Almaden Research Center.

¶Google, Inc., 1600 Amphitheatre Parkway, Mountain View, CA 94043 (siva@google.com). This author's work was done at the IBM Almaden Research Center.

with each partial ranking a "profile vector" and we define the distance between the partial rankings to be the L_1 distance between the corresponding profile vectors. In the second approach, we associate with each partial ranking the family of all full rankings that are obtained by breaking ties in all possible ways. The distance between partial rankings is then taken to be the Hausdorff distance between the corresponding sets of full rankings.[1] In addition to the four metrics we obtain by extending the Kendall tau distance and the Spearman footrule distance using these two approaches, we consider a method obtained by generalizing the Kendall tau distance where we vary a certain parameter. For some choices of the parameter, we obtain a metric, and for one natural choice, we obtain our Kendall profile metric. All the metrics we define admit efficient computation. These metrics are defined and discussed in section 3.

Having various metrics on partial rankings is good news, but exactly which one should a practitioner use to compare partial rankings? Furthermore, which one is best suited for formulating an aggregation problem for partial rankings? Our summary answer to these questions is that the exact choice does not matter much. Namely, following the lead of [16], we define two metrics to be *equivalent* if they are within constant multiples of each other. This notion was inspired by the Diaconis–Graham inequality [11], which says that the Kendall tau distance and the Spearman footrule distance are within a factor of two of each other. Our main theorem says that all of our metrics are equivalent in this sense. The methods where we generalize the Kendall tau distance by varying a certain parameter are easily shown to be equivalent to each other, and in particular to the profile version of the Kendall tau distance (since one choice of the parameter leads to the profile version). It is also simple to show that the Hausdorff versions of the Kendall tau distance and the Spearman footrule distance are equivalent and that the Hausdorff and the profile versions of the Kendall tau metric are equivalent. Proving equivalence for the profile metrics turns out to be rather tricky and requires us to uncover considerable structure inside partial rankings. We present these equivalence results in section 4.

Related work. The Hausdorff versions of the Kendall tau distance and the Spearman footrule distance are due to Critchlow [9]. Fagin, Kumar, and Sivakumar [16] studied a variation of these for top k lists. Kendall [23] defined two versions of the Kendall tau distance for partial rankings; one of these versions is a normalized version of our Kendall tau distance through profiles. Baggerly [5] defined two versions of the Spearman footrule distance for partial rankings; one of these versions is similar to our Spearman footrule metric through profiles. However, neither Kendall nor Baggerly proceeded significantly beyond simply providing the definition. Goodman and Kruskal [20] proposed an approach for comparing partial rankings, which was recently utilized [21] for evaluating strategies for similarity search on the Web. A serious disadvantage of Goodman and Kruskal's approach is that it is not always defined (this problem did not arise in the application of [21]).

Rank aggregation and partial rankings. As alluded to earlier, rank aggregation is the problem of combining several ranked lists of objects in a robust way to produce a single consensus ranking of the objects. In computer science, rank aggregation has proved to be a useful and powerful paradigm in several applications including metasearch [4, 12, 24, 25, 26, 29], combining experts [8], synthesizing rank functions from multiple indices [15], biological databases [28], similarity search [17], and classification [17, 24].

[1] The Hausdorff distance between two point sets A and B in a metric space with metric $d(\cdot,\cdot)$ is defined as $\max\{\max_{\gamma_1\in A}\min_{\gamma_2\in B} d(\gamma_1,\gamma_2), \max_{\gamma_2\in B}\min_{\gamma_1\in A} d(\gamma_1,\gamma_2)\}$.

There has been an extensive body of work in economics and computer science on providing a mathematical basis for aggregation of rankings. In the "axiomatic approach," one formulates a set of desiderata that the aggregation function is supposed to satisfy, and characterizes various aggregation functions in terms of the "axioms" they satisfy. The classical result of Arrow [2] shows that a small set of fairly natural requirements cannot be simultaneously achieved by any nontrivial aggregation function. For a comprehensive account of specific criteria satisfied by various aggregation methods, see the survey by Fishburn [18]. In the "metric approach," one starts with a metric on the underlying set of rankings (such as permutations or top k lists) and defines the aggregation problem as that of finding a consensus ranking (permutation or top k list, respectively) whose total distance to the given rankings is minimized. It is, of course, natural to study which axioms a given metric method satisfies, and indeed several such results are known (again, see Fishburn's survey [18]).

A prime consideration in the adoption of a metric aggregation method in computer science applications is whether it admits an efficient exact or provably approximate solution. Several metric methods with excellent properties (e.g., aggregating full lists with respect to the Kendall tau distance) turn out to be NP-hard to solve exactly [6, 12]; fortunately, results like the Diaconis–Graham inequality rescue us from this despair, since if two metrics are equivalent and one of them admits an efficient algorithm, we automatically obtain an efficient approximation algorithm for the other! This is one of the main reasons for our interest in obtaining equivalences between metrics.

While the work of [12, 16] and follow-up efforts offer a fairly clear picture on how to compare and aggregate full or top k lists, the context of database-centric applications poses a new, and rather formidable, challenge. As outlined earlier through the example of online commerce systems, as a result of nonnumeric/few-valued attributes, we encounter partial rankings much more than full rankings in some contexts. While it is possible to treat this issue heuristically by arbitrarily ordering the tied elements to produce a full ranking, a mathematically well-founded treatment becomes possible once we are equipped with metrics on partial rankings. By the equivalence outlined above, it follows that every constant-factor approximation algorithm for rank aggregation with respect to one of our metrics automatically yields a constant-factor approximation algorithm with respect to all of our metrics. These facts were crucially used in [14] to obtain approximation algorithms for the problem of aggregating partial rankings.

2. Preliminaries. *Bucket orders.* A *bucket order* is, intuitively, a (strict) linear order with ties. More formally, a bucket order is a transitive binary relation $\prec$ for which there are sets $\mathcal{B}_1, \ldots, \mathcal{B}_t$ (the *buckets*) that form a partition of the domain such that $x \prec y$ if and only if there are i, j with $i < j$ such $x \in \mathcal{B}_i$ and $y \in \mathcal{B}_j$. If $x \in \mathcal{B}_i$, we may refer to $\mathcal{B}_i$ as the *bucket of* x. We may say that bucket $\mathcal{B}_i$ *precedes* bucket $\mathcal{B}_j$ if $i < j$. Thus, $x \prec y$ if and only if the bucket of x precedes the bucket of y. We think of the members of a given bucket as "tied." A linear order is a bucket order where every bucket is of size 1. We now define the *position* of bucket $\mathcal{B}$, denoted $pos(\mathcal{B})$. Let $\mathcal{B}_1, \ldots, \mathcal{B}_t$ be the buckets in order (so that bucket $\mathcal{B}_i$ precedes bucket $\mathcal{B}_j$ when $i < j$). Then $pos(\mathcal{B}_i) = (\sum_{j<i} |\mathcal{B}_j|) + (|\mathcal{B}_i| + 1)/2$. Intuitively, $pos(\mathcal{B}_i)$ is the average location within bucket $\mathcal{B}_i$.

Comment on terminology.[2] A bucket order $\prec$ is *irreflexive*, that is, there is no x for which $x \prec x$ holds. The corresponding reflexive version $\preceq$ is defined by

[2] The authors are grateful to Bernard Monjardet for providing the information in this paragraph.

saying $x \preceq y$ precisely if either $x \prec y$ or $x = y$. What we call a bucket order is sometimes called a "weak order" (or "weak ordering") [1, 19]. But unfortunately, the corresponding reflexive version $\preceq$ is also sometimes called a weak order (or weak ordering) [2, 13, 27]. A bucket order is sometimes called a "strict weak order" (or "strict weak ordering") [7, 27]. The reflexive version is sometimes called a "complete preordering" [3] or a "total preorder" [7]. We are using the terminology bucket order because it is suggestive and unambiguous.

Partial ranking. Just as we can associate a ranking with a linear order (i.e., permutation), we associate a *partial ranking* $\boldsymbol{\sigma}$ with each bucket order, by letting $\boldsymbol{\sigma}(x) = pos(\mathcal{B})$ when $\mathcal{B}$ is the bucket of x. We refer to a partial ranking associated with a linear order as a *full ranking*. When it is not otherwise specified, we assume that all partial rankings have the same domain, denoted D. We say that *x is ahead of y in $\boldsymbol{\sigma}$* if $\boldsymbol{\sigma}(x) < \boldsymbol{\sigma}(y)$. We say that *$x$ and y are tied in $\boldsymbol{\sigma}$* if $\boldsymbol{\sigma}(x) = \boldsymbol{\sigma}(y)$. When we speak of the buckets of a partial ranking, we are referring to the buckets of the corresponding bucket order.

We define a *top k list* to be a partial ranking where the top k buckets are singletons, representing the top k elements, and the bottom bucket contains all other members of the domain. Note that in [16] there is no bottom bucket in a top k list. This is because in [16] each top k list has its own domain of size k, unlike our scenario where there is a fixed domain.

Given a partial ranking $\boldsymbol{\sigma}$ with domain D, we define its *reverse*, denoted $\boldsymbol{\sigma}^{\mathrm{R}}$, in the expected way. That is, for all $d \in D$, let $\boldsymbol{\sigma}^{\mathrm{R}}(d) = |D| + 1 - \boldsymbol{\sigma}(d)$.

We also define the notion of *swapping* in the normal way. If $a, b \in D$, then *swapping a and b in $\boldsymbol{\sigma}$* produces a new order $\boldsymbol{\sigma}'$, where $\boldsymbol{\sigma}'(a) = \boldsymbol{\sigma}(b)$, $\boldsymbol{\sigma}'(b) = \boldsymbol{\sigma}(a)$, and $\boldsymbol{\sigma}'(d) = \boldsymbol{\sigma}(d)$ for all $d \in D \setminus \{a, b\}$.

Refinements of partial rankings. Given two partial rankings $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$, both with domain D, we say that $\boldsymbol{\sigma}$ is a *refinement* of $\boldsymbol{\tau}$ and write $\boldsymbol{\sigma} \succeq \boldsymbol{\tau}$ if the following holds: for all $i, j \in D$, we have $\boldsymbol{\sigma}(i) < \boldsymbol{\sigma}(j)$ whenever $\boldsymbol{\tau}(i) < \boldsymbol{\tau}(j)$. Notice that when $\boldsymbol{\tau}(i) = \boldsymbol{\tau}(j)$, there is no order forced on $\boldsymbol{\sigma}$. When $\boldsymbol{\sigma}$ is a full ranking, we say that $\boldsymbol{\sigma}$ is a *full refinement* of $\boldsymbol{\tau}$. Given two partial rankings $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$, both with domain D, we frequently make use of a particular refinement of $\boldsymbol{\sigma}$ in which ties are broken according to $\boldsymbol{\tau}$. Define the *$\boldsymbol{\tau}$-refinement of $\boldsymbol{\sigma}$*, denoted $\boldsymbol{\tau} * \boldsymbol{\sigma}$, to be the refinement of $\boldsymbol{\sigma}$ with the following properties. For all $i, j \in D$, if $\boldsymbol{\sigma}(i) = \boldsymbol{\sigma}(j)$ and $\boldsymbol{\tau}(i) < \boldsymbol{\tau}(j)$, then $(\boldsymbol{\tau} * \boldsymbol{\sigma})(i) < (\boldsymbol{\tau} * \boldsymbol{\sigma})(j)$. If $\boldsymbol{\sigma}(i) = \boldsymbol{\sigma}(j)$ and $\boldsymbol{\tau}(i) = \boldsymbol{\tau}(j)$, then $(\boldsymbol{\tau} * \boldsymbol{\sigma})(i) = (\boldsymbol{\tau} * \boldsymbol{\sigma})(j)$. Notice that when $\boldsymbol{\tau}$ is in fact a full ranking, then $\boldsymbol{\tau} * \boldsymbol{\sigma}$ is also a full ranking. Also note that $*$ is an associative operation, so that if $\boldsymbol{\rho}$ is another partial ranking with domain D, it makes sense to talk about $\boldsymbol{\rho} * \boldsymbol{\tau} * \boldsymbol{\sigma}$.

Notation. When f and g are functions with the same domain D, we denote the L_1 distance between f and g by $L_1(f, g)$. Thus, $L_1(f, g) = \sum_{i \in D} |f(i) - g(i)|$.

2.1. Metrics, near metrics, and equivalence. A binary function d is called *symmetric* if $d(x, y) = d(y, x)$ for all x, y in the domain, and it is called *regular* if $d(x, y) = 0$ if and only if $x = y$. A *distance measure* is a nonnegative, symmetric, regular binary function. A *metric* is a distance measure d that satisfies the *triangle inequality*: $d(x, z) \leq d(x, y) + d(y, z)$ for all x, y, z in the domain.

The definitions and results in this section were derived in [16], in the context of comparing top k lists. Two seemingly different notions of a "near metric" were defined in [16]: their first notion of near metric is based on "relaxing" the polygonal inequality that a metric is supposed to satisfy.

DEFINITION 1 (near metric). *A distance measure on partial rankings with domain D is a* near metric *if there is a constant c, independent of the size of D, such that the distance measure satisfies the relaxed polygonal inequality: $d(x,z) \le c(d(x,x_1) + d(x_1,x_2) + \cdots + d(x_{n-1},z))$ for all $n > 1$ and $x,z,x_1,\ldots,x_{n-1} \in D$.*

It makes sense to say that the constant c is independent of the size of D when, as in [16], each of the distance measures considered is actually a family, parameterized by D. We need to make an assumption that c is independent of the size of D, since otherwise we are simply considering distance measures over finite domains, where there is always such a constant c.

The other notion of near metric given in [16] is based on bounding the distance measure above and below by positive constant multiples of a metric. It was shown that both the notions of near metrics coincide.[3] This theorem inspired a definition of what it means for a distance measure to be "almost" a metric, and a robust notion of "similar" or "equivalent" distance measures. We modify the definitions in [16] slightly to fit our scenario, where there is a fixed domain D.

DEFINITION 2 (equivalent distance measures). *Two distance measures d and d' between partial rankings with domain D are* equivalent *if there are positive constants c_1 and c_2, independent of the size of D, such that $c_1 d'(\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2) \le d(\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2) \le c_2 d'(\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2)$ for every pair $\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2$ of partial rankings.*

It is clear that the above definition leads to an equivalence relation (i.e., reflexive, symmetric, and transitive). It follows from [16] that a distance measure is equivalent to a metric if and only if it is a near metric.

2.2. Metrics on full rankings.

We now review two well-known notions of metrics on full rankings, namely the Kendall tau distance and the Spearman footrule distance.

Let $\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2$ be two full rankings with domain D. The *Spearman footrule distance* is simply the L_1 distance $L_1(\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2)$. The definition of the Kendall tau distance requires a little more work.

Let $\mathcal{P} = \{\{i,j\} \mid i \ne j \text{ and } i,j \in D\}$ be the set of unordered pairs of distinct elements. The *Kendall tau distance* between full rankings is defined as follows. For each pair $\{i,j\} \in \mathcal{P}$ of distinct members of D, if i and j are in the same order in $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$, then let the penalty $\bar{K}_{i,j}(\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2) = 0$; and if i and j are in the opposite order (such as i being ahead of j in $\boldsymbol{\sigma}_1$ and j being ahead of i in $\boldsymbol{\sigma}_2$), then let $\bar{K}_{i,j}(\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2) = 1$. The Kendall tau distance is given by $K(\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2) = \sum_{\{i,j\}\in\mathcal{P}} \bar{K}_{i,j}(\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2)$. The Kendall tau distance turns out to be equal to the number of exchanges needed in a bubble sort to convert one full ranking to the other.

Diaconis and Graham [11] proved a classical result, which states that for every two full rankings σ_1,σ_2,

$$K(\sigma_1,\sigma_2) \le F(\sigma_1,\sigma_2) \le 2K(\sigma_1,\sigma_2). \tag{1}$$

Thus, the Kendall tau distance and the Spearman footrule distance are equivalent metrics for full rankings.

3. Metrics for comparing partial rankings.

In this section we define metrics on partial rankings. The first set of metrics is based on profile vectors (section 3.1). As part of this development, we consider variations of the Kendall tau distance where

[3] This result would not hold if instead of relaxing the polygonal inequality, we simply relaxed the triangle inequality.

we vary a certain parameter. The second set of metrics is based on the Hausdorff distance (section 3.2). Section 3.3 compares these metrics (when the partial rankings are top k lists) with the distance measures for top k lists that are developed in [16].

3.1. Metrics based on profiles. Let $\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2$ be two partial rankings with domain D. We now define a family of generalizations of the Kendall tau distance to partial rankings. These are based on a generalization [16] of the Kendall tau distance to top k lists.

Let p be a fixed parameter, with $0 \leq p \leq 1$. Similar to our definition of $\bar{K}_{i,j}(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2)$ for full rankings $\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2$, we define a penalty $\bar{K}^{(p)}_{i,j}(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2)$ for partial rankings $\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2$ for $\{i, j\} \in \mathcal{P}$. There are three cases.

Case 1. i and j are in different buckets in both $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$. If i and j are in the same order in $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$ (such as $\boldsymbol{\sigma}_1(i) > \boldsymbol{\sigma}_1(j)$ and $\boldsymbol{\sigma}_2(i) > \boldsymbol{\sigma}_2(j)$), then let $\bar{K}^{(p)}_{i,j}(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2) = 0$; this corresponds to "no penalty" for $\{i, j\}$. If i and j are in the opposite order in $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$ (such as $\boldsymbol{\sigma}_1(i) > \boldsymbol{\sigma}_1(j)$ and $\boldsymbol{\sigma}_2(i) < \boldsymbol{\sigma}_2(j)$), then let the penalty $\bar{K}^{(p)}_{i,j}(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2) = 1$.

Case 2. i and j are in the same bucket in both $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$. We then let the penalty $\bar{K}^{(p)}_{i,j}(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2) = 0$. Intuitively, both partial rankings agree that i and j are tied.

Case 3. i and j are in the same bucket in one of the partial rankings $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$, but in different buckets in the other partial ranking. In this case, we let the penalty $\bar{K}^{(p)}_{i,j}(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2) = p$.

Based on these cases, define $K^{(p)}$, the *Kendall distance with penalty parameter* p, as follows:

$$K^{(p)}(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2) = \sum_{\{i,j\} \in \mathcal{P}} \bar{K}^{(p)}_{i,j}(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2).$$

We now discuss our choice of penalty in Cases 2 and 3. In Case 2, where i and j are in the same bucket in both $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$, what if we had defined there to be a positive penalty $\bar{K}^{(p)}_{i,j}(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2) = q > 0$? Then if $\boldsymbol{\sigma}$ were an arbitrary partial ranking that had some bucket of size at least 2, we would have had $K^{(p)}(\boldsymbol{\sigma}, \boldsymbol{\sigma}) \geq q > 0$. So $K^{(p)}$ would not have been a metric, or even a distance measure, since we would have lost the property that $K^{(p)}(\boldsymbol{\sigma}, \boldsymbol{\sigma}) = 0$. The next proposition shows the effect of the choice of p in Case 3.

PROPOSITION 3. *$K^{(p)}$ is a metric when $1/2 \leq p \leq 1$, is a near metric when $0 < p < 1/2$, and is not a distance measure when $p = 0$.*

Proof. Let us first consider the case $p = 0$. We now show that $K^{(0)}$ is not even a distance measure. Let the domain D have exactly two elements a and b. Let $\boldsymbol{\tau}_1$ be the full ranking where a precedes b, let $\boldsymbol{\tau}_2$ be the partial ranking where a and b are in the same bucket, and let $\boldsymbol{\tau}_3$ be the full ranking where b precedes a. Then $K^{(0)}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) = 0$ even though $\boldsymbol{\tau}_1 \neq \boldsymbol{\tau}_2$. So indeed, $K^{(0)}$ is not a distance measure. Note also that the near triangle inequality is violated badly in this example, since $K^{(0)}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) = 0$ and $K^{(0)}(\boldsymbol{\tau}_2, \boldsymbol{\tau}_3) = 0$, but $K^{(0)}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_3) = 1$.

It is easy to see that $K^{(p)}$ is a distance measure for every p with $0 < p \leq 1$. We now show that $K^{(p)}$ does not satisfy the triangle inequality when $0 < p < 1/2$ and satisfies the triangle inequality when $1/2 \leq p \leq 1$. Let $\boldsymbol{\tau}_1$, $\boldsymbol{\tau}_2$, and $\boldsymbol{\tau}_3$ be as in our previous example. Then $K^{(p)}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) = p$, $K^{(p)}(\boldsymbol{\tau}_2, \boldsymbol{\tau}_3) = p$, and $K^{(p)}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_3) = 1$. So the triangle inequality fails for $0 < p < 1/2$, since $K^{(p)}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_3) > K^{(p)}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) + K^{(p)}(\boldsymbol{\tau}_2, \boldsymbol{\tau}_3)$. On the other hand, the triangle inequality holds for $1/2 \leq p \leq 1$, since then it is easy to verify that $\bar{K}^{(p)}_{i,j}(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_3) \leq \bar{K}^{(p)}_{i,j}(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2) + \bar{K}^{(p)}_{i,j}(\boldsymbol{\sigma}_2, \boldsymbol{\sigma}_3)$ for every i, j, and so $K^{(p)}(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_3) \leq K^{(p)}(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2) + K^{(p)}(\boldsymbol{\sigma}_2, \boldsymbol{\sigma}_3)$.

We now show that $K^{(p)}$ is a near metric when $0 < p < 1/2$. It is easy to verify that if $0 < p < p' \leq 1$, then $K^{(p)}(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2) \leq K^{(p')}(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2) \leq (p'/p)K^{(p)}(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2)$. Hence, all of the distance measures $K^{(p)}$ are equivalent whenever $0 < p$. As noted earlier, it follows from [16] that a distance measure is equivalent to a metric if and only if it is a near metric. Since $K^{(p)}$ is equivalent to the metric $K^{(1/2)}$ when $0 < p$, we conclude that in this case, $K^{(p)}$ is a near metric. □

It is worth stating formally the following simple observation from the previous proof.

PROPOSITION 4. *All of the distance measures $K^{(p)}$ are equivalent whenever* $0 < p \leq 1$.

For the rest of the paper, we focus on the natural case $p = 1/2$, which corresponds to an "average" penalty for two elements i and j that are tied in one partial ranking but not in the other partial ranking. We show that $K^{(1/2)}$ is equivalent to the other metrics we define. It thereby follows from Proposition 4 that each of the distance measures $K^{(p)}$ for $0 < p \leq 1$, and in particular the metrics $K^{(p)}$ for $1/2 \leq p \leq 1$, is equivalent to these other metrics.

We now show there is an alternative interpretation for $K^{(1/2)}$ in terms of a "profile." Let $\mathcal{O} = \{(i, j) : \; i \neq j \text{ and } i, j \in D\}$ be the set of ordered pairs of distinct elements in the domain D. Let $\boldsymbol{\sigma}$ be a partial ranking (as usual, with domain D). For $(i, j) \in \mathcal{O}$, define p_{ij} to be $1/4$ if $\boldsymbol{\sigma}(i) < \boldsymbol{\sigma}(j)$, to be 0 if $\boldsymbol{\sigma}(i) = \boldsymbol{\sigma}(j)$, and to be $-1/4$ if $\boldsymbol{\sigma}(i) > \boldsymbol{\sigma}(j)$. Define the *K-profile* of $\boldsymbol{\sigma}$ to be the vector $\langle p_{ij} : (i, j) \in \mathcal{O}\rangle$ and $K_{\text{prof}}(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2)$ to be the L_1 distance between the K-profiles of $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$. It is easy to verify that $K_{\text{prof}} = K^{(1/2)}$.[4] It is also easy to see that the K-profile of $\boldsymbol{\sigma}$ uniquely determines $\boldsymbol{\sigma}$.

It is clear how to generalize the Spearman footrule distance to partial rankings—we simply take it to be $L_1(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2)$, just as before. We refer to this value as $F_{\text{prof}}(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2)$, for reasons we now explain. Let us define the *F-profile* of a partial ranking $\boldsymbol{\sigma}$ to be the vector of values $\boldsymbol{\sigma}(i)$. So the F-profile is indexed by D, whereas the K-profile is indexed by $\mathcal{O}$. Just as the K_{prof} value of two partial rankings (or of the corresponding bucket orders) is the L_1 distance between their K-profiles, the F_{prof} value of two partial rankings (or of the corresponding bucket orders) is the L_1 distance between their F-profiles. Since K_{prof} and F_{prof} are L_1 distances, and since the K-profile and the F-profile each uniquely determine the partial ranking, it follows that K_{prof} and F_{prof} are both metrics.

3.2. The Hausdorff metrics. Let A and B be finite sets of objects and let d be a metric on objects. The *Hausdorff distance* between A and B is given by

$$d_{\text{Haus}}(A, B) = \max\left\{\max_{\gamma_1 \in A} \min_{\gamma_2 \in B} d(\gamma_1, \gamma_2), \max_{\gamma_2 \in B} \min_{\gamma_1 \in A} d(\gamma_1, \gamma_2)\right\}. \tag{2}$$

Although this looks fairly nonintuitive, it is actually quite natural, as we now explain. The quantity $\min_{\gamma_2 \in B} d(\gamma_1, \gamma_2)$ is the distance between γ_1 and the set B. Therefore, the quantity $\max_{\gamma_1 \in A} \min_{\gamma_2 \in B} d(\gamma_1, \gamma_2)$ is the maximal distance of a member of A from the set B. Similarly, the quantity $\max_{\gamma_2 \in B} \min_{\gamma_1 \in A} d(\gamma_1, \gamma_2)$ is the

[4] The reason that the values of p_{ij} in the K-profile are $1/4$, 0, and $-1/4$ rather than $1/2$, 0, and $-1/2$ is that each pair $\{i, j\}$ with $i \neq j$ is counted twice, once as (i, j) and once as (j, i).

maximal distance of a member of B from the set A. Therefore, the Hausdorff distance between A and B is the maximal distance of a member of A or B from the other set. Thus, A and B are within Hausdorff distance s of each other precisely if every member of A and B is within distance s of some member of the other set. The Hausdorff distance is well known to be a metric.

Critchlow [9] used the Hausdorff distance to define a metric, which we now define, between partial rankings. Given a metric d that gives the distance $d(\gamma_1, \gamma_2)$ between full rankings γ_1 and γ_2, define the distance d_{Haus} between partial rankings $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$ to be

$$d_{\text{Haus}}(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2) = \max\left\{ \max_{\gamma_1 \succeq \boldsymbol{\sigma}_1} \min_{\gamma_2 \succeq \boldsymbol{\sigma}_2} d(\gamma_1, \gamma_2), \max_{\gamma_2 \succeq \boldsymbol{\sigma}_2} \min_{\gamma_1 \succeq \boldsymbol{\sigma}_1} d(\gamma_1, \gamma_2) \right\}, \tag{3}$$

where γ_1 and γ_2 are full rankings. In particular, when d is the footrule distance, this gives us a metric between partial rankings that we call F_{Haus}, and when d is the Kendall distance, this gives us a metric between partial rankings that we call K_{Haus}. Both F_{Haus} and K_{Haus} are indeed metrics, since they are special cases of the Hausdorff distance.

The next theorem, which is due to Critchlow (but which we state using our notation), gives a complete characterization of F_{Haus} and K_{Haus}. For the sake of completeness, we prove this theorem in the appendix.[5]

THEOREM 5 (see [9]). *Let $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$ be partial rankings, let $\boldsymbol{\sigma}^{\text{R}}$ be the reverse of $\boldsymbol{\sigma}$, and let $\boldsymbol{\tau}^{\text{R}}$ be the reverse of $\boldsymbol{\tau}$. Let ρ be any full ranking. Then*

$$\begin{aligned} F_{\text{Haus}}(\boldsymbol{\sigma}, \boldsymbol{\tau}) = \max\{&F(\rho * \boldsymbol{\tau}^{\text{R}} * \boldsymbol{\sigma}, \rho * \boldsymbol{\sigma} * \boldsymbol{\tau}), \\ &F(\rho * \boldsymbol{\tau} * \boldsymbol{\sigma}, \rho * \boldsymbol{\sigma}^{\text{R}} * \boldsymbol{\tau})\}, \\ K_{\text{Haus}}(\boldsymbol{\sigma}, \boldsymbol{\tau}) = \max\{&K(\rho * \boldsymbol{\tau}^{\text{R}} * \boldsymbol{\sigma}, \rho * \boldsymbol{\sigma} * \boldsymbol{\tau}), \\ &K(\rho * \boldsymbol{\tau} * \boldsymbol{\sigma}, \rho * \boldsymbol{\sigma}^{\text{R}} * \boldsymbol{\tau})\}. \end{aligned}$$

Theorem 5 gives us a simple algorithm for computing $F_{\text{Haus}}(\boldsymbol{\sigma}, \boldsymbol{\tau})$ and $K_{\text{Haus}}(\boldsymbol{\sigma}, \boldsymbol{\tau})$: we simply pick an arbitrary full ranking ρ and do the computations given in Theorem 5. Thus, let $\sigma_1 = \rho * \boldsymbol{\tau}^{\text{R}} * \boldsymbol{\sigma}$, let $\tau_1 = \rho * \boldsymbol{\sigma} * \boldsymbol{\tau}$, let $\sigma_2 = \rho * \boldsymbol{\tau} * \boldsymbol{\sigma}$, and let $\tau_2 = \rho * \boldsymbol{\sigma}^{\text{R}} * \boldsymbol{\tau}$. Theorem 5 tells us that $F_{\text{Haus}}(\boldsymbol{\sigma}, \boldsymbol{\tau}) = \max\{F(\sigma_1, \tau_1), F(\sigma_2, \tau_2)\}$ and $K_{\text{Haus}}(\boldsymbol{\sigma}, \boldsymbol{\tau}) = \max\{K(\sigma_1, \tau_1), K(\sigma_2, \tau_2)\}$. It is interesting that the same pairs, namely (σ_1, τ_1) and (σ_2, τ_2), are the candidates for exhibiting the Hausdorff distance for both F and K. Note that the only role that the arbitrary full ranking ρ plays is to arbitrarily break ties (in the same way for $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$) for pairs (i, j) of distinct elements that are in the same bucket in both $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$. A way to describe the pair (σ_1, τ_1) intuitively is as follows: break the ties in $\boldsymbol{\sigma}$ based on the reverse of the ordering in $\boldsymbol{\tau}$, break the ties in $\boldsymbol{\tau}$ based on the ordering in $\boldsymbol{\sigma}$, and break any remaining ties arbitrarily (but in the same way in both). A similar description applies to the pair (σ_2, τ_2).

The algorithm just described for computing $F_{\text{Haus}}(\boldsymbol{\sigma}, \boldsymbol{\tau})$ and $K_{\text{Haus}}(\boldsymbol{\sigma}, \boldsymbol{\tau})$ is based on creating pairs (σ_1, τ_1) and (σ_2, τ_2), one of which must exhibit the Hausdorff distance. The next theorem gives a direct algorithm for computing $K_{\text{Haus}}(\boldsymbol{\sigma}, \boldsymbol{\tau})$ that we make use of later.

THEOREM 6. *Let $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$ be partial rankings. Let S be the set of pairs $\{i, j\}$ of distinct elements such that i and j appear in the same bucket of $\boldsymbol{\sigma}$ but in different*

[5]Our proof arose when, unaware of Critchlow's result, we derived and proved this theorem.

buckets of $\boldsymbol{\tau}$*, let* T *be the set of pairs* $\{i, j\}$ *of distinct elements such that* i *and* j *appear in the same bucket of* $\boldsymbol{\tau}$ *but in different buckets of* $\boldsymbol{\sigma}$*, and let* U *be the set of pairs* $\{i, j\}$ *of distinct elements that are in different buckets of both* $\boldsymbol{\sigma}$ *and* $\boldsymbol{\tau}$ *and are in a different order in* $\boldsymbol{\sigma}$ *and* $\boldsymbol{\tau}$*. Then* $K_{\mathrm{Haus}}(\boldsymbol{\sigma}, \boldsymbol{\tau}) = |U| + \max\{|S|, |T|\}$*.*

Proof. As before, let $\sigma_1 = \rho * \boldsymbol{\tau}^{\mathrm{R}} * \boldsymbol{\sigma}$, let $\tau_1 = \rho * \boldsymbol{\sigma} * \boldsymbol{\tau}$, let $\sigma_2 = \rho * \boldsymbol{\tau} * \boldsymbol{\sigma}$, and let $\tau_2 = \rho * \boldsymbol{\sigma}^{\mathrm{R}} * \boldsymbol{\tau}$. It is straightforward to see that the set of pairs $\{i, j\}$ of distinct elements that are in a different order in σ_1 and τ_1 is exactly the union of the disjoint sets U and S. Therefore, $K(\sigma_1, \tau_1) = |U| + |S|$. Identically, the set of pairs $\{i, j\}$ of distinct elements that are in a different order in σ_2 and τ_2 is exactly the union of the disjoint sets U and T, and hence $K(\sigma_2, \tau_2) = |U| + |T|$. But by Theorem 5, we know that $K_{\mathrm{Haus}}(\boldsymbol{\sigma}, \boldsymbol{\tau}) = \max\{K(\sigma_1, \tau_1), K(\sigma_2, \tau_2)\} = \max\{|U| + |S|, |U| + |T|\}$. The result follows immediately. ∎

3.3. Metrics in this paper for top $\boldsymbol{k}$ **lists vs. distance measures defined in** [10]**.** Metrics on partial rankings naturally induce metrics on top k lists. We now compare (a) the metrics on top k lists that are induced by our metrics on partial rankings with (b) the distance measures on top k lists that were introduced in [16]. Recall that for us, a top k list is a partial ranking consisting of k singleton buckets, followed by a bottom bucket of size $|D| - k$. However, in [16], a top k list is a bijection of a domain ("the top k elements") onto $\{1, \ldots, k\}$. Let $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$ be top k lists (of our form). Define the *active domain for* $\boldsymbol{\sigma}$*,* $\boldsymbol{\tau}$ to be the union of the elements in the top k buckets of $\boldsymbol{\sigma}$ and the elements in the top k buckets of $\boldsymbol{\tau}$. In order to make our scenario compatible with the scenario of [16], we assume during our comparison that the domain D equals the active domain for $\boldsymbol{\sigma}$, $\boldsymbol{\tau}$. Our definitions of $K^{(p)}$, F_{Haus}, and K_{Haus} are then exactly the same in the two scenarios. (Unlike the situation in section 3.1, even the case $p = 0$ gives a distance measure, since the unpleasant case where $K^{(0)}(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2) = 0$ even though $\boldsymbol{\sigma}_1 \neq \boldsymbol{\sigma}_2$ does not arise for top k lists $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$.) Nevertheless, $K^{(p)}$, F_{Haus}, and K_{Haus} are only near metrics in [16] in spite of being metrics for us. This is because, in [16], the active domain varies depending on which pair of top k lists is being compared.

Our definition of $K_{\mathrm{prof}}(\boldsymbol{\sigma}, \boldsymbol{\tau})$ is equivalent to the definition of $K_{\mathrm{avg}}(\boldsymbol{\sigma}, \boldsymbol{\tau})$ in [16], namely the average value of $K(\sigma, \tau)$ over all full rankings σ, τ with domain D, where $\sigma \succeq \boldsymbol{\sigma}$ and $\tau \succeq \boldsymbol{\tau}$. It is interesting to note that if $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$ were not top k lists but arbitrary partial rankings, then K_{avg} would not be a distance measure, since $K_{\mathrm{avg}}(\boldsymbol{\sigma}, \boldsymbol{\sigma})$ can be strictly positive if $\boldsymbol{\sigma}$ is an arbitrary partial ranking.

Let ℓ be a real number greater than k. The *footrule distance with location parameter* ℓ, denoted $F^{(\ell)}$, is defined by treating each element that is not among the top k elements as if it were in position ℓ, and then taking the L_1 distance [16]. More formally, let $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$ be top k lists (of our form). Define the function $f_{\boldsymbol{\sigma}}$ with domain D by letting $f_{\boldsymbol{\sigma}}(i) = \boldsymbol{\sigma}(i)$ if $1 \leq \boldsymbol{\sigma}(i) \leq k$, and $f_{\boldsymbol{\sigma}}(i) = \ell$ otherwise. Similarly, define the function $f_{\boldsymbol{\tau}}$ with domain D by letting $f_{\boldsymbol{\tau}}(i) = \boldsymbol{\tau}(i)$ if $1 \leq \boldsymbol{\tau}(i) \leq k$, and $f_{\boldsymbol{\tau}}(i) = \ell$ otherwise. Then $F^{(\ell)}(\boldsymbol{\sigma}, \boldsymbol{\tau})$ is defined to be $L_1(f_{\boldsymbol{\sigma}}, f_{\boldsymbol{\tau}})$. It is straightforward to verify that $F_{\mathrm{prof}}(\boldsymbol{\sigma}, \boldsymbol{\tau}) = F^{(\ell)}(\boldsymbol{\sigma}, \boldsymbol{\tau})$ for $\ell = (|D| + k + 1)/2$.

4. Equivalence between the metrics. In this section we prove our main theorem, which says that our four metrics are equivalent.

THEOREM 7. *The metrics* F_{prof}*,* K_{prof}*,* F_{Haus}*, and* K_{Haus} *are all equivalent, that is, within constant multiples of each other.*

Proof. First, we show

$$K_{\mathrm{Haus}}(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2) \leq F_{\mathrm{Haus}}(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2) \leq 2K_{\mathrm{Haus}}(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2). \tag{4}$$

The proof of this equivalence between F_{Haus} and K_{Haus} uses the robustness of the Hausdorff definition with respect to equivalent metrics. It is fairly easy, and is given in section 4.1.

Next, we show

$$K_{\text{prof}}(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2) \leq F_{\text{prof}}(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2) \leq 2K_{\text{prof}}(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2). \tag{5}$$

We note that (5) is much more complicated to prove than (4) and is also much more complicated to prove than the Diaconis–Graham inequality (1). The proof involves two main concepts: "reflecting" each partial ranking so that every element has a mirror image and using the notion of "nesting," which means that the interval spanned by an element and its image in one partial ranking sits inside the interval spanned by the same element and its image in the other partial ranking. The proof is presented in section 4.2.

We note that the equivalences given by (4) and (5) are interesting in their own right.

Finally, we show in section 4.3 that

$$K_{\text{prof}}(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2) \leq K_{\text{Haus}}(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2) \leq 2K_{\text{prof}}(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2). \tag{6}$$

This is proved using Theorem 6.

Using (4), (5), and (6), the proof is complete, since (4) tells us that the two Hausdorff metrics are equivalent, (5) tells us that the two profile metrics are equivalent, and (6) tells us that some Hausdorff metric is equivalent to some profile metric. □

4.1. Equivalence of F_{Haus} and K_{Haus}. In this section, we prove the simple result that the Diaconis–Graham inequalities (1) extend to F_{Haus} and K_{Haus}. We begin with a lemma. In this lemma, for metric d, we define d_{Haus} as in (2), and similarly for metric d'.

LEMMA 8. *Assume that d and d' are metrics where there is a constant c such that $d \leq c \cdot d'$. Then $d_{\text{Haus}} \leq c \cdot d'_{\text{Haus}}$.*

Proof. Let A and B be as in (2). Assume without loss of generality (by reversing A and B if necessary) that $d_{\text{Haus}}(A, B) = \max_{\gamma_1 \in A} \min_{\gamma_2 \in B} d(\gamma_1, \gamma_2)$. Find γ_1 in A that maximizes $\min_{\gamma_2 \in B} d(\gamma_1, \gamma_2)$, and γ_2 in B that minimizes $d(\gamma_1, \gamma_2)$. Therefore, $d_{\text{Haus}}(A, B) = d(\gamma_1, \gamma_2)$. Find γ_2' in B that minimizes $d'(\gamma_1, \gamma_2')$. (There is such a γ_2' since by assumption on the definition of Hausdorff distance, A and B are finite sets.) Then $d_{\text{Haus}}(A, B) = d(\gamma_1, \gamma_2) \leq d(\gamma_1, \gamma_2')$, since γ_2 minimizes $d(\gamma_1, \gamma_2)$. Also $d(\gamma_1, \gamma_2') \leq c \cdot d'(\gamma_1, \gamma_2')$, by assumption on d and d'. Finally $c \cdot d'(\gamma_1, \gamma_2') \leq c \cdot d'_{\text{Haus}}(A, B)$, by definition of d'_{Haus} and the fact that γ_2' minimizes $d'(\gamma_1, \gamma_2')$. Putting these inequalities together, we obtain $d_{\text{Haus}}(A, B) \leq c \cdot d'_{\text{Haus}}(A, B)$, which completes the proof. □

We can now show that the Diaconis–Graham inequalities (1) extend to F_{Haus} and K_{Haus}.

THEOREM 9. *Let $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$ be partial rankings. Then $K_{\text{Haus}}(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2) \leq F_{\text{Haus}}(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2) \leq 2K_{\text{Haus}}(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2)$.*

Proof. The first inequality $K_{\text{Haus}}(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2) \leq F_{\text{Haus}}(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2)$ follows from the first Diaconis–Graham inequality $K(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2) \leq F(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2)$ and Lemma 8, where we let the roles of d, d', and c be played by K, F, and 1, respectively. The second inequality $F_{\text{Haus}}(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2) \leq 2K_{\text{Haus}}(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2)$ follows from the second Diaconis–Graham inequality $F(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2) \leq 2K(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2)$ and Lemma 8, where we let the roles of d, d', and c be played by F, K, and 2, respectively. □

4.2. Equivalence of F_{prof} and K_{prof}. In order to generalize the Diaconis–Graham inequalities to F_{prof} and K_{prof}, we convert a pair of partial rankings into full rankings (on an enlarged domain) in such a way that the F_{prof} distance between the partial rankings is precisely 4 times the F distance between the full rankings, and the K_{prof} distance between the partial rankings is precisely 4 times the K distance between the full rankings. Given a domain D, produce a "duplicate set" $D^\sharp = \{i^\sharp : i \in D\}$. Given a partial ranking $\boldsymbol{\sigma}$ with domain D, produce a new partial ranking $\boldsymbol{\sigma}^\sharp$, with domain $D \cup D^\sharp$, as follows. Modify the bucket order associated with $\boldsymbol{\sigma}$ by putting $i^\sharp$ in the same bucket as i for each $i \in D$. We thereby double the size of every bucket. Let $\boldsymbol{\sigma}^\sharp$ be the partial ranking associated with this new bucket order. Since $i^\sharp$ is in the same bucket as i, we have $\boldsymbol{\sigma}^\sharp(i) = \boldsymbol{\sigma}^\sharp(i^\sharp)$. We now show that $\boldsymbol{\sigma}^\sharp(i) = 2\boldsymbol{\sigma}(i) - 1/2$ for all i in D.

Fix i in D, let p be the number of elements j in D such that $\boldsymbol{\sigma}(j) < \boldsymbol{\sigma}(i)$, and let q be the number of elements k in D such that $\boldsymbol{\sigma}(k) = \boldsymbol{\sigma}(i)$. By the definition of the ranking associated with a bucket order, we have

$$\boldsymbol{\sigma}(i) = p + (q+1)/2. \tag{7}$$

Since each bucket doubles in size for the bucket order associated with $\boldsymbol{\sigma}^\sharp$, we similarly have

$$\boldsymbol{\sigma}^\sharp(i) = 2p + (2q+1)/2. \tag{8}$$

It follows easily from (7) and (8) that $\boldsymbol{\sigma}^\sharp(i) = 2\boldsymbol{\sigma}(i) - 1/2$, as desired.

We need to obtain a full ranking from the partial ranking $\boldsymbol{\sigma}^\sharp$. First, for every full ranking π with domain D, define a full ranking $\pi^\dagger$ with domain $D \cup D^\sharp$ as follows:

$$\pi^\dagger(d) = \pi(d) \quad \text{for all } d \in D,$$
$$\pi^\dagger(d^\sharp) = 2|D| + 1 - \pi(d) \quad \text{for all } d \text{ in } D$$

so that $\pi^\dagger$ ranks elements of D in the same order as π, elements of $D^\sharp$ in the reverse order of π, and all elements of D before all elements of $D^\sharp$.

We define $\boldsymbol{\sigma}_\pi = \pi^\dagger * (\boldsymbol{\sigma}^\sharp)$. For instance, suppose $\mathcal{B}$ is a bucket of $\boldsymbol{\sigma}^\sharp$ containing the items $a, b, c, a^\sharp, b^\sharp, c^\sharp$, and suppose that π orders the items $\pi(a) < \pi(b) < \pi(c)$. Then $\boldsymbol{\sigma}_\pi$ will contain the sequence $a, b, c, c^\sharp, b^\sharp, a^\sharp$. Also notice that in this example, $\frac{1}{2}(\boldsymbol{\sigma}_\pi(a) + \boldsymbol{\sigma}_\pi(a^\sharp)) = \frac{1}{2}(\boldsymbol{\sigma}_\pi(b) + \boldsymbol{\sigma}_\pi(b^\sharp)) = \frac{1}{2}(\boldsymbol{\sigma}_\pi(c) + \boldsymbol{\sigma}_\pi(c^\sharp)) = pos(\mathcal{B})$. In fact, because of this "reflected-duplicate" property, we see that in general, for every $d \in D$,

$$\frac{1}{2}(\boldsymbol{\sigma}_\pi(d) + \boldsymbol{\sigma}_\pi(d^\sharp)) = \boldsymbol{\sigma}^\sharp(d) = \boldsymbol{\sigma}^\sharp(d^\sharp) = 2\boldsymbol{\sigma}(d) - 1/2. \tag{9}$$

The following lemma shows that no matter what order π we choose, the Kendall distance between $\boldsymbol{\sigma}_\pi$ and $\boldsymbol{\tau}_\pi$ is exactly 4 times the K_{prof} distance between $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$.

LEMMA 10. *Let $\boldsymbol{\sigma}, \boldsymbol{\tau}$ be partial rankings, and let π be any full ranking on the same domain. Then $K(\boldsymbol{\sigma}_\pi, \boldsymbol{\tau}_\pi) = 4K_{\text{prof}}(\boldsymbol{\sigma}, \boldsymbol{\tau})$.*

Proof. Assume that i and j are in D. Let us consider the cases in the definition of $K^{(p)}$ (recall that K_{prof} equals $K^{(p)}$ when $p = 1/2$).

Case 1. i and j are in different buckets in both $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$. If i and j are in the same order in $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$, then the pair $\{i, j\}$ contributes no penalty to $K_{\text{prof}}(\boldsymbol{\sigma}, \boldsymbol{\tau})$, and no pair of members of the set $\{i, j, i^\sharp, j^\sharp\}$ contribute any penalty to $K(\boldsymbol{\sigma}_\pi, \boldsymbol{\tau}_\pi)$. If i and j are in the opposite order in $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$, then the pair $\{i, j\}$ contributes a penalty of 1 to

$K_{\text{prof}}(\boldsymbol{\sigma}, \boldsymbol{\tau})$, and the pairs among $\{i, j, i^\sharp, j^\sharp\}$ that contribute a penalty to $K(\boldsymbol{\sigma}_\pi, \boldsymbol{\tau}_\pi)$ are precisely $\{i, j\}$, $\{i^\sharp, j^\sharp\}$, $\{i, j^\sharp\}$, and $\{i^\sharp, j\}$, each of which contributes a penalty of 1.

Case 2. i and j are in the same bucket in both $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$. Then the pair $\{i, j\}$ contributes no penalty to $K_{\text{prof}}(\boldsymbol{\sigma}, \boldsymbol{\tau})$, and no pair of members of the set $\{i, j, i^\sharp, j^\sharp\}$ contribute any penalty to $K(\boldsymbol{\sigma}_\pi, \boldsymbol{\tau}_\pi)$.

Case 3. i and j are in the same bucket in one of the partial rankings $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$, but in different buckets in the other partial ranking. Then the pair $\{i, j\}$ contributes a penalty of 1/2 to $K_{\text{prof}}(\boldsymbol{\sigma}, \boldsymbol{\tau})$. Assume without loss of generality that i and j are in the same bucket in $\boldsymbol{\sigma}$ and that $\boldsymbol{\tau}(i) < \boldsymbol{\tau}(j)$. There are now two subcases, depending on whether $\pi(i) < \pi(j)$ or $\pi(j) < \pi(i)$. In the first subcase, when $\pi(i) < \pi(j)$, we have

$$\boldsymbol{\sigma}_\pi(i) < \boldsymbol{\sigma}_\pi(j) < \boldsymbol{\sigma}_\pi(j^\sharp) < \boldsymbol{\sigma}_\pi(i^\sharp)$$

and

$$\boldsymbol{\tau}_\pi(i) < \boldsymbol{\tau}_\pi(i^\sharp) < \boldsymbol{\tau}_\pi(j) < \boldsymbol{\tau}_\pi(j^\sharp).$$

So the pairs among $\{i, j, i^\sharp, j^\sharp\}$ that contribute a penalty to $K(\boldsymbol{\sigma}_\pi, \boldsymbol{\tau}_\pi)$ are precisely $\{i^\sharp, j\}$ and $\{i^\sharp, j^\sharp\}$, each of which contribute a penalty of 1.

In the second subcase, when $\pi(j) < \pi(i)$, we have

$$\boldsymbol{\sigma}_\pi(j) < \boldsymbol{\sigma}_\pi(i) < \boldsymbol{\sigma}_\pi(i^\sharp) < \boldsymbol{\sigma}_\pi(j^\sharp)$$

and

$$\boldsymbol{\tau}_\pi(i) < \boldsymbol{\tau}_\pi(i^\sharp) < \boldsymbol{\tau}_\pi(j) < \boldsymbol{\tau}_\pi(j^\sharp).$$

So the pairs among $\{i, j, i^\sharp, j^\sharp\}$ that contribute a penalty to $K(\boldsymbol{\sigma}_\pi, \boldsymbol{\tau}_\pi)$ are precisely $\{i, j\}$ and $\{i^\sharp, j\}$, each of which contribute a penalty of 1.

In all cases, the amount of penalty contributed to $K(\boldsymbol{\sigma}_\pi, \boldsymbol{\tau}_\pi)$ is 4 times the amount of penalty contributed to $K_{\text{prof}}(\boldsymbol{\sigma}, \boldsymbol{\tau})$. The lemma then follows. □

Notice that Lemma 10 holds for every choice of π. The analogous statement is not true for F_{prof}. In that case, we need to choose π specifically for the pair of partial rankings we are given. In particular, we need to avoid a property we call "nesting."

Given fixed $\boldsymbol{\sigma}, \boldsymbol{\tau}$, we say that an element $d \in D$ is *nested* with respect to π if either

$$[\boldsymbol{\sigma}_\pi(d), \boldsymbol{\sigma}_\pi(d^\sharp)] \sqsubset [\boldsymbol{\tau}_\pi(d), \boldsymbol{\tau}_\pi(d^\sharp)]$$
$$\text{or} \quad [\boldsymbol{\tau}_\pi(d), \boldsymbol{\tau}_\pi(d^\sharp)] \sqsubset [\boldsymbol{\sigma}_\pi(d), \boldsymbol{\sigma}_\pi(d^\sharp)],$$

where the notation $[s, t] \sqsubset [u, v]$ for numbers s, t, u, v means that $[s, t] \subseteq [u, v]$ and $s \neq u$ and $t \neq v$. It is sometimes convenient to write $[u, v] \sqsupset [s, t]$ for $[s, t] \sqsubset [u, v]$.

The following lemma shows us why we want to avoid nesting.

LEMMA 11. *Given partial rankings $\boldsymbol{\sigma}, \boldsymbol{\tau}$ and full ranking π, suppose that there are no elements that are nested with respect to π. Then $F(\boldsymbol{\sigma}_\pi, \boldsymbol{\tau}_\pi) = 4F_{\text{prof}}(\boldsymbol{\sigma}, \boldsymbol{\tau})$.*

Proof. Assume $d \in D$. By assumption, d is not nested with respect to π. We now show that

$$\begin{aligned} &|\boldsymbol{\sigma}_\pi(d) - \boldsymbol{\tau}_\pi(d)| + |\boldsymbol{\sigma}_\pi(d^\sharp) - \boldsymbol{\tau}_\pi(d^\sharp)| \\ &= |\boldsymbol{\sigma}_\pi(d) - \boldsymbol{\tau}_\pi(d) + \boldsymbol{\sigma}_\pi(d^\sharp) - \boldsymbol{\tau}_\pi(d^\sharp)|. \end{aligned} \tag{10}$$

There are three cases, depending on whether $\boldsymbol{\sigma}_\pi(d) = \boldsymbol{\tau}_\pi(d)$, $\boldsymbol{\sigma}_\pi(d) < \boldsymbol{\tau}_\pi(d)$, or $\boldsymbol{\sigma}_\pi(d) > \boldsymbol{\tau}_\pi(d)$.

If $\boldsymbol{\sigma}_\pi(d) = \boldsymbol{\tau}_\pi(d)$, then (10) is immediate. If $\boldsymbol{\sigma}_\pi(d) < \boldsymbol{\tau}_\pi(d)$, then necessarily $\boldsymbol{\sigma}_\pi(d^\sharp) \leq \boldsymbol{\tau}_\pi(d^\sharp)$, since d is not nested. But then the left-hand side and right-hand side of (10) are each $\boldsymbol{\tau}_\pi(d) - \boldsymbol{\sigma}_\pi(d) + \boldsymbol{\tau}_\pi(d^\sharp) - \boldsymbol{\sigma}_\pi(d^\sharp)$, and so (10) holds. If $\boldsymbol{\sigma}_\pi(d) > \boldsymbol{\tau}_\pi(d)$, then necessarily $\boldsymbol{\sigma}_\pi(d^\sharp) \geq \boldsymbol{\tau}_\pi(d^\sharp)$, since d is not nested. But then the left-hand side and right-hand side of (10) are each $\boldsymbol{\sigma}_\pi(d) - \boldsymbol{\tau}_\pi(d) + \boldsymbol{\sigma}_\pi(d^\sharp) - \boldsymbol{\tau}_\pi(d^\sharp)$, and so once again, (10) holds.

From (9) we obtain $\boldsymbol{\sigma}_\pi(d) + \boldsymbol{\sigma}_\pi(d^\sharp) = 4\boldsymbol{\sigma}(d) - 1$. Similarly, we have $\boldsymbol{\tau}_\pi(d) + \boldsymbol{\tau}_\pi(d^\sharp) = 4\boldsymbol{\tau}(d) - 1$. Therefore

$$|\boldsymbol{\sigma}_\pi(d) - \boldsymbol{\tau}_\pi(d) + \boldsymbol{\sigma}_\pi(d^\sharp) - \boldsymbol{\tau}_\pi(d^\sharp)| = 4|\boldsymbol{\sigma}(d) - \boldsymbol{\tau}(d)|. \tag{11}$$

From (10) and (11) we obtain

$$|\boldsymbol{\sigma}_\pi(d) - \boldsymbol{\tau}_\pi(d)| + |\boldsymbol{\sigma}_\pi(d^\sharp) - \boldsymbol{\tau}_\pi(d^\sharp)| = 4|\boldsymbol{\sigma}(d) - \boldsymbol{\tau}(d)|.$$

Hence,

$$\begin{aligned} F(\boldsymbol{\sigma}_\pi, \boldsymbol{\tau}_\pi) &= \sum_{d \in D} (|\boldsymbol{\sigma}_\pi(d) - \boldsymbol{\tau}_\pi(d)| + |\boldsymbol{\sigma}_\pi(d^\sharp) - \boldsymbol{\tau}_\pi(d^\sharp)|) \\ &= \sum_{d \in D} 4|\boldsymbol{\sigma}(d) - \boldsymbol{\tau}(d)| \\ &= 4F_{\text{prof}}(\boldsymbol{\sigma}, \boldsymbol{\tau}). \qquad \square \end{aligned}$$

In the proof of the following lemma, we show that in fact, there is always a full ranking π with no nested elements.

LEMMA 12. *Let $\boldsymbol{\sigma}, \boldsymbol{\tau}$ be partial rankings. Then there exists a full ranking π on the same domain such that $F(\boldsymbol{\sigma}_\pi, \boldsymbol{\tau}_\pi) = 4F_{\text{prof}}(\boldsymbol{\sigma}, \boldsymbol{\tau})$.*

Proof. By Lemma 11, we need only show that there is some full ranking π with no nested elements. Assume that every full ranking has a nested element; we shall derive a contradiction. For a full ranking π, we say that its *first nest* is $\min_d \pi(d)$, where d is allowed to range over all nested elements of π. Choose π to be a full ranking whose first nest is as large as possible.

Let a be the element such that $\pi(a)$ is the first nest of π. By definition, a is nested. Without loss of generality, assume that

$$[\boldsymbol{\sigma}_\pi(a), \boldsymbol{\sigma}_\pi(a^\sharp)] \sqsupset [\boldsymbol{\tau}_\pi(a), \boldsymbol{\tau}_\pi(a^\sharp)]. \tag{12}$$

The intuition behind the proof is the following. We find an element b such that it appears in the left-side interval but not in the right-side interval of (12). We swap a and b in the ordering π and argue that b is not nested in this new ordering. Furthermore, we also argue that no element occurring before a in π becomes nested due to the swap. Hence, we produce a full ranking whose first nest—if it has a nested element at all—is later than the first nest of π, a contradiction. We now proceed with the formal details.

Define the sets S_1 and S_2 as follows:

$$S_1 = \left\{ d \in D \setminus \{a\} \ \ | \ \ [\boldsymbol{\sigma}_\pi(a), \boldsymbol{\sigma}_\pi(a^\sharp)] \sqsupset [\boldsymbol{\sigma}_\pi(d), \boldsymbol{\sigma}_\pi(d^\sharp)] \right\} \text{ and}$$

$$S_2 = \left\{ d \in D \setminus \{a\} \ \ | \ \ [\boldsymbol{\sigma}_\pi(a), \boldsymbol{\sigma}_\pi(a^\sharp)] \sqsupset [\boldsymbol{\tau}_\pi(d), \boldsymbol{\tau}_\pi(d^\sharp)] \right\}.$$

We now show that $S_1 \setminus S_2$ is nonempty. This is because $|S_1| = \frac{1}{2}|[\boldsymbol{\sigma}_\pi(a), \boldsymbol{\sigma}_\pi(a^\sharp)]| - 1$, while $|S_2| \leq \frac{1}{2}|[\boldsymbol{\sigma}_\pi(a), \boldsymbol{\sigma}_\pi(a^\sharp)]| - 2$, since $[\boldsymbol{\sigma}_\pi(a), \boldsymbol{\sigma}_\pi(a^\sharp)] \sqsupset [\boldsymbol{\tau}_\pi(a), \boldsymbol{\tau}_\pi(a^\sharp)]$ but a is not counted in S_2. Choose b in $S_1 \setminus S_2$. Note that the fact that $b \in S_1$ implies that a and b are in the same bucket for $\boldsymbol{\sigma}$. It further implies that $\pi(a) < \pi(b)$.

We now show that a and b are in different buckets for $\boldsymbol{\tau}$. Suppose that a and b were in the same bucket for $\boldsymbol{\tau}$. Then since $\pi(a) < \pi(b)$, we would have $\boldsymbol{\tau}_\pi(a) < \boldsymbol{\tau}_\pi(b)$ and $\boldsymbol{\tau}_\pi(a^\sharp) > \boldsymbol{\tau}_\pi(b^\sharp)$. That is, $[\boldsymbol{\tau}_\pi(a), \boldsymbol{\tau}_\pi(a^\sharp)] \sqsupset [\boldsymbol{\tau}_\pi(b), \boldsymbol{\tau}_\pi(b^\sharp)]$. If we combine this fact with (12), we obtain $[\boldsymbol{\sigma}_\pi(a), \boldsymbol{\sigma}_\pi(a^\sharp)] \sqsupset [\boldsymbol{\tau}_\pi(a), \boldsymbol{\tau}_\pi(a^\sharp)] \sqsupset [\boldsymbol{\tau}_\pi(b), \boldsymbol{\tau}_\pi(b^\sharp)]$. This contradicts the fact that $b \notin S_2$. Hence, a and b must be in different buckets for $\boldsymbol{\tau}$.

Now, produce π' by swapping a and b in π. Since $\pi(a) < \pi(b)$, we see that $\pi'(b) = \pi(a) < \pi(b) = \pi'(a)$. We wish to prove that the first nest for π'—if it has a nested element at all—is larger than the first nest for π, which gives our desired contradiction. We do so by showing that b is unnested for π' and further, that d is unnested for π' for all d in D such that $\pi'(d) < \pi'(b)$. In order to prove this, we need to examine the effect of swapping a and b in π.

We first consider $\boldsymbol{\sigma}$. We know that a and b appear in the same bucket of $\boldsymbol{\sigma}$. Let $\mathcal{B}_{ab}$ be the bucket of $\boldsymbol{\sigma}$ that contains both a and b. Swapping a and b in π has the effect of swapping the positions of a and b in $\boldsymbol{\sigma}_\pi$ (so in particular $\boldsymbol{\sigma}_{\pi'}(b) = \boldsymbol{\sigma}_\pi(a)$), swapping the positions of $a^\sharp$ and $b^\sharp$ in $\boldsymbol{\sigma}_\pi$ (so in particular $\boldsymbol{\sigma}_{\pi'}(b^\sharp) = \boldsymbol{\sigma}_\pi(a^\sharp)$) and leaving all other elements d and $d^\sharp$ in $\mathcal{B}_{ab}$ in the same place (so $\boldsymbol{\sigma}_\pi(d) = \boldsymbol{\sigma}_{\pi'}(d)$ and $\boldsymbol{\sigma}_\pi(d^\sharp) = \boldsymbol{\sigma}_{\pi'}(d^\sharp)$). Since $\boldsymbol{\sigma}_{\pi'}(b) = \boldsymbol{\sigma}_\pi(a)$ and $\boldsymbol{\sigma}_{\pi'}(b^\sharp) = \boldsymbol{\sigma}_\pi(a^\sharp)$, and since two closed intervals of numbers are equal precisely if their left endpoints and their right endpoints are equal, we have

$$[\boldsymbol{\sigma}_{\pi'}(b), \boldsymbol{\sigma}_{\pi'}(b^\sharp)] = [\boldsymbol{\sigma}_\pi(a), \boldsymbol{\sigma}_\pi(a^\sharp)]. \tag{13}$$

Now, let $\mathcal{B}$ be a bucket of $\boldsymbol{\sigma}$ other than $\mathcal{B}_{ab}$. Then swapping a and b in π has no effect (as far as $\boldsymbol{\sigma}_\pi$ is concerned) on the elements in $\mathcal{B}$, since the relative order of all elements in $\mathcal{B}$ is precisely the same with or without the swap. That is, $\boldsymbol{\sigma}_\pi(d) = \boldsymbol{\sigma}_{\pi'}(d)$ and $\boldsymbol{\sigma}_\pi(d^\sharp) = \boldsymbol{\sigma}_{\pi'}(d^\sharp)$ for all d in $\mathcal{B}$. But we noted earlier that these same two equalities hold for all elements d in $\mathcal{B}_{ab}$ other than a and b. Therefore, for all elements d other than a or b (whether or not these elements are in $\mathcal{B}_{ab}$), we have

$$[\boldsymbol{\sigma}_{\pi'}(d), \boldsymbol{\sigma}_{\pi'}(d^\sharp)] = [\boldsymbol{\sigma}_\pi(d), \boldsymbol{\sigma}_\pi(d^\sharp)]. \tag{14}$$

We now consider $\boldsymbol{\tau}$. We know that a and b appear in different buckets of $\boldsymbol{\tau}$. Let $\mathcal{B}$ be a bucket of $\boldsymbol{\tau}$ containing neither a nor b (if there is such a bucket). As with $\boldsymbol{\sigma}$, we see that elements in $\mathcal{B}$ are unaffected by swapping a and b in π. That is, $\boldsymbol{\tau}_\pi(d) = \boldsymbol{\tau}_{\pi'}(d)$ and $\boldsymbol{\tau}_\pi(d^\sharp) = \boldsymbol{\tau}_{\pi'}(d^\sharp)$ for all d in $\mathcal{B}$.

Now, let $\mathcal{B}_a$ be the bucket of $\boldsymbol{\tau}$ containing a (but not b). Notice that for all d in $\mathcal{B}_a$ such that $\pi(d) < \pi(a)$, we have $\pi(d) = \pi'(d)$. Hence, the relative order among these most highly ranked elements of $\mathcal{B}_a$ remains the same. Therefore, $\boldsymbol{\tau}_\pi(d) = \boldsymbol{\tau}_{\pi'}(d)$ and $\boldsymbol{\tau}_\pi(d^\sharp) = \boldsymbol{\tau}_{\pi'}(d^\sharp)$ for all d in $\mathcal{B}_a$ such that $\pi(d) < \pi(a)$. Furthermore, $\pi'(a) > \pi(a)$, and so a is still ranked after all the aforementioned d's in $\boldsymbol{\tau}_{\pi'}$. Hence, $\boldsymbol{\tau}_\pi(a) \leq \boldsymbol{\tau}_{\pi'}(a)$ and $\boldsymbol{\tau}_\pi(a^\sharp) \geq \boldsymbol{\tau}_{\pi'}(a^\sharp)$. That is,

$$[\boldsymbol{\tau}_{\pi'}(a), \boldsymbol{\tau}_{\pi'}(a^\sharp)] \subseteq [\boldsymbol{\tau}_\pi(a), \boldsymbol{\tau}_\pi(a^\sharp)]. \tag{15}$$

Finally, let $\mathcal{B}_b$ be the bucket of $\boldsymbol{\tau}$ that contains b (but not a). As before, for all d in $\mathcal{B}_b$ such that $\pi(d) < \pi(a)$, we have $\pi(d) = \pi'(d)$. Hence, the relative order among these most highly ranked elements of $\mathcal{B}_b$ remains the same. Therefore, $\boldsymbol{\tau}_\pi(d) = \boldsymbol{\tau}_{\pi'}(d)$

and $\boldsymbol{\tau}_\pi(d^\sharp) = \boldsymbol{\tau}_{\pi'}(d^\sharp)$ for all d in $\mathcal{B}_b$ such that $\pi(d) < \pi(a)$. That is, for every d such that $\pi(d) < \pi(a)$ (i.e., every d such that $\pi'(d) < \pi'(b)$), we have

$$[\boldsymbol{\tau}_{\pi'}(d), \boldsymbol{\tau}_{\pi'}(d^\sharp)] = [\boldsymbol{\tau}_\pi(d), \boldsymbol{\tau}_\pi(d^\sharp)]. \tag{16}$$

Furthermore, $\pi'(b) < \pi(b)$, and so b is still ranked before all d' in $\mathcal{B}_b$ such that $\pi(b) < \pi(d') = \pi'(d')$. Hence, $\boldsymbol{\tau}_\pi(b) \geq \boldsymbol{\tau}_{\pi'}(b)$ and $\boldsymbol{\tau}_\pi(b^\sharp) \leq \boldsymbol{\tau}_{\pi'}(b^\sharp)$. That is,

$$[\boldsymbol{\tau}_{\pi'}(b), \boldsymbol{\tau}_{\pi'}(b^\sharp)] \supseteq [\boldsymbol{\tau}_\pi(b), \boldsymbol{\tau}_\pi(b^\sharp)]. \tag{17}$$

From (14) and (16), we see that d remains unnested for all d such that $\pi'(d) < \pi'(b)$. So we need only show that b is unnested for π' to finish the proof. If b were nested for π', then either $[\boldsymbol{\sigma}_{\pi'}(b), \boldsymbol{\sigma}_{\pi'}(b^\sharp)] \sqsupset [\boldsymbol{\tau}_{\pi'}(b), \boldsymbol{\tau}_{\pi'}(b^\sharp)]$ or $[\boldsymbol{\tau}_{\pi'}(b), \boldsymbol{\tau}_{\pi'}(b^\sharp)] \sqsupset [\boldsymbol{\sigma}_{\pi'}(b), \boldsymbol{\sigma}_{\pi'}(b^\sharp)]$. First, suppose that $[\boldsymbol{\sigma}_{\pi'}(b), \boldsymbol{\sigma}_{\pi'}(b^\sharp)] \sqsupset [\boldsymbol{\tau}_{\pi'}(b), \boldsymbol{\tau}_{\pi'}(b^\sharp)]$. Then

$$\begin{aligned}
[\boldsymbol{\sigma}_\pi(a), \boldsymbol{\sigma}_\pi(a^\sharp)] &= [\boldsymbol{\sigma}_{\pi'}(b), \boldsymbol{\sigma}_{\pi'}(b^\sharp)] \quad \text{by (13)} \\
&\sqsupset [\boldsymbol{\tau}_{\pi'}(b), \boldsymbol{\tau}_{\pi'}(b^\sharp)] \quad \text{by supposition} \\
&\supseteq [\boldsymbol{\tau}_\pi(b), \boldsymbol{\tau}_\pi(b^\sharp)] \quad \text{by (17).}
\end{aligned}$$

But this contradicts the fact that $b \notin S_2$. Now, suppose that $[\boldsymbol{\tau}_{\pi'}(b), \boldsymbol{\tau}_{\pi'}(b^\sharp)] \sqsupset [\boldsymbol{\sigma}_{\pi'}(b), \boldsymbol{\sigma}_{\pi'}(b^\sharp)]$. Then

$$\begin{aligned}
[\boldsymbol{\tau}_{\pi'}(b), \boldsymbol{\tau}_{\pi'}(b^\sharp)] &\sqsupset [\boldsymbol{\sigma}_{\pi'}(b), \boldsymbol{\sigma}_{\pi'}(b^\sharp)] \quad \text{by supposition} \\
&= [\boldsymbol{\sigma}_\pi(a), \boldsymbol{\sigma}_\pi(a^\sharp)] \quad \text{by (13)} \\
&\sqsupset [\boldsymbol{\tau}_\pi(a), \boldsymbol{\tau}_\pi(a^\sharp)] \quad \text{by (12)} \\
&\supseteq [\boldsymbol{\tau}_{\pi'}(a), \boldsymbol{\tau}_{\pi'}(a^\sharp)] \quad \text{by (15).}
\end{aligned}$$

But this implies that a and b are in the same bucket for $\boldsymbol{\tau}$, a contradiction. Hence, b must not be nested for π', which was to be shown. ∎

We can now prove our desired theorem that F_{prof} and K_{prof} are equivalent.

THEOREM 13. *Let $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$ be partial rankings. Then $K_{\text{prof}}(\boldsymbol{\sigma}, \boldsymbol{\tau}) \leq F_{\text{prof}}(\boldsymbol{\sigma}, \boldsymbol{\tau}) \leq 2K_{\text{prof}}(\boldsymbol{\sigma}, \boldsymbol{\tau})$.*

Proof. Given $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$, let π be the full ranking guaranteed by Lemma 12. Then we have

$$\begin{aligned}
K_{\text{prof}}(\boldsymbol{\sigma}, \boldsymbol{\tau}) &= 4K(\boldsymbol{\sigma}_\pi, \boldsymbol{\tau}_\pi) \quad \text{by Lemma 10} \\
&\leq 4F(\boldsymbol{\sigma}_\pi, \boldsymbol{\tau}_\pi) \quad \text{by (1)} \\
&= F_{\text{prof}}(\boldsymbol{\sigma}, \boldsymbol{\tau}) \quad \text{by Lemma 12.}
\end{aligned}$$

And similarly,

$$\begin{aligned}
F_{\text{prof}}(\boldsymbol{\sigma}, \boldsymbol{\tau}) &= 4F(\boldsymbol{\sigma}_\pi, \boldsymbol{\tau}_\pi) \quad \text{by Lemma 12} \\
&\leq 8K(\boldsymbol{\sigma}_\pi, \boldsymbol{\tau}_\pi) \quad \text{by (1)} \\
&= 2K_{\text{prof}}(\boldsymbol{\sigma}, \boldsymbol{\tau}) \quad \text{by Lemma 10.}
\end{aligned}$$ ∎

4.3. Equivalence of K_{Haus} and K_{prof}. We now prove (6), which is the final step in proving Theorem 7.

THEOREM 14. *Let $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$ be partial rankings. Then $K_{\text{prof}}(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2) \leq K_{\text{Haus}}(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2) \leq 2K_{\text{prof}}(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2)$.*

Proof. As in Theorem 6 (where we let $\boldsymbol{\sigma}_1$ play the role of $\boldsymbol{\sigma}$, and let $\boldsymbol{\sigma}_2$ play the role of $\boldsymbol{\tau}$), let S be the set of pairs $\{i,j\}$ of distinct elements such that i and j appear in the same bucket of $\boldsymbol{\sigma}_1$ but in different buckets of $\boldsymbol{\sigma}_2$, let T be the set of pairs $\{i,j\}$ of distinct elements such that i and j appear in the same bucket of $\boldsymbol{\sigma}_2$ but in different buckets of $\boldsymbol{\sigma}_1$, and let U be the set of pairs $\{i,j\}$ of distinct elements that are in different buckets of both $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$ and are in a different order in $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$. By Theorem 6, we know that $K_{\mathrm{Haus}}(\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2) = |U| + \max\{|S|,|T|\}$. It follows from the definition of K_{prof} that $K_{\mathrm{prof}}(\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2) = |U| + \frac{1}{2}|S| + \frac{1}{2}|T|$. The theorem now follows from the straightforward inequalities $|U|+\frac{1}{2}|S|+\frac{1}{2}|T| \le |U|+\max\{|S|,|T|\} \le 2(|U|+\frac{1}{2}|S|+\frac{1}{2}|T|)$. □

This concludes the proof that all our metrics are equivalent.

5. An alternative representation. Let $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}'$ be partial rankings. Assume that the buckets of $\boldsymbol{\sigma}$ are, in order, $\mathcal{B}_1,\ldots,\mathcal{B}_t$, and the buckets of $\boldsymbol{\sigma}'$ are, in order, $\mathcal{B}'_1,\ldots,\mathcal{B}'_{t'}$. Critchlow [9] defines n_{ij} (for $1 \le i \le t$ and $1 \le j \le t'$) to be $|\mathcal{B}_i \cap \mathcal{B}'_j|$. His main theorem gives formulas for $K_{\mathrm{Haus}}(\boldsymbol{\sigma},\boldsymbol{\sigma}')$ and $F_{\mathrm{Haus}}(\boldsymbol{\sigma},\boldsymbol{\sigma}')$ (and for other Hausdorff measures) in terms of the n_{ij}'s. His formula for $K_{\mathrm{Haus}}(\boldsymbol{\sigma},\boldsymbol{\sigma}')$ is particularly simple, and is given by the following theorem.

THEOREM 15 (see [9]). *Let $\boldsymbol{\sigma}$, $\boldsymbol{\sigma}'$, and the n_{ij}'s be as above. Then*

$$K_{\mathrm{Haus}}(\boldsymbol{\sigma},\boldsymbol{\sigma}') = \max\left\{ \sum_{i<i',\ \ j\ge j'} n_{ij}n_{i'j'},\ \sum_{i\le i',\ \ j>j'} n_{ij}n_{i'j'} \right\}.$$

It is straightforward to derive Theorem 6 from Theorem 15, and to derive Theorem 15 from Theorem 6, by using the simple fact that if S,T,U are as in Theorem 6, then

$$|U| = \sum_{i<i',\ \ j>j'} n_{ij}n_{i'j'},$$

$$|S| = \sum_{i=i',\ \ j>j'} n_{ij}n_{i'j'},$$

$$|T| = \sum_{i<i',\ \ j=j'} n_{ij}n_{i'j'}.$$

Let us define the *Critchlow profile* of the pair $(\boldsymbol{\sigma},\boldsymbol{\sigma}')$ to be a $t \times t'$ matrix, where t is the number of buckets of $\boldsymbol{\sigma}$, t' is the number of buckets of $\boldsymbol{\sigma}'$, and the (i,j)th entry is n_{ij}. We noted that Critchlow gives formulas for $K_{\mathrm{Haus}}(\boldsymbol{\sigma},\boldsymbol{\sigma}')$ and $F_{\mathrm{Haus}}(\boldsymbol{\sigma},\boldsymbol{\sigma}')$ in terms of the Critchlow profile. The reader may find it surprising that the Critchlow profile contains enough information to compute $K_{\mathrm{Haus}}(\boldsymbol{\sigma},\boldsymbol{\sigma}')$ and $F_{\mathrm{Haus}}(\boldsymbol{\sigma},\boldsymbol{\sigma}')$. The following theorem implies that this "surprise" is true not just about K_{Haus} and F_{Haus}, but about every function d (not even necessarily a metric) whose arguments are a pair of partial rankings, as long as d is "name-independent" (that is, the answer is the same when we rename the elements). Before we state the theorem, we need some more terminology. The theorem says that the Critchlow profile "uniquely determines $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}'$, up to renaming of the elements." What this means is that if $(\boldsymbol{\sigma},\boldsymbol{\sigma}')$ has the same Critchlow profile as $(\boldsymbol{\tau},\boldsymbol{\tau}')$, then the pair $(\boldsymbol{\sigma},\boldsymbol{\sigma}')$ is isomorphic to the pair $(\boldsymbol{\tau},\boldsymbol{\tau}')$. That is, there is a one-to-one function f from the common domain D onto itself such that $\boldsymbol{\sigma}(i) = \boldsymbol{\tau}(f(i))$ and $\boldsymbol{\sigma}'(i) = \boldsymbol{\tau}'(f(i))$ for every i in D. Intuitively, the pair $(\boldsymbol{\tau},\boldsymbol{\tau}')$ is obtained from the pair $(\boldsymbol{\sigma},\boldsymbol{\sigma}')$ by the renaming function f.

THEOREM 16. *The Critchlow profile uniquely determines $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}'$, up to renaming of the elements.*

Proof. We first give an informal proof. The only relevant information about an element is which $\mathcal{B}_i$ it is in and which $\mathcal{B}'_j$ it is in. So the only information that matters about the pair $\boldsymbol{\sigma}, \boldsymbol{\sigma}'$ of partial rankings is, for each i, j, how many elements are in $\mathcal{B}_i \cap \mathcal{B}'_j$. That is, we can reconstruct $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}'$, up to renaming of the elements, by knowing only the Critchlow profile.

More formally, let $(\boldsymbol{\sigma}, \boldsymbol{\sigma}')$ and $(\boldsymbol{\tau}, \boldsymbol{\tau}')$ each be pairs of partial rankings with the same Critchlow profile. That is, assume that the buckets of $\boldsymbol{\sigma}$ are, in order, $\mathcal{B}_1, \dots, \mathcal{B}_t$, the buckets of $\boldsymbol{\sigma}'$ are, in order, $\mathcal{B}'_1, \dots, \mathcal{B}'_{t'}$, the buckets of $\boldsymbol{\tau}$ are, in order, $\mathcal{C}_1, \dots, \mathcal{C}_t$, and the buckets of $\boldsymbol{\tau}'$ are, in order, $\mathcal{C}'_1, \dots, \mathcal{C}_{t'}$, where $|\mathcal{B}_i \cap B'_j| = |\mathcal{C}_i \cap C'_j|$ for each i, j. (Note that the number t of buckets of $\boldsymbol{\sigma}$ is the same as the number of buckets of $\boldsymbol{\tau}$, and similarly the number t' of buckets of $\boldsymbol{\sigma}'$ is the same as the number of buckets of $\boldsymbol{\tau}'$; this follows from the assumption that $(\boldsymbol{\sigma}, \boldsymbol{\sigma}')$ and $(\boldsymbol{\tau}, \boldsymbol{\tau}')$ have the same Critchlow profile.) Let f_{ij} be a one-to-one mapping of $\mathcal{B}_i \cap \mathcal{B}'_j$ onto $\mathcal{C}_i \cap \mathcal{C}'_j$ (such an f_{ij} exists because $|\mathcal{B}_i \cap \mathcal{B}'_j| = |\mathcal{C}_i \cap \mathcal{C}'_j|$). Let f be the function obtained by taking the union of the functions f_{ij} (we think of functions as sets of ordered pairs, so it is proper to take the union). It is easy to see that $(\boldsymbol{\sigma}, \boldsymbol{\sigma}')$ and $(\boldsymbol{\tau}, \boldsymbol{\tau}')$ are isomorphic under the isomorphism f. This proves the theorem. □

The Critchlow profile differs in several ways from the K-profile and the F-profile, as defined in section 3.1. First, the K-profile and the F-profile are each profiles of a single partial ranking, whereas the Critchlow profile is a profile of a pair of partial rankings. Second, from the K-profile of $\boldsymbol{\sigma}$ we can completely reconstruct $\boldsymbol{\sigma}$ (not just up to renaming of elements, but completely), and a similar comment applies to the F-profile. On the other hand, from the Critchlow profile we can reconstruct the pair $(\boldsymbol{\sigma}, \boldsymbol{\sigma}')$ only up to a renaming of elements. Thus, the Critchlow profile "loses information," whereas the K-profile and F-profile do not.

6. Conclusions. In this paper we consider metrics between partial rankings. We define four natural metrics between partial rankings. We obtain efficient polynomial time algorithms to compute these metrics. We also show that these metrics are all within constant multiples of each other.

Appendix. Proof of Theorem 5. In this appendix, we prove Theorem 5. First, we state a fact that we use several times.

LEMMA 17. *Suppose $a \le b$ and $c \le d$. Then $|a-c| + |b-d| \le |a-d| + |b-c|$.*

Proof. To see this, first note that by symmetry, we can assume, without loss of generality, that $a \le c$. Now there are three cases: $a \le b \le c \le d$, $a \le c \le b \le d$, and $a \le c \le d \le b$. In the first case (when $a \le b \le c \le d$), it is easy to verify that both the left-hand side and the right-hand side of the inequality equal $|a-b| + 2|b-c| + |c-d|$, and so the left-hand side and the right-hand side are equal. In both the second case (when $a \le c \le b \le d$) and the third case (when $a \le c \le d \le b$), it is easy to verify that the right-hand side equals $|a-c| + 2|b-c| + |b-d|$, which exceeds the left-hand side by $2|b-c|$. □

We next show a simple lemma.

LEMMA 18. *Let π be a full ranking, and let $\boldsymbol{\sigma}$ be a partial ranking. Suppose that $\pi \neq \boldsymbol{\sigma}$. Then there exist i, j such that $\pi(j) = \pi(i) + 1$ while $\boldsymbol{\sigma}(j) \le \boldsymbol{\sigma}(i)$. If $\boldsymbol{\sigma}$ is in fact a full ranking, then $\boldsymbol{\sigma}(j) < \boldsymbol{\sigma}(i)$.*

Proof. For each m with $1 \le m \le |D|$, let d_m be the member of the domain D, where $\pi(d_m) = m$. Thus, $D = \{d_1, \dots, d_{|D|}\}$ and $\pi(d_1) < \pi(d_2) < \cdots < \pi(d_{|D|})$. If $\boldsymbol{\sigma}(d_\ell) < \boldsymbol{\sigma}(d_{\ell+1})$ for all ℓ, then we would have $K_{\text{prof}}(\boldsymbol{\sigma}, \pi) = 0$, contradicting the fact that $\pi \neq \boldsymbol{\sigma}$. Hence, there must be some ℓ for which $\boldsymbol{\sigma}(d_{\ell+1}) \leq \boldsymbol{\sigma}(d_\ell)$. Setting $i = d_\ell$ and $j = d_{\ell+1}$ gives us the lemma.

If $\boldsymbol{\sigma}$ is a full ranking, then $\boldsymbol{\sigma}(j) \neq \boldsymbol{\sigma}(i)$, showing $\boldsymbol{\sigma}(j) < \boldsymbol{\sigma}(i)$. ∎

The next two lemmas will be helpful in obtaining a characterization of the Hausdorff distance.

LEMMA 19. *Let σ be a full ranking, and let $\boldsymbol{\tau}$ be a partial ranking. Then the quantity $F(\sigma, \tau)$, taken over all full refinements $\tau \succeq \boldsymbol{\tau}$, is minimized for $\tau = \sigma * \boldsymbol{\tau}$. Similarly, the quantity $K(\sigma, \tau)$, taken over all full refinements $\tau \succeq \boldsymbol{\tau}$, is minimized for $\tau = \sigma * \boldsymbol{\tau}$.*

Proof. First, note that if τ is a full ranking with $\tau \succeq \boldsymbol{\tau}$, then there is a full ranking π such that $\tau = \boldsymbol{\tau} * \pi$. We show that $F(\sigma, \sigma * \boldsymbol{\tau}) \leq F(\sigma, \pi * \boldsymbol{\tau})$ and $K(\sigma, \sigma * \boldsymbol{\tau}) \leq K(\sigma, \pi * \boldsymbol{\tau})$ for every full ranking π. The lemma will then follow. Let

$$U = \{\pi \mid \pi \text{ is a full ranking and } F(\sigma, \sigma * \boldsymbol{\tau}) > F(\sigma, \pi * \boldsymbol{\tau})\},$$

$$V = \{\pi \mid \pi \text{ is a full ranking and } K(\sigma, \sigma * \boldsymbol{\tau}) > K(\sigma, \pi * \boldsymbol{\tau})\},$$

and let $S = U \cup V$. If S is empty, then we are done. So suppose not; we derive a contradiction. Over all full rankings $\pi \in S$, choose π to be a full ranking that minimizes $K(\sigma, \pi)$. In other words, choose a full ranking in S that is as close to σ as possible, according to the Kendall distance.

Clearly $\sigma \notin S$, and so $\pi \neq \sigma$ (since $\pi \in S$). Since $\pi \neq \sigma$, Lemma 18 guarantees that we can find a pair i, j such that $\pi(j) = \pi(i) + 1$, but $\sigma(j) < \sigma(i)$. Produce π' by swapping i and j in π. Clearly, π' has one fewer inversion with respect to σ than π does. Hence, $K(\sigma, \pi') < K(\sigma, \pi)$. If we can show that $\pi' \in S$, then we obtain our desired contradiction, since π is the full ranking in S that minimizes $K(\sigma, \pi)$. So we need only show that $\pi' \in S$.

If i and j are in different buckets for $\boldsymbol{\tau}$, then $\pi' * \boldsymbol{\tau} = \pi * \boldsymbol{\tau}$. Hence, $F(\sigma, \pi' * \boldsymbol{\tau}) = F(\sigma, \pi * \boldsymbol{\tau})$ and $K(\sigma, \pi' * \boldsymbol{\tau}) = K(\sigma, \pi * \boldsymbol{\tau})$. So if $\pi \in U$, then $\pi' \in U$, and if $\pi \in V$, then $\pi' \in V$. In either case, $\pi' \in S$, and we are done.

On the other hand, assume that i and j are in the same bucket for $\boldsymbol{\tau}$. Then $\pi' * \boldsymbol{\tau}(i) = \pi * \boldsymbol{\tau}(j)$ and $\pi' * \boldsymbol{\tau}(j) = \pi * \boldsymbol{\tau}(i)$. Furthermore, since $\pi(i) < \pi(j)$ and i and j are in the same bucket for $\boldsymbol{\tau}$, we have $\pi * \boldsymbol{\tau}(i) < \pi * \boldsymbol{\tau}(j)$, while $\sigma(j) < \sigma(i)$.

Either $\pi \in U$ or $\pi \in V$. First, consider the case where $\pi \in U$. We have

$$\begin{aligned} &|\pi' * \boldsymbol{\tau}(j) - \sigma(j)| + |\pi' * \boldsymbol{\tau}(i) - \sigma(i)| \\ &= |\pi * \boldsymbol{\tau}(i) - \sigma(j)| + |\pi * \boldsymbol{\tau}(j) - \sigma(i)| \\ &\leq |\pi * \boldsymbol{\tau}(i) - \sigma(i)| + |\pi * \boldsymbol{\tau}(j) - \sigma(j)|, \end{aligned} \tag{18}$$

where the inequality follows from Lemma 17 with $a = \pi * \boldsymbol{\tau}(i)$, $b = \pi * \boldsymbol{\tau}(j)$, $c = \sigma(j)$, and $d = \sigma(i)$. We also have $|\pi' * \boldsymbol{\tau}(d) - \sigma(d)| = |\pi * \boldsymbol{\tau}(d) - \sigma(d)|$ for all $d \in D \setminus \{i, j\}$, since $\pi' * \boldsymbol{\tau}$ and $\pi * \boldsymbol{\tau}$ agree everywhere but at i and j. If we sum over all d (where we make use of (18) for $d = i$ and $d = j$), we obtain $F(\sigma, \pi' * \boldsymbol{\tau}) \leq F(\sigma, \pi * \boldsymbol{\tau})$. Since $\pi \in U$, we have $F(\sigma, \pi * \boldsymbol{\tau}) < F(\sigma, \sigma * \boldsymbol{\tau})$. Combining these last two inequalities, we obtain $F(\sigma, \pi' * \boldsymbol{\tau}) < F(\sigma, \sigma * \boldsymbol{\tau})$. Therefore, $\pi' \in U$, and so $\pi' \in S$, which was to be shown.

Now consider the case where $\pi \in V$. Since $\pi(j) = \pi(i) + 1$ and since i and j are in the same bucket of $\boldsymbol{\tau}$, we have $\pi * \boldsymbol{\tau}(j) = \pi * \boldsymbol{\tau}(i) + 1$. Similarly, $\pi' * \boldsymbol{\tau}(i) = \pi' * \boldsymbol{\tau}(j) + 1$. And as we noted earlier, $\pi * \boldsymbol{\tau}$ and $\pi' * \boldsymbol{\tau}$ agree everywhere except at i and j. In other words, $\pi' * \boldsymbol{\tau}$ is just $\pi * \boldsymbol{\tau}$, with the *adjacent* elements i and j swapped. Since

$\sigma(i) > \sigma(j)$ we see that $\pi' * \boldsymbol{\tau}$ has exactly one fewer inversion with respect to σ than $\pi * \boldsymbol{\tau}$ does. Hence, $K(\sigma, \pi' * \boldsymbol{\tau}) < K(\sigma, \pi * \boldsymbol{\tau})$. Since $\pi \in V$, we have $K(\sigma, \pi * \boldsymbol{\tau}) < K(\sigma, \sigma * \boldsymbol{\tau})$. Combining these last two inequalities, we obtain $K(\sigma, \pi' * \boldsymbol{\tau}) < K(\sigma, \sigma * \boldsymbol{\tau})$. Therefore, $\pi' \in V$, and so $\pi' \in S$, which was to be shown. □

LEMMA 20. *Let $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$ be partial rankings, and let ρ be any full ranking. Then the quantity $F(\sigma, \sigma * \boldsymbol{\tau})$, taken over all full refinements $\sigma \succeq \boldsymbol{\sigma}$, is maximized when $\sigma = \rho * \boldsymbol{\tau}^{\mathrm{R}} * \boldsymbol{\sigma}$. Similarly, the quantity $K(\sigma, \sigma * \boldsymbol{\tau})$, taken over all full refinements $\sigma \succeq \boldsymbol{\sigma}$, is maximized when $\sigma = \rho * \boldsymbol{\tau}^{\mathrm{R}} * \boldsymbol{\sigma}$.*

Proof. First, note that for any full refinement $\sigma \succeq \boldsymbol{\sigma}$, there is some full ranking π such that $\sigma = \pi * \boldsymbol{\sigma}$. We show that for all full rankings π,

$$\begin{aligned} F(\rho * \boldsymbol{\tau}^{\mathrm{R}} * \boldsymbol{\sigma}, \rho * \boldsymbol{\tau}^{\mathrm{R}} * \boldsymbol{\sigma} * \boldsymbol{\tau}) &\geq F(\pi * \boldsymbol{\sigma}, \pi * \boldsymbol{\sigma} * \boldsymbol{\tau}) \\ \text{and } K(\rho * \boldsymbol{\tau}^{\mathrm{R}} * \boldsymbol{\sigma}, \rho * \boldsymbol{\tau}^{\mathrm{R}} * \boldsymbol{\sigma} * \boldsymbol{\tau}) &\geq K(\pi * \boldsymbol{\sigma}, \pi * \boldsymbol{\sigma} * \boldsymbol{\tau}). \end{aligned}$$

The lemma will then follow.

Let $U = \{\text{full } \pi \mid F(\rho * \boldsymbol{\tau}^{\mathrm{R}} * \boldsymbol{\sigma}, \rho * \boldsymbol{\tau}^{\mathrm{R}} * \boldsymbol{\sigma} * \boldsymbol{\tau}) < F(\pi * \boldsymbol{\sigma}, \pi * \boldsymbol{\sigma} * \boldsymbol{\tau})\}$, let $V = \{\text{full } \pi \mid K(\rho * \boldsymbol{\tau}^{\mathrm{R}} * \boldsymbol{\sigma}, \rho * \boldsymbol{\tau}^{\mathrm{R}} * \boldsymbol{\sigma} * \boldsymbol{\tau}) < K(\pi * \boldsymbol{\sigma}, \pi * \boldsymbol{\sigma} * \boldsymbol{\tau})\}$, and let $S = U \cup V$. If S is empty, then we are done. So suppose not; we derive a contradiction. Over all full rankings $\pi \in S$, choose π to be the full ranking that minimizes $K(\rho * \boldsymbol{\tau}^{\mathrm{R}}, \pi)$.

Clearly $\rho * \boldsymbol{\tau}^{\mathrm{R}} \notin S$, and so $\pi \neq \rho * \boldsymbol{\tau}^{\mathrm{R}}$ (since $\pi \in S$). Since $\pi \neq \rho * \boldsymbol{\tau}^{\mathrm{R}}$, Lemma 18 guarantees that we can find a pair i, j such that $\pi(j) = \pi(i) + 1$, but $\rho * \boldsymbol{\tau}^{\mathrm{R}}(j) < \rho * \boldsymbol{\tau}^{\mathrm{R}}(i)$. Produce π' by swapping i and j. Clearly, π' has one fewer inversion with respect to $\rho * \boldsymbol{\tau}^{\mathrm{R}}$ than π does. Hence, $K(\rho * \boldsymbol{\tau}^{\mathrm{R}}, \pi') < K(\rho * \boldsymbol{\tau}^{\mathrm{R}}, \pi)$. We now show that $\pi' \in S$, producing a contradiction.

If i and j are in different buckets for $\boldsymbol{\sigma}$, then $\pi' * \boldsymbol{\sigma} = \pi * \boldsymbol{\sigma}$. Hence, $F(\pi' * \boldsymbol{\sigma}, \pi' * \boldsymbol{\sigma} * \boldsymbol{\tau}) = F(\pi * \boldsymbol{\sigma}, \pi * \boldsymbol{\sigma} * \boldsymbol{\tau})$ and $K(\pi' * \boldsymbol{\sigma}, \pi' * \boldsymbol{\sigma} * \boldsymbol{\tau}) = K(\pi * \boldsymbol{\sigma}, \pi * \boldsymbol{\sigma} * \boldsymbol{\tau})$. So if $\pi \in U$, then $\pi' \in U$, and if $\pi \in V$, then $\pi' \in V$. In either case, $\pi' \in S$, and we are done.

Likewise, if i and j are in the same bucket for both $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$, then swapping i and j in π swaps their positions in both $\pi * \boldsymbol{\sigma} * \boldsymbol{\tau}$ and $\pi * \boldsymbol{\sigma}$ and leaves all other elements in their same positions in both $\pi * \boldsymbol{\sigma} * \boldsymbol{\tau}$ and $\pi * \boldsymbol{\sigma}$. So again, we see $F(\pi' * \boldsymbol{\sigma}, \pi' * \boldsymbol{\sigma} * \boldsymbol{\tau}) = F(\pi * \boldsymbol{\sigma}, \pi * \boldsymbol{\sigma} * \boldsymbol{\tau})$ and $K(\pi' * \boldsymbol{\sigma}, \pi' * \boldsymbol{\sigma} * \boldsymbol{\tau}) = K(\pi * \boldsymbol{\sigma}, \pi * \boldsymbol{\sigma} * \boldsymbol{\tau})$. As before, $\pi' \in S$.

The only remaining situation is when i and j are in the same bucket for $\boldsymbol{\sigma}$, but in different buckets for $\boldsymbol{\tau}$. Let us consider this situation. First of all, $\pi' * \boldsymbol{\sigma}$ is just $\pi * \boldsymbol{\sigma}$ with the adjacent elements i and j swapped, since i and j are in the same bucket for $\boldsymbol{\sigma}$. Second, $\pi' * \boldsymbol{\sigma} * \boldsymbol{\tau} = \pi * \boldsymbol{\sigma} * \boldsymbol{\tau}$ since i and j are in different buckets for $\boldsymbol{\tau}$.

Since $\pi(i) < \pi(j)$, we have $\pi * \boldsymbol{\sigma}(i) < \pi * \boldsymbol{\sigma}(j)$. Further, $\boldsymbol{\tau}(i) < \boldsymbol{\tau}(j)$ since $\rho * \boldsymbol{\tau}^{\mathrm{R}}(j) < \rho * \boldsymbol{\tau}^{\mathrm{R}}(i)$ and $\rho * \boldsymbol{\tau}^{\mathrm{R}}$ is a refinement of the reverse of $\boldsymbol{\tau}$. Since $\boldsymbol{\tau}(i) < \boldsymbol{\tau}(j)$, we have $\pi * \boldsymbol{\sigma} * \boldsymbol{\tau}(i) < \pi * \boldsymbol{\sigma} * \boldsymbol{\tau}(j)$.

Either $\pi \in U$ or $\pi \in V$. Let us first examine the case that $\pi \in U$. Substituting $a = \pi * \boldsymbol{\sigma}(i)$, $b = \pi * \boldsymbol{\sigma}(j)$, $c = \pi * \boldsymbol{\sigma} * \boldsymbol{\tau}(i)$, $d = \pi * \boldsymbol{\sigma} * \boldsymbol{\tau}(j)$ in Lemma 17 gives us

$$\begin{aligned} &|\pi * \boldsymbol{\sigma}(i) - \pi * \boldsymbol{\sigma} * \boldsymbol{\tau}(i)| + |\pi * \boldsymbol{\sigma}(j) - \pi * \boldsymbol{\sigma} * \boldsymbol{\tau}(j)| \\ &\leq |\pi * \boldsymbol{\sigma}(i) - \pi * \boldsymbol{\sigma} * \boldsymbol{\tau}(j)| + |\pi * \boldsymbol{\sigma}(j) - \pi * \boldsymbol{\sigma} * \boldsymbol{\tau}(i)| \\ &= |\pi' * \boldsymbol{\sigma}(j) - \pi' * \boldsymbol{\sigma} * \boldsymbol{\tau}(j)| + |\pi' * \boldsymbol{\sigma}(i) - \pi' * \boldsymbol{\sigma} * \boldsymbol{\tau}(i)|, \end{aligned} \tag{19}$$

where the equality follows from the facts that (a) $\pi * \boldsymbol{\sigma}(i) = \pi' * \boldsymbol{\sigma}(j)$ and $\pi * \boldsymbol{\sigma}(j) = \pi' * \boldsymbol{\sigma}(i)$ since $\pi' * \boldsymbol{\sigma}$ is just $\pi * \boldsymbol{\sigma}$ with the adjacent elements i and j swapped, and (b) $\pi' * \boldsymbol{\sigma} * \boldsymbol{\tau} = \pi * \boldsymbol{\sigma} * \boldsymbol{\tau}$. Also, since $\pi' * \boldsymbol{\sigma}$ is just $\pi * \boldsymbol{\sigma}$ with the adjacent elements i and j swapped, $|\pi' * \boldsymbol{\sigma}(d) - \pi' * \boldsymbol{\sigma} * \boldsymbol{\tau}(d)| = |\pi * \boldsymbol{\sigma}(d) - \pi * \boldsymbol{\sigma} * \boldsymbol{\tau}(d)|$ for all $d \in D \setminus \{i, j\}$. If we sum over all d (where we make use of (19) for $d = i$ and $d = j$), we obtain $F(\pi * \boldsymbol{\sigma}, \pi * \boldsymbol{\sigma} * \boldsymbol{\tau}) \leq F(\pi' * \boldsymbol{\sigma}, \pi' * \boldsymbol{\sigma} * \boldsymbol{\tau})$. Since $\pi \in U$, we have that $F(\rho * \boldsymbol{\tau}^{\mathrm{R}} * \boldsymbol{\sigma}, \rho * \boldsymbol{\tau}^{\mathrm{R}} * \boldsymbol{\sigma} * \boldsymbol{\tau}) < F(\pi * \boldsymbol{\sigma}, \pi * \boldsymbol{\sigma} * \boldsymbol{\tau})$. Combining these last two inequalities, we obtain $F(\rho * \boldsymbol{\tau}^{\mathrm{R}} * \boldsymbol{\sigma}, \rho * \boldsymbol{\tau}^{\mathrm{R}} * \boldsymbol{\sigma} * \boldsymbol{\tau}) < F(\pi' * \boldsymbol{\sigma}, \pi' * \boldsymbol{\sigma} * \boldsymbol{\tau})$. Therefore, $\pi' \in U$, and so $\pi' \in S$, which was to be shown.

We now examine the case that $\pi \in V$. From above, we see that $\pi' * \boldsymbol{\sigma} * \boldsymbol{\tau} = \pi * \boldsymbol{\sigma} * \boldsymbol{\tau}$, while $\pi' * \boldsymbol{\sigma}$ and $\pi * \boldsymbol{\sigma}$ differ only by swapping the adjacent elements i and j. Since, as shown above, $\pi' * \boldsymbol{\sigma}(i) > \pi' * \boldsymbol{\sigma}(j)$ while $\pi' * \boldsymbol{\sigma} * \boldsymbol{\tau}(i) < \pi' * \boldsymbol{\sigma} * \boldsymbol{\tau}(j)$, we see that there is exactly one more inversion between $\pi' * \boldsymbol{\sigma}$ and $\pi' * \boldsymbol{\sigma} * \boldsymbol{\tau}$ than between $\pi * \boldsymbol{\sigma}$ and $\pi * \boldsymbol{\sigma} * \boldsymbol{\tau}$. Hence, $K(\pi * \boldsymbol{\sigma}, \pi * \boldsymbol{\sigma} * \boldsymbol{\tau}) < K(\pi' * \boldsymbol{\sigma}, \pi' * \boldsymbol{\sigma} * \boldsymbol{\tau})$. By our assumption, $\pi \in V$, and so $K(\rho * \boldsymbol{\tau}^{\mathrm{R}} * \boldsymbol{\sigma}, \rho * \boldsymbol{\tau}^{\mathrm{R}} * \boldsymbol{\sigma} * \boldsymbol{\tau}) < K(\pi * \boldsymbol{\sigma}, \pi * \boldsymbol{\sigma} * \boldsymbol{\tau})$. Combining these last two inequalities, we obtain $K(\rho * \boldsymbol{\tau}^{\mathrm{R}} * \boldsymbol{\sigma}, \rho * \boldsymbol{\tau}^{\mathrm{R}} * \boldsymbol{\sigma} * \boldsymbol{\tau}) < K(\pi' * \boldsymbol{\sigma}, \pi' * \boldsymbol{\sigma} * \boldsymbol{\tau})$. Therefore, $\pi' \in V$, and so $\pi' \in S$, which was to be shown. □

We can now prove Theorem 5. We prove the theorem for F_{Haus}. The proof for K_{Haus} is analogous. Recall that

$$F_{\mathrm{Haus}}(\boldsymbol{\sigma}, \boldsymbol{\tau}) = \max\left\{\max_{\sigma} \min_{\tau} F(\sigma, \tau), \max_{\tau} \min_{\sigma} F(\sigma, \tau)\right\},$$

where throughout this proof, σ and τ range through all full refinements of $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$, respectively. We show $\max_{\sigma} \min_{\tau} F(\sigma, \tau) = F(\rho * \boldsymbol{\tau}^{\mathrm{R}} * \boldsymbol{\sigma}, \rho * \boldsymbol{\sigma} * \boldsymbol{\tau})$. A similar argument shows that $\max_{\tau} \min_{\sigma} F(\sigma, \tau) = F(\rho * \boldsymbol{\tau} * \boldsymbol{\sigma}, \rho * \boldsymbol{\sigma}^{\mathrm{R}} * \boldsymbol{\tau})$. The claim about F_{Haus} in the statement of the theorem follows easily.

Think for now of $\sigma \succeq \boldsymbol{\sigma}$ as fixed. Then by Lemma 19, the quantity $F(\sigma, \tau)$, where τ ranges over all full refinements of $\boldsymbol{\tau}$, is minimized when $\tau = \sigma * \boldsymbol{\tau}$. That is, $\min_{\tau} F(\sigma, \tau) = F(\sigma, \sigma * \boldsymbol{\tau})$.

By Lemma 20, the quantity $F(\sigma, \sigma * \boldsymbol{\tau})$, where σ ranges over all full refinements of $\boldsymbol{\sigma}$, is maximized when $\sigma = \rho * \boldsymbol{\tau}^{\mathrm{R}} * \boldsymbol{\sigma}$. Hence, $\max_{\sigma} \min_{\tau} F(\sigma, \tau) = F(\rho * \boldsymbol{\tau}^{\mathrm{R}} * \boldsymbol{\sigma}, \rho * \boldsymbol{\tau}^{\mathrm{R}} * \boldsymbol{\sigma} * \boldsymbol{\tau})$. Since $\rho * \boldsymbol{\tau}^{\mathrm{R}} * \boldsymbol{\sigma} * \boldsymbol{\tau} = \rho * \boldsymbol{\sigma} * \boldsymbol{\tau}$, we have $\max_{\sigma} \min_{\tau} F(\sigma, \tau) = F(\rho * \boldsymbol{\tau}^{\mathrm{R}} * \boldsymbol{\sigma}, \rho * \boldsymbol{\sigma} * \boldsymbol{\tau})$, as we wanted.

REFERENCES

[1] F. Aleskerov and B. Monjardet, *Utility Maximization, Choice, and Preference*, Stud. Econom. Theory 16, Springer-Verlag, Berlin, 2002.

[2] K. J. Arrow, *Social Choice and Individual Values*, John Wiley and Sons, New York, 1951.

[3] K. J. Arrow and M. D. Intriligator, *Handbook of Mathematical Economics*, North-Holland, Amsterdam, 1982.

[4] J. A. Aslam and M. Montague, *Models for metasearch*, in Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, New Orleans, 2001, pp. 276–284.

[5] K. A. Baggerly, *Visual Estimation of Structure in Ranked Data*, Ph.D. thesis, Rice University, Houston, TX, 1995.

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

[7] D. S. Bridges and G. B. Mehta, *Representations of Preference Orderings*, Lecture Notes in Econom. and Math. Systems 422, Springer-Verlag, Heidelberg, Berlin, New York, 1995.

[8] W. W. Cohen, R. E. Schapire, and Y. Singer, *Learning to order things*, J. Artificial Intelligence Res., 10 (1999), pp. 243–270.

[9] D. E. Critchlow, *Metric Methods for Analyzing Partially Ranked Data*, Lecture Notes in Statist. 34, Springer-Verlag, Berlin, 1980.

[10] P. Diaconis, *Group Representation in Probability and Statistics*, IMS Lecture Series Monogr. Ser. 11, Institute of Mathematical Statistics, Hayward, CA, 1988.

[11] P. Diaconis and R. Graham, *Spearman's footrule as a measure of disarray*, J. Roy. Statist. Soc. Ser. B, 39 (1977), pp. 262–268.

[12] C. Dwork, R. Kumar, M. Naor, and D. Sivakumar, *Rank aggregation methods for the web*, in Proceedings of the 10th International World Wide Web Conference, Hong Kong, 2001, pp. 613–622.
[13] J. Eatwell, M. Milgate, and P. Newman, *The New Palgrave: A Dictionary of Economics*, MacMillan, London, 1987.
[14] R. Fagin, R. Kumar, M. Mahdian, D. Sivakumar, and E. Vee, *Comparing and aggregating rankings with ties*, in Proceedings of the 23rd ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, Denver, 2004, pp. 47–58.
[15] R. Fagin, R. Kumar, K. McCurley, J. Novak, D. Sivakumar, J. Tomlin, and D. Williamson, *Searching the workplace web*, in Proceedings of the 12th International World Wide Web Conference, Budapest, 2003, pp. 366–375.
[16] R. Fagin, R. Kumar, and D. Sivakumar, *Comparing top k lists*, SIAM J. Discrete Math., 17 (2003), pp. 134–160.
[17] R. Fagin, R. Kumar, and D. Sivakumar, *Efficient similarity search and classification via rank aggregation*, in Proceedings of the 2003 ACM SIGMOD International Conference on Management of Data, San Diego, 2003, pp. 301–312.
[18] P. Fishburn, *Condorcet social choice functions*, SIAM J. Appl. Math., 33 (1977), pp. 469–489.
[19] P. C. Fishburn, *Interval Orders and Interval Graphs: A Study of Partially Ordered Sets*, John Wiley and Sons, New York, 1985.
[20] L. A. Goodman and W. H. Kruskal, *Measures of association for cross classification*, J. Amer. Statist. Assoc., 49 (1954), pp. 732–764.
[21] T. H. Haveliwala, A. Gionis, D. Klein, and P. Indyk, *Evaluating strategies for similarity search on the web*, in Proceedings of the 11th International World Wide Web Conference, Honolulu, 2002, pp. 432–442.
[22] M. Kendall and J. D. Gibbons, *Rank Correlation Methods*, Edward Arnold, London, 1990.
[23] M. G. Kendall, *The treatment of ties in ranking problems*, Biometrika, 33 (1945), pp. 239–251.
[24] G. Lebanon and J. D. Lafferty, *Cranking: Combining rankings using conditional probability models on permutations*, in Proceedings of the 19th International Conference on Machine Learning, Sydney, Australia, 2002, pp. 363–370.
[25] M. Montague and J. A. Aslam, *Condorcet fusion for improved retrieval*, in Proceedings of the 11th International Conference on Information and Knowledge Management, McLean, VA, 2002, pp. 538–548.
[26] M. E. Renda and U. Straccia, *Web metasearch: Rank vs. score based rank aggregation methods*, in Proceedings of the 18th Annual Symposium on Applied Computing, Melbourne, FL, 2003, pp. 841–846.
[27] F. S. Roberts, *Measurement Theory, with Applications to Decisionmaking, Utility, and the Social Sciences*, Encyclopedia Math. Appl., Addison-Wesley, Reading, MA, 1979.
[28] J. Sese and S. Morishita, *Rank aggregation method for biological databases*, Genome Informatics, 12 (2001), pp. 506–507.
[29] R. R. Yager and V. Kreinovich, *On how to merge sorted lists coming from different web search tools*, Soft Computing Research Journal, 3 (1999), pp. 83–88.