Generalized Distances between Rankings

Ravi Kumar Sergei Vassilvitskii
Yahoo! Research
701 First Avenue
Sunnyvale, CA 94089.
{ravikumar,sergei}@yahoo-inc.com

ABSTRACT

Spearman's footrule and Kendall's tau are two well established distances between rankings. They, however, fail to take into account concepts crucial to evaluating a result set in information retrieval: element relevance and positional information. That is, changing the rank of a highly-relevant document should result in a higher penalty than changing the rank of an irrelevant document; a similar logic holds for the top versus the bottom of the result ordering. In this work, we extend both of these metrics to those with position and element weights, and show that a variant of the Diaconis–Graham inequality still holds — the generalized two measures remain within a constant factor of each other for all permutations.

We continue by extending the element weights into a distance metric between elements. For example, in search evaluation, swapping the order of two nearly duplicate results should result in little penalty, even if these two are highly relevant and appear at the top of the list. We extend the distance measures to this more general case and show that they remain within a constant factor of each other.

We conclude by conducting simple experiments on web search data with the proposed measures. Our experiments show that the weighted generalizations are more robust and consistent with each other than their unweighted counterparts.

Categories and Subject Descriptors. H.3.m [**Information Storage and Retrieval**]: Miscellaneous

General Terms. Algorithms, Experimentation, Theory

Keywords. Permutation distances, Kendall tau, Spearman footrule, metrics, retrieval measures

1. INTRODUCTION

The study of metrics for information retrieval is as old as the field itself, after all, it is impossible to evaluate a result without some sort of a measure. There are dozens of different metrics used in the literature, from classical ones like Kendall's tau and Spearman's footrule, neoclassical ones like MAP and NDCG, and much newer propositions, for example, rank distance [4], ERR [5], and others [1, 7, 21]. The preponderance of measures leads to a natural second order question: how does one measure the measures?

WWW 2010, April 26–30, 2010, Raleigh, North Carolina, USA.
ACM 978-1-60558-799-8/10/04.

Evaluating metrics is indeed a difficult problem. There have been two approaches generally taken by the community. The first is axiomatic. Here the approach is to establish basic properties that the metric should obey, and then derive a measure that satisfies these properties. Kendall's tau and Spearman's footrule serve as leading examples of this approach; a general version of DCG can also be seen in this light, if the axioms talk about a decreasing importance with rank and relevance.

A different approach is to exhibit a problem with one of the standard and accepted metrics, and propose a variation that fixes the problem. The new metric is then shown to generally correlate well with the old metric, except for the problematic examples, where it performs significantly better. Overtime, variations on these variations are introduced, and we are left with dozens of metrics to chose from and report in experimental evaluations.

In this paper we take an axiomatic approach. We argue that there are several criteria that any modern evaluation metric should satisfy. We describe these in turn.

(1) *Richness.* By now it has become evident that there are three important factors to keep in mind when computing the distance between two orderings. The metric should support element weights, which represent the relevance of a particular document or result to the query. Intuitively, an error on a high-weight element should be more significant than an error on a low-weight element. In a similar vein, the position of the element in the list plays a large role with respect to the efficacy of the metric: errors at the top of the list are costlier than errors at the tail of the list, and any new metric should be reflective of this fact. Finally, diversity, and, more generally, interaction between results has been recognized as an important part of result evaluation. For example, returning five identical results may not lead to high user satisfaction even if individually each one of those results is extremely relevant.

(2) *Simplicity.* One of the reasons metrics like Kendall's tau remain prominent, even when their shortfalls have been widely recognized, is their inherent simplicity. It is natural to count the total number of inversions, or look at the total ℓ_1 distance, as in the case of Spearman's Footrule. However, as the metrics get richer, this simplicity is often lost. Even the authors themselves lament that their new "measure is not intuitive" [4].

(3) *Generalization.* Notwithstanding the richness criteria, any proposed metric should collapse to a natural metric in cases where the richer criteria do not play a role. For example if all of the element weights are set to 1, or all pairwise

interactions are the same, the metric should simplify to one of the classic and well-known metrics. In this sense, the new metric must generalize already existing metrics.

(4) *Basic properties.* The proposed evaluation measure should satisfy some basic properties that make it easier to reason about. For example, it should be scale-free: that is scaling all of the weights by a the same constant factor should not change the solution. It should be invariant under relabeling of the elements, and, ideally, should follow the triangle inequality: for two permutations σ and τ, the metric M should satisfy $M(\sigma, \tau) \leq M(\sigma) + M(\tau)$.

(5) *Correlation with other metrics.* Finally, a suite of metrics all capturing the same effect in different ways is much more powerful than a single new metric. Typically, all of these measures have good intuition underlying them, however the formal statement for the intuition takes on different forms. Therefore if two metrics are trying to capture the same effect, but disagree greatly on some of the examples, it typically implies that *neither* is fully capturing the effect in question. One of the best examples of this kind of correlation is the fact that while Kendall's tau and Spearman's footrule take very different ways of measuring distances between permutations, Diaconis and Graham [9] showed that the two measures are equivalent up to a factor of 2. The correlation also gives freedom when trying to build on top of these metrics. As we illustrate in Section 5 sometimes solving the problem with respect to one optimization metric is NP-hard, while using an equivalent metric leads to a polynomial time solution.

1.1 Our contributions

In this work we enrich the space measured by Spearman's footrule and Kendall's tau to satisfy the first condition above. We give new formulations of these metrics that capture element weights, position weights, and pairwise distances between permutations, while at the same time retaining their classical form. The generalized footrule distance can be seen as an ℓ_1 distance on the right metric space, and the generalized Kendall's tau has a term for every inverted pair. We show that even though they take on very different forms, the generalized versions of these two metrics remain within a factor of three of each other, and collapse to their classical variants when all of the weights and pairwise distances are set to 1.

2. RELATED WORK

The subject of distances between permutations (or ranked lists) has a rich and long history. Perhaps the two most popular measure are Spearman's footrule distance [20], which measures the ℓ_1 distance between ranks and Kendall's tau [15], which measures the total number of pairwise inversions. In a celebrated result, Diaconis and Graham [9] showed that these two measures are always within a factor of two from each other. See the book by Diaconis [8] for discussions on other metrics on permutations.

In the context of web information retrieval many new measures have been developed over the years. Average precision (AP) and reciprocal rank (RR) are two popular metrics that can be seen as distances on permutations of binary strings (where each document is marked to be either relevant or not relevant). To quantify the importance of the top few positions over the bottom ranks, the set of measures was expanded to compare only the top-k rankings, for a given value of the cut-off k [12]. More generally, a metric like NDCG [14] uses a scoring function that decays with the rank of the document. Recently Yilmaz et al. [21] and Carterette [4] argued that the lack of this feature makes Kendall's tau is a poor metric because, because it penalizes equally inversions near the head and near the tail of a list. In a new take on a difficult related problem, D'Alberto and Dasdan [7] address the sparsity issue often prevalent in web result evaluation and describe distance measures on incomplete permutations; see also [3, 11].

In addition to decreasing with the rank, NDCG also combines the relevance (or weight) of the document into the final score. This variant has been considered in the statistics literature as well. For example Shieh et al. [18, 19] give a version of Kendall's tau distance, where every inversion is weighted in proportion to the product of the two element weights. However, this work does not focus on its "metric"-like properties. In fact, [18] work with a generalization of Spearman rho (a L_2 version of footrule) that takes position weights into account. Once again, there is no formal analysis of its properties (or its relationship to a Kendall variant). Sculley [17] proposed extending Kendall's tau to incorporate item-item similarity using a notion of "similarity projections"; his extension, however, does not appear amenable to our analysis.

A growing area of research has recently focused on incorporating diversity into search metrics. The cascade model [6], and the follow up work by Agrawal et al. [1] and Chapelle et al. [5] show that earlier results have a profound effect on the perceived relevance of results further in the list and present several new metrics to take that into account.

A different strand of research uses distances between permutations to solve the rank aggregation problem: given a number of different orderings of results, find the most agreeable ordering. This problem has been extensively studied by Dwork et al. [10] and Fagin et al. [13], and has a close relation to the Feedback Arc Set (FAS) problem on tournament graphs [2, 16].

3. BASIC DEFINITIONS

Let $[n] = \{1, \ldots, n\}$ be a universe of elements. Let S_n be the set of permutations on $[n]$ and for $\sigma \in S_n$, let $\sigma(i)$ denote the rank of the element i.

There are two well-known metrics that evaluate the distance between two permutations $\sigma, \tau \in S_n$: the Spearman's footrule distance $F(\sigma, \tau)$ and the Kendall's tau $K(\sigma, \tau)$. Before we state the definitions, we note that these metrics do not depend on the actual identity of elements (such metrics are called *invariant*). Hence, it suffices to consider $F(\sigma) = F(\sigma, 1)$ and $K(\sigma) = K(\sigma, 1)$, where 1 is the identity permutation. Now we proceed with the formal definitions.

(1) The *Spearman's footrule* distance is given by

$$F(\sigma) = \sum_i |i - \sigma(i)|; \qquad (1)$$

this measures the total element-wise displacement from the identity permutation.

(2) The *Kendall's tau* is given by

$$K(\sigma) = \sum_{(i,j): i > j} [\sigma(i) < \sigma(j)]\,; \qquad (2)$$

this measures the total number of pairwise *inversions*.

In a celebrated result, Diaconis and Graham [9] showed[1] that these metrics differ by at most a constant factor (such pair of metrics are said to be *equivalent*).

Theorem 1 (Diaconis–Graham (DG) inequality). *For all $\sigma \in S_n$, $K(\sigma) \le F(\sigma) \le 2K(\sigma)$.*

This inequality is in fact tight.

All the metrics we will define in the paper are invariant metrics. Note that if M is an invariant metric on S_n, then M is symmetric provided $\forall \sigma, M(\sigma) = M(\sigma^{-1})$. Indeed, $M(\sigma,\tau) = M(\sigma\tau^{-1}) = M((\sigma\tau^{-1})^{-1}) = M(\tau\sigma^{-1}) = M(\tau,\sigma)$, where the first and last equalities follow from invariance and the second equality follows from the premise. Also, to make our exposition simpler, we will not normalize our metrics to be in $[0,1]$ or $[-1,1]$.

3.1 Element weights

For $i \in [n]$, let $w_i > 0$ be the *weight of an element*; let $w = w_1, \ldots, w_n$. For the remainder of the paper, we assume that the weights are integral, i.e., $w_i \in \mathbf{Z}^+$; all of the results, however, follow for non-integral weights as well.

There are many ways to define a weighted analog of Kendall's tau (K_w) and Spearman's footrule (F_w). For example, to define K_w, one can stipulate that an inversion of elements i and j should have a penalty proportional to some average of their weights, say $\frac{w_i+w_j}{2}$, or to their product, $w_i w_j$. Alternatively, one may define F_w where each displacement is scaled by the weight of the element i that is displaced, say, $w_i|i - \sigma(i)|$. We strive to find a variant of Kendall's tau and Spearman's footrule so that the measures continue to be equivalent, up to a small constant factor.

We define the weighted version K_w of Kendall's tau by penalizing each inversion proportionally to the product of the weights of the two elements being inverted:

$$K_w(\sigma) = \sum_{i>j} w_i w_j \left[\sigma(i) < \sigma(j)\right]. \tag{3}$$

The weighted footrule F_w is defined to be:

$$F_w(\sigma) = \sum_i w_i \left| \sum_{j:j\le i} w_j - \sum_{j:\sigma(j)\le\sigma(i)} w_j \right|. \tag{4}$$

The inner term measures the sum of weights of elements that span the displacement of element i.

Note that if all of the weights are uniformly unit, (3) and (4) collapse to (2) and (1) respectively.

Example. Consider the permutation $\sigma([abc]) = [bca]$, and let $w_a = 1, w_b = 2$, and $w_c = 3$. Then $K_w(\sigma) = 1{\cdot}2+1{\cdot}3 = 5$ and $F_w(\sigma) = 1\cdot(2+3)+2\cdot 1+3\cdot 1 = 10$. Notice that in this case $K_w(\sigma)$ and $F_w(\sigma)$ are a factor of 2 apart. As we will show later, these two metrics are always within a factor of two of each other.

3.2 Position weights

As we discussed earlier, in addition to element weights, we wish to define a distance that penalizes inversions early in the permutation more than inversions late in the permutation. We therefore introduce *position weights* to differentiate between inversions occurring near the head or the tail of a permutation.

In order to study the effect of position weights we first model the *cost of a swap* between two adjacent positions. Let δ_i be the cost of swapping an element at position $i-1$ with an element at position i; let $\delta = \delta_1, \ldots, \delta_n$. In the traditional version of Kendall and footrule metrics, $\delta_i = 1$ for all i, and thus all swaps have the same cost. As with element weights, we assume that $\delta_i \ge 0$ for all i. Let $p_1 = 1$ and for $1 < i \le n$, $p_i = \sum_{j=1}^{i-1}\delta_j$. For notational purposes, let

$$\bar{p}_i(\sigma) = \frac{p_i - p_{\sigma(i)}}{i - \sigma(i)},$$

be the average cost that i encountered in moving from position i to position $\sigma(i)$, where $\bar{p}_i = 1$ if $i = \sigma(i)$. By the monotonicity of the p_i's, we have $\bar{p}_i(\sigma) > 0$ for all i.

We are now ready to define the position weighted versions Spearman's footrule (F_δ) and Kendall's tau (K_δ):

$$K_\delta(\sigma) = \sum_{i<j} \bar{p}_i(\sigma)\bar{p}_j(\sigma)\left[\sigma(i) > \sigma(j)\right]. \tag{5}$$

$$F_\delta(\sigma) = \sum_i \bar{p}_i(\sigma) \left| \sum_{j:j\le i} \bar{p}_j(\sigma) - \sum_{j:\sigma(j)\le\sigma(i)} \bar{p}_j(\sigma) \right|. \tag{6}$$

Note that if $\forall i, \delta_i = 1$, then $\forall i, \bar{p}_i(\sigma) = 1$. Hence for unit swap costs, (5) and (6) collapse to (2) and (1) respectively.

Example.. Let $\sigma([abc]) = [bca]$ as before, and let $\delta_1 = 1, \delta_2 = 0.5$. Then $\bar{p}_a = 0.75, \bar{p}_b = 1$, and $\bar{p}_c = 0.5$. We get $K_\delta(\sigma) = 1.125$ and $F_\delta(\sigma) = 2.25$.

3.3 Element similarities

While position weights address the question of swaps occurring near the beginning or an end of a permutation, many times the importance of the swap crucially depends on the *similarity* of the elements being swapped. In an extreme case, swapping two identical elements should result in no change to the metric, whereas swapping two radically different elements should result in a large effect, even if the weights are small and the elements are in adjacent positions. To formally describe this, we define a distance metric on elements. Let $D : [n] \times [n]$ be a non-empty metric on S_n and let D_{ij} be the cost of a swap of elements i and j; clearly, D is a metric and for three elements i, j, k, we have

$$D_{ik} \le D_{ij} + D_{jk}. \tag{7}$$

The generalization of Kendall's tau (K_D) and Spearman's footrule (S_D) to incorporate element distances is most evident in the former, where we scale each inversion by the distance between the pair of elements inverted.

$$K_D(\sigma) = \sum_{i<j} D_{ij}\left[\sigma(i) > \sigma(j)\right]. \tag{8}$$

And, likewise, we define

$$F'_D(\sigma) = \sum_i \left| \sum_{j:j\le i} D_{ij} - \sum_{j:\sigma(j)\le\sigma(i)} D_{ij} \right|. \tag{9}$$

[1] Actually, they show something stronger, but we will not be concerned about the stronger version.

Observe that when D is the unit metric, (8) and (9) collapse to (2) and (1) respectively. Notice that the definition of F'_D above is asymmetric and we symmetrize it as

$$F_D(\sigma) = \frac{1}{2}(F'_D(\sigma) + F'_D(\sigma^{-1})). \quad (10)$$

3.4 The general definition

Finally, we can combine element weights, position weights, and element similarities into one single definition.

Definition 2 (Generalized Kendall's tau). *Given w, δ, D, for $\sigma \in S_n$, let*

$$K^* = K_{w,\delta,D}(\sigma) = \sum_{i<j} w_i w_j \bar{p}_i \bar{p}_j D_{ij} \,[\sigma(i) > \sigma(j)]\,. \quad (11)$$

Definition 3 (Generalized Spearman's footrule). *Given w, δ, D, for $\sigma \in S_n$, let*

$$F'^* = F'_{w,\delta,D}(\sigma) = \sum_i w_i \bar{p}_i(\sigma) \cdot \left| \sum_{j:j\le i} w_j \bar{p}_j(\sigma) D_{ij} - \sum_{j:\sigma(j)\le\sigma(i)} w_j \bar{p}_j(\sigma) D_{ij} \right|. \quad (12)$$

The above footrule version can be symmetrized as before and from now on, we will only work with the symmetric version F^*.

4. PROPERTIES

4.1 Basic characteristics

We first show a reduction from the element-weighted case to the unweighted case.

Lemma 4. *Given w, for any σ, there is a τ such that $K_w(\sigma) = K(\tau)$ and $F_w(\sigma) = F(\tau)$.*

Proof. Recall that all of the weights are assumed to be integral. Therefore we can divide an element i of weight $w_i > 0$ into w_i sub-elements $i'_1, \ldots, i'_{w_i}$, each of weight 1. Consider a permutation τ that keeps all of the sub-elements in the same order, but reorders these blocks based on σ. Formally, $\tau(i'_k) = k + \sum_{j:\sigma(j)<\sigma(i)}$. (For example, if $\sigma([abc]) = ([bca])$ and $w_a = 1, w_b = 2, w_c = 3$, then the corresponding τ is: $\tau([a'_1 b'_1 b'_2 c'_1 c'_2 c'_3]) = [b'_1 b'_2 c'_1 c'_2 c'_3 a'_1]$).) In what follows, denote by $pos(i) = \sum_{j:j<i} w_j$.

First, consider the set of inversions produced by σ. Each inversion $\langle i, j\rangle$ in σ results in $w_i \cdot w_j$ inversions in τ, since every sub-element of i is inverted with every sub-element of j. This implies that $K_w(\sigma) = K(\tau)$. In addition,

$$\begin{aligned} F(\tau) &= \sum_i \sum_{j=1}^{w_i} |pos(i) + k - (k + \sum_{j:\sigma(j)<\sigma(i)} w_j)| \\ &= \sum_i \sum_{j=1}^{w_i} |pos(i) - \sum_{j:\sigma(j)<\sigma(i)} w_j| \\ &= \sum_i w_i |pos(i) - \sum_{j:\sigma(j)<\sigma(i)} w_j| \\ &= F_w(\sigma). \quad \square \end{aligned}$$

Next we show how to transform the position weight case into an element weight case.

Lemma 5. *Given δ, for any σ, there is a w such that $K_\delta(\sigma) = K_w(\sigma)$ and $F_\delta(\sigma) = F_w(\sigma)$.*

Proof. Fix σ and let $w_i = \bar{p}_i(\sigma)$. Then, from (5) and (3), we have $K_\delta(\sigma) = K_w(\sigma)$. Likewise, from (6) and (4), we have $F_\delta(\sigma) = F_w(\sigma)$. It is important to note that the element weights here are not oblivious — they actually depend on the permutation σ that is being considered. $\square$

Finally we show that the metrics are scale invariant.

Lemma 6. *Given w, δ, and a metric D, let $w' = c_1 \cdot w$, $\delta' = c_2 \cdot \delta$, and $D' = c_3 \cdot D$ for some constants $c_1, c_2, c_3 > 0$. Then for any σ, $K_{w',\delta',D'}(\sigma) = c \cdot K_{w,\delta,D}(\sigma)$ and $F_{w',\delta',D'}(\sigma) = F_{w,\delta,D}(\sigma)$, where $c = c_1^2 c_2^2 c_3$.*

4.2 Metric properties

It is easy to see that both K^* and F^* do not depend on the actual identity of the elements, as long as relabeling an element affects neither its weight nor its similarity to other elements. Hence, these are invariant distances. Since we assume $w > 0, \delta > 0$, and D to a non-empty metric, $K^*(\sigma) = 0 = F^*(\sigma)$ if and only if $\sigma = 1$. Furthermore, by our definition, we have $K^*(\sigma) = K^*(\sigma^{-1})$ and $F^*(\sigma) = F^*(\sigma^{-1})$. This guarantees symmetry.

For triangle inequality, notice that Lemma 4 and Lemma 5 show that $K_w, F_w, K_\delta, F_\delta$ satisfy the triangle inequality. It only remains to show the triangle inequality for the element similarity case, i.e., K_D and F_D also satisfy the triangle inequality.

First, we consider the element similarity version of Kendall's tau. Interestingly, the proof does not use the metric property of D.

Lemma 7. *Given D, for all σ, τ, $K_D(\sigma, \tau) \le K_D(\sigma) + K_D(\tau)$.*

Proof. We will use (8) and show that the triangle inequality holds point-wise. Indeed, consider a pair $\langle i, j\rangle$ of inversions between σ and τ. Without loss of generality, let $\sigma(i) > \sigma(j)$ and $\tau(j) < \tau(i)$. If $i < j$, then, $\langle i, j\rangle$ is inverted in σ and if $i > j$, then $\langle i, j\rangle$ is inverted in τ. $\square$

We next prove that the Spearman footrule with element similarities satisfies the triangle inequality. Interestingly, the triangle inequality is not satisfied point-wise, i.e., it is easy to construct examples such that for each fixed i, the term in (12) fails to satisfy the triangle inequality. Unlike the Kendall tau case, this proof crucially uses the fact that D is a metric (otherwise, there are simple counter-examples).

Lemma 8. *Given a metric D, for all σ, τ, $F_D(\sigma, \tau) \le F_D(\sigma) + F_D(\tau)$.*

Proof. Recall that our goal is to prove that for two permutations σ, τ, $F_D(\sigma, \tau) \le F_D(\sigma) + F_D(\tau)$. We will do that by examining the occurrences of a specific D_{ij} on both the LHS and RHS of the desired inequality.

First, we define an unordered interval $I(a, b)$ to be $[a, b]$ if $a < b$ and $[b, a]$ otherwise. We define the *box* of an element i with respect to $\sigma, \tau \in S_n$ to be the multi-set

$$\begin{aligned} \texttt{box}_{\sigma,\tau}(i) = &\{j : \sigma(j) \in I(\sigma(i), \tau(i))\} \,\cup \\ &\{j : \tau(j) \in I(\sigma(i), \tau(i))\}, \end{aligned}$$

where the union is a multi-set union (the multiplicity of each element is at most 2). If $\tau = 1$, we simply abbreviate it as $\texttt{box}_\sigma(i)$.

Informally, $\mathtt{box}_\sigma(i)$ consists of all the elements that do not cancel out in the inner summation in (9), for a fixed i. In fact, each occurrence of D_{ij} in the expression for $F_D(\sigma)$ is the sum of the multiplicity of j in $\mathtt{box}_\sigma(i)$ and the multiplicity of i in $\mathtt{box}_\sigma(j)$. Moreover, it is easy to verify that if $\langle i, j\rangle$ is inverted between σ and τ, then $\mathtt{box}_\sigma(i) \cap \mathtt{box}_\sigma(j) \neq \emptyset$. Note that the converse of this statement is not necessarily true.

Now, consider a specific D_{ij}. We first claim that in any expression for $F_D(\sigma, \tau)$, the term D_{ij} can occur 0, 2, 3, or 4 times. Clearly if $\mathtt{box}_{\sigma,\tau}(i) \cap \mathtt{box}_{\sigma,\tau}(j) = \emptyset$, then the term D_{ij} never occurs. If $\sigma(i) = \tau(j)$ and $\tau(i) = \sigma(j)$, then it can be seen that D_{ij} occurs four times. If only one of these conditions is satisfied, then D_{ij} occurs thrice and if neither is satisfied, but the boxes of i and j intersect, then D_{ij} occurs twice. Our goal now is to show that each occurrence of D_{ij} on the LHS is matched by an occurrence of value D_{ij} or more on the RHS.

Suppose D_{ij} occurs twice on the LHS. If $\mathtt{box}_\sigma(i)\cap\mathtt{box}_\sigma(j) \neq \emptyset$, then we will have at least two occurrences of D_{ij} on the RHS. If $\mathtt{box}_\sigma(i) \cap \mathtt{box}_\sigma(j) = \emptyset$, then it follows that $\langle i, j\rangle$ is not inverted in σ. In this case, one can check that $\mathtt{box}_\tau(i) \cap \mathtt{box}_\tau(j) \neq \emptyset$, yielding at least two occurrences of D_{ij} on the RHS. Note that in the latter case, $\langle i, j\rangle$ need not be inverted in τ, but the boxes will intersect.

Suppose D_{ij} occurs thrice on the LHS. While the analysis for the two occurrence case can still be applied here to match two occurrences of D_{ij}, we still need to account for the third occurrence of D_{ij}. Now observe that there is some k such that $\sigma(i) = k = \tau(j)$ (the case when $\sigma(j) = k = \tau(i)$ is similar). It follows $i \in \mathtt{box}_\sigma(k)$ and $j \in \mathtt{box}_\tau(k)$. Therefore, both the terms D_{ik} and D_{kj} will occur on the RHS. (Note that these terms cannot "claimed" elsewhere since k is uniquely defined by both i and j.) By the triangle inequality (7), we can account for the third occurrence of D_{ij} on the LHS by the term $D_{ik} + D_{kj} \geq D_{ij}$ on the RHS.

The case when D_{ij} occurs four times on the LHS is similar. We will have $\sigma(i) = k_1 = \tau(j)$ and $\sigma(j) = k_2 = \tau(i)$ and the triangle inequality (7) has to be applied twice. □

4.3 Equivalence

In this section our goal is establish the equivalence relationship between F^* and K^*. We proceed to do this by a series of reductions. We first show that the Diaconis–Graham inequality holds "as is" between K_w and F_w and between K_δ and F_δ.

In the case of element weights, we show that K_w and F_w are equivalent up to a factor of 2.

Theorem 9. *For all σ and w, $K_w(\sigma) \leq F_w(\sigma) \leq 2K_w(\sigma)$.*

Proof. This follows from the reduction in Lemma 4 and then applying Theorem 1. □

Next, we show that K_δ and F_δ are equivalent up to a factor of 2.

Theorem 10. *For all σ and δ, $K_\delta(\sigma) \leq F_\delta(\sigma) \leq 2K_\delta(\sigma)$.*

Proof. This follows from Lemma 5 and then applying Theorem 9. □

Moreover, the factor of 2 is the best possible. Consider a three element unweighted example, $\sigma([abc]) = [bca]$. Then the footrule distance is 4, while Kendall's tau is 2.

In the case of element similarities, recall that for any two elements i and j the cost of inverting these two elements is D_{ij}. We show that the K_D and F_D differ by at most a factor of 3 from each other.

We begin with a simple classification of the inversions in K_D.

(1) Type I: an inversion $\langle i, j\rangle$ is a *type I* inversion if $\sigma(j) \in [i, \sigma(i)]$ or $\sigma(i) \in [j, \sigma(j)]$. Note that in this case the term D_{ij} appears *both* in the calculation of Kendall's tau and in Spearman's footrule.

(2) Type II: an inversion $\langle i, j\rangle$ is a *type II* inversion if $\sigma(j) < i < j < \sigma(i)$. In this case the term D_{ij} appears in the expression for K_D but not in F_D.

We now develop a notion of a valid mapping that will be helpful. Let B_i be the set of elements that form inversions of type II with i: $B_i = \{j : \sigma(j) < i < j < \sigma(i)\}$. These are the *bad* elements for i. On the other hand, for a given i, consider the elements $G_i = \{k : \sigma(k) \geq i\}$. These are the *good* elements, and we will map each bad element $j \in B_i$ to one good element $k \in G_i$.

Definition 11. *A mapping $\mu : \cup_i B_i \to \cup_i G_i$ is* valid *if the following four properties hold:*

(a) $\forall i, j, j' \in B_i \quad \mu_i(j) \neq \mu_i(j')$,

(b) $\forall i, i', j \in B_i \cap B_{i'} \quad \mu_i(j) \neq \mu_{i'}(j)$,

(c) $\forall i, j \in B_i \quad \sigma(\mu_i(j)) \in [i, \sigma(i))$, *and*

(d) $\forall i, j \in B_i \quad \sigma(\mu_i(j)) \in [j, \sigma(j))$.

We emphasize that the mapping μ is different for each permutation σ.

We now proceed to define a mapping μ and prove that it is valid. Let B_i and G_i as before, and let $\beta_i(j) = |\{j' \in B_i, j' < j\}|$, i.e., the number of elements j' in B_i that precede j. Similarly, let $\gamma_i(k) = \{k' \in G_i, \sigma(k') < \sigma(k)\}$, i.e., the number of elements in G_i that precede k. We define the mapping μ as:

$$\mu_i(j) = \{k : \beta_i(j) = \gamma_i(k)\}.$$

Lemma 12. *The mapping μ is valid.*

Proof. To show property (a), fix i and consider $j, j' \in B_i$. Since the elements j and j' have different β_i, $\mu_i(j) \neq \mu_i(j')$. For properties (c) and (d), observe that for all $k \in G_i$, $k > i > \sigma(j)$. On the other hand, by definition of β_i and γ_i, $\sigma_i(j) \leq j$. Therefore $\sigma_i(j) \leq j < \sigma(i)$.

It remains to prove property (b). Consider i, i' with $i > i'$ and an element j so that $j \in B_i$ and $j \in B_{i'}$. We claim that $\mu_i(j) \neq \mu_{i'}(j)$. Suppose by contradiction that the two were equal. Then, there is some element $k = \mu_i(j) = \mu_{i'}(j)$, and therefore $\beta_i(j) = \gamma_i(k)$ and $\beta_{i'}(j) = \gamma_{i'}(k)$. Equivalently, $\beta_{i'}(j) - \beta_i(j) = \gamma_{i'}(k) - \gamma_i(k)$. From the definition of γ_i, we can conclude that $\gamma_{i'}(k) - \gamma_i(k) = i - i'$.

On the other hand, we will show that $\beta_{i'}(j) - \beta_i(j) < i - i'$, and hence property (b) holds. In what follows we say that an element j' *contributes* to $\beta_i(j)$ if $j' \in B_i$ and $j' < j$.

Consider the elements that contribute to $\beta_{i'}(j)$. Since all of these elements are in $B_{i'}$ they must lie in the range of (i', j). We split this range into three groups: (i', i), the element i, and (i, j). To maximize $\beta_{i'}(j)$, we can assume that *all* of the elements from the first group contribute to it. The total number of such elements is $i - i' - 1$. Now consider the element i that forms the second group. Since $j \in B_i \cap B_{i'}$, we know that $i', i < j < \sigma(i'), \sigma(i)$. Since $i' < i$

we have $i' < i < \sigma(i')$, and therefore $i \notin B_{i'}$. Thus i does not contribute to $\beta_{i'}(j)$.

Finally, we show that every element in the third group that contributes to $\beta_{i'}(j)$ also contributes to $\beta_i(j)$. Therefore the contribution is canceled out when looking at the difference of $\beta_{i'}(j) - \beta_i(j)$. Formally we must prove that if $j' \in B_{i'} \cap (i, j)$, then $j' \in B_i$. To show that any such $j' \in B_i$ we must show that (i) $\sigma(j') < i$, which follows since $\sigma(j') < i' < i$ and (ii) $i < j'$, which is true since $j' \in (i, j)$, and (iii) $j' < \sigma(i)$, which is true since $j' < j < \sigma(i)$.

Therefore any element that contributes to $\beta_{i'}(j)$ but not to $\beta_i(j)$ must come from the interval (i', i) and there are at most $i' - i - 1$ such elements. $\square$

Now, we are ready to show that the generalized Kendall tau is never much larger than the Spearman footrule distance.

Lemma 13. *For all σ and metrics D, $K_D(\sigma) \le 3F_D(\sigma)$.*

Proof. The basic idea is to explicitly map each term of K_D to a term of F_D so that no term of F_D has more than three terms of K_D mapped to it.

We charge the inversions of different types separately. For each $\langle i, j\rangle$ inversion of type I the D_{ij} term appears both in K_D and in F_D.

Now suppose that $\langle i, j\rangle$ is an inversion of type II, then the term D_{ij} appears in K_D but not in F_D. Since D defines a metric, by the triangle inequality (7), we can conclude that for any $\mu_i(j)$,

$$D_{ij} \le D_{i\mu_i(j)} + D_{\mu_i(j)j}.$$

Since μ is a valid mapping, property (c) ensures that $\mu_i(j) \in (i, \sigma(i)$ and thus the term $D_{i\mu_i(j)}$ appears in F_D. Similarly, since $\mu_i(j) \in (j, \sigma(j))$, the term $D_{\mu_i(j)j}$ appears in F_D. Thus we have can charge each inversion of type II to two terms that appear in F_D. To complete the proof, observe that no term in F_D can be charged more than three times: the first time by the inversions of type I, and then at most twice by the type II inversions. The latter claim follows from properties (a) and (b) of μ. $\square$

Having shown that $K_D < 3F_D$, we proceed to showing the converse. The proof follows the same structure as the proof of Lemma 13. We begin by categorizing the types of terms that appear in F_D into those that appear in K_D as well, and those that do not. For the latter group, we construct a mapping ν that maps each term to two inversions in K_D so that no inversion is mapped to at most twice.

We start by dividing all terms that appear in $F_D(\sigma)$ into two categories.

(1) Type I term: a pair $\langle i, j\rangle$ is a term of *type I* if $\sigma(j) \in [i, \sigma(i))$ and $j > i$. Then $\langle i, j\rangle$ form an inversion and the term D_{ij} appears both in F_D and in K_D.

(2) Type II term: a pair $\langle i, j\rangle$ is a term of *type II* if $\sigma(j) \in [i, \sigma(i))$ and $j < i$. Then the term D_{ij} appears in F_D but does not appear in K_D since i and j are

We will map each term of type II to a pair of inversions. To define the mapping, let $P_i = \{j : j < i \le \sigma(j) < \sigma(i)\}$ (these are the items that form type II terms with i) and $Q_i = \{k : \sigma(k) \le i < k\}$.

Definition 14. *A mapping $\nu : \cup_i P_i \to \cup_i Q_i$ is* valid *if it satisfies the following four properties:*

(a) $\forall i, j, j' \in P_i \quad \nu_i(j) \ne \nu_i(j')$.

(b) $\forall i, i', j \in P_i \cap P_{i'} \quad \nu_i(j) \ne \nu_{i'}(j)$.

(c) $\forall i, j \in P_i \quad i$ and $\nu_i(j)$ form an inversion.

(d) $\forall i, j \in P_i \quad j$ and $\nu_i(j)$ form an inversion.

As before, we construct a mapping ν and prove that it is a valid mapping. Let P_i and Q_i as above. For an element $j \in P_i$ denote by $\pi_i(j) = |\{j' \in P_i : i \le \sigma(j') < \sigma(j)\}|$ the number of elements in P_i that precede $\sigma(j)$. For an element $k \in Q_i$ denote by $\xi_i(k) = |\{k' \in Q_i | k' < k\}|$. We define the mapping ν as

$$\nu_i(j) = \{k : \pi_i(j) = \xi_i(k)\}.$$

Lemma 15. *The mapping ν is valid.*

Proof. To show property (a), fix i and consider $j, j' \in Q_i$. Sine j and j' have different π_i values, $\nu_i(j) \ne \nu_{i'}(j)$.

Now consider i, i' with $i > i'$ and an element $j \in Q_i \cap Q_{i'}$. We claim that $\nu_i(j) \ne \nu_{i'}(j)$. Suppose otherwise, and let $k = \nu_i(j) = \nu_{i'}(j)$. Then $\pi_{i'}(j) - \pi_i(j) = \xi_{i'}(k) - \xi_i(k)$. From the definition of π, we have that $\pi_{i'}(j) - \pi_i(j) = i - i'$.

On the other hand, consider the maximum value of $\xi_{i'}(k) - \xi_i(k)$. The elements that contribute to $\xi_i(k)$ can be divided into three groups. Those in the interval (i, i'), the element i' and those in the interval (i', k). To maximize $\xi_i(k)$, we can assume that all of the elements in the first interval contribute to it; there are $i - i' - 1$ such elements. The element i is not in $Q_{i'}$ therefore it does not contribute to the value of $\xi_{i'}(k)$. Finally, for elements $k' \in (i, k)$, if any such $k' \in Q_{i'}$ then it is also in Q_i, and therefore contributes equally to $\xi_i(k)$ and to $\xi_{i'}(k)$. Therefore the maximum value for $\xi_i(k) - \xi_{i'}(k) = i - i' - 1$.

It is easy to check that properties (c) and (d) follow. $\square$

We can now show that the generalized Spearman distance is never much larger than the generalized Kendall tau.

Lemma 16. *For all σ and metrics D, $F_D(\sigma) \le 3K_D(\sigma)$.*

Proof. We charge each term separately. For terms of type I, D_{ij} appears both in F_D and in K_D. Let D_{ij} be a term of type II. By the triangle inequality (7), we can conclude that $D_{ij} \le D_{i\nu_i(j)} + D_{\nu_i(j)j}$ for any $\nu_i(j)$. Furthermore, since ν is a valid metric, $\nu_i(j)$ forms inversions with both i and j, and therefore the terms $D_{i\nu_i(j)}$ and $D_{\nu_i(j)j}$ appear in K_D. Finally, any term D_{ij} can be charged at most three times — once by terms of type I and at most twice by terms of type II. The proof is complete. $\square$

Finally, putting all of these together, we obtain the main result.

Theorem 17 (Generalized DG inequality). *For all σ, $(1/3)F^*(\sigma) \le K^*(\sigma) \le 3F^*(\sigma)$.*

We note that the left inequality is tight by giving an example where $F^* = 3K^*$. On the other hand we conjecture that $K^* \le F^*$, but proving that remains an open problem. To show that $F^* = 3K^*$, consider a simple permutation on three elements $\sigma([abc]) = [bca]$. There are two inversions (namely a is inverted with both b and c) and thus $K^* = D_{ab} + D_{ac}$. On the other hand $F^* = D_{ab} + D_{ac} + D_{bc} + D_{ca}$. Setting $D_{ac} = D_{bc} = 1$ and $D_{bc} = 0$ ensures that D is a metric and $K^* = 1$ while $F^* = 3$.

5. DISCUSSION

We have presented a generalized version of Kendall's tau and Spearman's footrule metrics that take into account the element weights, their position in the permutation and pairwise distance when computing the distance between two permutations. Importantly, these two new metrics address the desiderata put forth in the Introduction. They are rich enough to handle the modern requirements on a metric, especially in the context of web search results. They remain relatively straightforward: for example, the generalized Kendall's tau simply scales the score of each inversion by the pairwise distance between the inverted element and their element and position weights. They are indeed true generalizations: when all of the weights, and distances are set to 1, we recover the original Kendall's tau and footrule distances. Although not at first evident, we proved in Section 4.1 that the two measures satisfy all of the basic properties — they are right invariant, scale free, and satisfy the triangle inequality. Finally, we showed that these two metrics are equivalent, and regardless of the setting of the element and position weights and the instantiation of the metric space D, they will always be within a factor of three from each other. Below, we give a practical application of this invariance.

5.1 Generalized rank aggregation

Consider, the rank aggregation problem: given a set of disparate rankings: $\sigma_1, \ldots, \sigma_k$, find one that minimizes the total disagreement. Since there are two ways to measure disagreement we can define two versions of the problem. In the footrule version, we strive to find a permutation σ_F^*, that minimizes the total footrule distance $\sum_{i=1}^{k} F(\sigma^{-1}\sigma_F^*)$. In the Kendall version, we wish for a permutation σ_K^* that minimizes the total Kendall distance: $\sum_{i=1}^{k} K(\sigma^{-1}\sigma_K^*)$. A priori, both problems look daunting: a direct approach to finding σ_K^* is to solve the feedback arc set problem (FAS). Here we have a node for each element, and a directed edge between i and j with the weight set to the number of permutations in which i precedes j. Solving the FAS problem on the resulting tournament graph (since there is an edge between any pair of nodes) would lead to an optimal σ_K^*. Unfortunately the FAS problem is NP-hard, even on tournament graphs [2, 16], so we can at best hope for an approximation; even so while a PTAS for this problem exists, it is far from practical. On the other hand, Dwork et al. [10] show how to find σ_F^* in polynomial time via a simple minimum-cost perfect matching algorithm. The equivalence of F and K allows us to to conclude that σ_F^* is a non-trivial 2-approximation to σ_K^*. A trivial 2-approximation exists because the distance is a metric: one of the input rankings achieves this factor.

The equivalence can work in the opposite direction as well. Consider the same rank aggregation problem with generalized footrule and Kendall's tau. The addition of element, position and distance weights means the minimum cost matching formulation used by [13] no longer applies. However, the FAS formulation still holds — the cost of each directed (i, j) edge becomes the total contribution to the (i, j)th term of K^* by the rankings where $i > j$. Now we can find an approximation to σ_K^* either by using the PTAS in [16] or the Quick-Select algorithm in [2].

6. EXPERIMENTS

In general it is not easy to evaluate a new ranking function, especially due to the lack of a suitable ground truth. Instead of focusing on how our new measures correlate with widely-used relevance measures such as mean-average precision or DCG, we focus on the *robustness* of our measures. In particular, we study how our generalizations utilize the additional information (such as the element weights, or the position weights) available to them in order to provide a better understanding on rankings. We also focus on a specific application in web search: finding queries where a particular search engine performs badly.

6.1 Data

The data we use for our experiments are from two real-world sources. Our first dataset, called D_1, is a subset of search clicks from Yahoo! search, for a week period in September 2009. For each query with at least a thousand total clicks, we track the number of clicks at each position and retain these counts only for the top ten positions. The resulting dataset has about 80,000 queries. Note that the number of clicks naturally induces a permutation (there may be some ties, but these will be broken arbitrarily) on $\{1, \ldots, 10\}$.

We will use this permutation in our experiments, with three different types of position weights, in addition to the unweighted case (called UNIT).

(1) DCG: Here, we set

$$\delta_i = \frac{1}{\log(i+1)} - \frac{1}{\log(i+2)}.$$

This corresponds to weighing each position by the standard position weight used in DCG computations.

(2) CTR: Here, we set $\delta_i = \mathrm{ctr}_i - \mathrm{ctr}_{i+1}$, where ctr_i is the aggregated click-through rate at position i. This corresponds to penalizing each swap by the amount of the click-through rate lost. Table 1 shows the actual values we used in our experiments.

(3) TOPK: Here, we use a step function to simulate the top k ranking. For $i \leq 5$, we set $\delta_i = 1$ and for $i > 5$, we set $\delta_i = 0$. This means there is no penalty for swapping after the fifth position.

Our second dataset, called D_2, is a subset of human-labeled judgments for a number of queries. Each query has the top five results judged into four scales from 0 (meaning worst) to 3 (meaning best). This data has about 140,000 queries and its associated judgments. The judgments induce a permutation (actually, a partial order, but we will break ties arbitrarily) on the top five positions; we will use this permutation in our experiments. The editorial labels give a natural weighting for elements (called EDIT). As before, we use UNIT to denote the unweighted case.

6.2 Tightness of the equivalence

Our first experiment concerns the tightness of the inequality given by Theorem 17. We are interested in how much the Kendall's tau and Spearman footrule vary (as a ratio) for various choices of the position weights. To do this, we use the dataset D_1. The top panel in Figure 1 shows the scatter plot of Kendall vs footrule values, for the DCG position weight. The two envelopes are the lines $y = x$ and $y = 2x$. From the plot we can see that if the Kendall value itself is small, the footrule value is nearly twice as big, with the majority of the points near the upper envelope. On the

Position i	1	2	3	4	5	6	7	8	9	10
ctr_i	.488	.146	.089	.066	.051	.041	.033	.029	.027	.027

Table 1: Click-through rates used in our experiments as position weights.

other hand, if the Kendall's tau value is itself large, then the footrule value is closer to the Kendall's tau. In contrast, the bottom panel in Figure 1 shows the same plot for D_1, using the unweighted versions of the metrics. Here, the points are always distributed more towards the upper envelope showing a greater discrepancy between the two measures. This can be explained by the fact that when the Kendall's tau value is large, the permutations are very different, and DCG position weights reduce the effect in the lower positions. This is precisely what we want in a ranking measure.

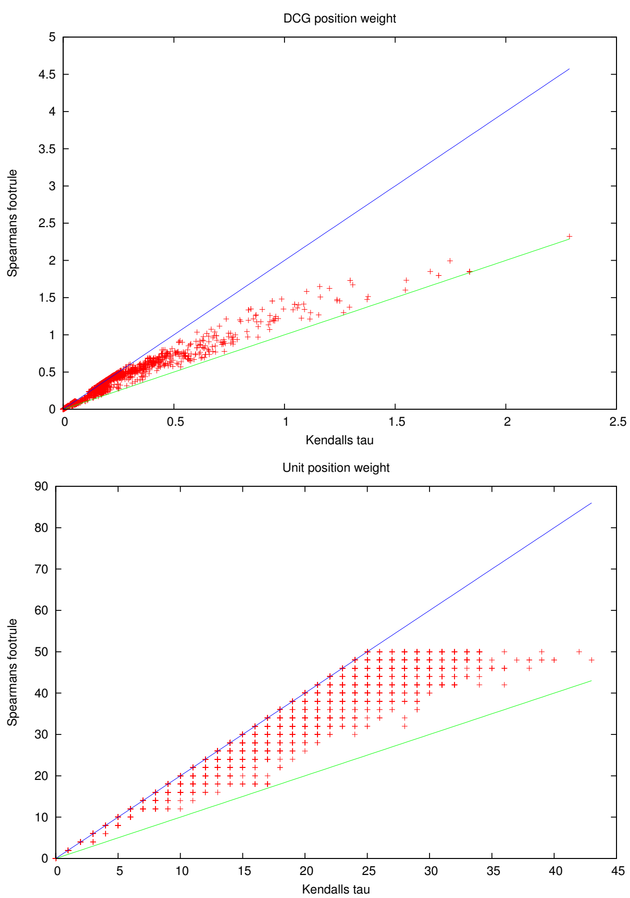

Figure 1: The ratio of the measures for D_1 with DCG position weight and with unit weights.

6.3 Correlation among the measures

In practice, should one use Kendall's tau or the footrule measure? Both have their advantages and disadvantages, ranging from computability to interpretability. A point to note is that even though Theorem 17 shows that generalized measures are equivalent, it is not clear how much these measures themselves are correlated: given two permutations σ, τ, how correlated are the four quantities $K^*(\sigma)$, $K^*(\tau)$, $F^*(\sigma)$, $F^*(\tau)$? Ideally, we would like to show that the weighted versions are better correlated with one another in practice than their unweighted counterparts. We study the error in correlation for two sequences $X = \{x_i\}$ and $Y = \{y_i\}$, defined as follows:

$$\epsilon(X, Y) = 1 - \frac{E[XY] - \mu(X)\mu(Y)}{\sigma(X)\sigma(Y)},$$

where $\mu(X) = E[X], \sigma^2(X) = E[(X - E[X])^2]$.

In our application, for a particular setting of either the position weights or element weights, we take X to be the Spearman's footrule values and Y to be the Kendall's tau values; for such a choice, we define Z to be the point-wise ratio $Z = \{x_i/y_i\}$. We study $E[Z]$, the average ratio of the footrule to Kendall's tau values, $\sigma(Z)$, its standard deviation, and $\epsilon(X, Y)$, the correlation error between the two values. Table 2 shows these values for our datasets and various weighting schemes.

	$E[Z]$	$\sigma(Z)$	$\epsilon(X,Y)$
		D_1	
UNIT	1.756	0.248	0.040
DCG	1.756	0.273	0.022
CTR	1.806	0.268	0.026
TOPK	1.767	0.262	0.026
		D_2	
UNIT	1.799	0.354	0.058
EDIT	1.721	0.342	0.040

Table 2: Correlation between weighted Kendall and footrule measures.

As we can see from the results, all of the weighted versions have much less correlation error than the unweighted versions, for both D_1 and D_2. The average value of the ratio is away from 2 (the upper bound) in all cases. Interestingly, the variance is more for the weighted measures, this, once again, can be explained by their emphasis on the higher positions.

6.4 Finding the worst queries

In this section we illustrate the robustness of our weighted measures via a simple application: the goal is to find the worst-performing queries in a search engine. For D_1, these are the queries whose click ordering is furthest away from the result ordering. For D_2, these are the queries whose editorial judgments are disagree the most with the result ordering.

We wish to illustrate the robustness of the weighted measures by asking the following question: suppose one were to use the unweighted Kendall's tau or footrule measure to pull out the k most egregious queries? How sensitive is this to the choice of the metric, i.e., whether we use Kendall's tau or the footrule measure? In other words, how consistent are the top k (worst) sets returned by Kendall's tau and footrule measures?

We measure this consistency by computing the normalized set intersection of the top k queries deemed worst by Kendall's tau and the top k queries deemed worst by the footrule measure. We compare this intersection with that of using the weighted counterparts. The top panel of Figure 2 shows the result for D_1 for the DCG position weight. It is easy to see that the weighted versions return very similar queries. This holds for even moderately large value of k; of course, for large enough k, these curves will start converging. A similar pattern holds for D_2. The bottom panel of Figure 2 shows the results.

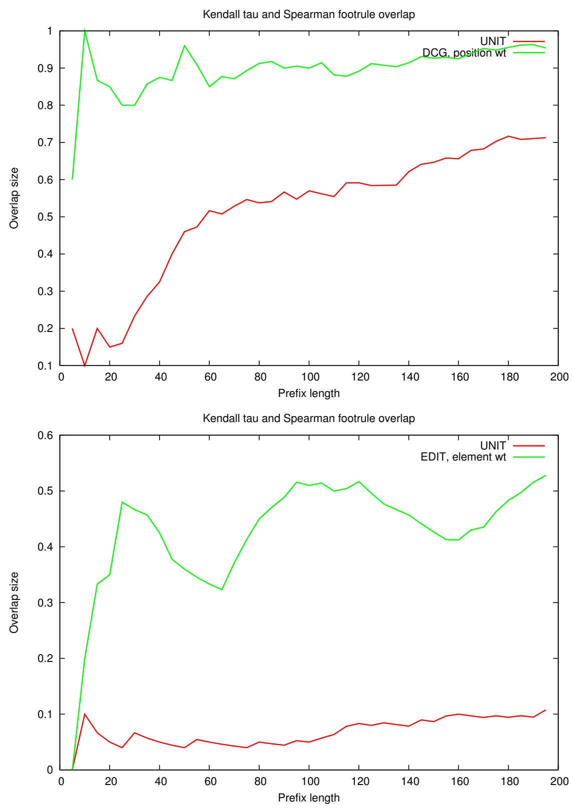

Figure 2: Correlation between the unweighted and weighted measures for D_1 with DCG position weights and for D_2 with EDIT element weights.

7. CONCLUSIONS

We presented generalizations of the Spearman's footrule and Kendall's tau distance metrics to take into account element weights, position weights, and the relative similarity between elements. These generalizations encode some of the common metrics used in practice. For example, by setting d_{ij} to be 1 between the pairs of elements in the top k position, 0 for elements in positions $k+1, \ldots, n$, and p for the distances between the two partitions, we recover a top-k metric for Kendall's tau.

We remark that our definitions lead to an intriguing metric between two orderings of nodes in a graph. Given a weighted graph $G = (V, E)$, let D_{ij} be the shortest distance between two nodes $i, j \in V$. The distance between any two orderings can be computed using the F_D or K_D metrics that we have defined. As an application, consider a co-authorship graph, and rank nodes (i.e., authors) based on how often certain keywords, (e.g., internet, routing, advertising, etc) appears in their titles. A footrule or Kemeny optimal rank aggregation might lead to finding interesting co-authorship patterns.

Acknowledgments

We thank Olivier Chappelle and Kunal Punera for sharing their data, Ali Dasdan and Paolo D'Alberto for sharing an early version of their manuscript, and D. Sivakumar for helpful conversations.

8. REFERENCES

[1] R. Agrawal, S. Gollapudi, A. Halverson, and S. Ieong. Diversifying search results. In *Proc. 2nd WSDM*, pages 5–14, 2009.
[2] N. Ailon, M. Charikar, and A. Newman. Aggregating inconsistent information: Ranking and clustering. *J. ACM*, 55(5), 2008.
[3] C. Buckley and E. Voorhees. Retrieval evaluation with incomplete information. In *Proc. 27th SIGIR*, pages 25–32, 2004.
[4] B. Carterette. On rank correlation and the distance between rankings. In *Proc. 32nd SIGIR*, pages 436–443, 2009.
[5] O. Chapelle, D. Metlzer, Y. Zhang, and P. Grinspan. Expected reciprocal rank for graded relevance. In *Proc. 18th CIKM*, pages 621–630, 2009.
[6] N. Craswell, O. Zoeter, M. Taylor, and B. Ramsey. An experimental comparison of click position-bias models. In *Proc. 1st WSDM*, pages 87–94, 2008.
[7] P. D'Alberto and A. Dasdan. Content-based search engine rank correlation. Manuscript, 2009.
[8] P. Diaconis. *Group Representation in Probability and Statistics*. Number 11 in IMS Lecture Series. Institute of Mathematical Statistics, 1988.
[9] P. Diaconis and R. Graham. Spearman footrule as a measure of disarray. *Journal of the Royal Statistical Society B (Methodological)*, 39(2):262–268, 1977.
[10] C. Dwork, R. Kumar, M. Naor, and D. Sivakumar. Rank aggregation methods for the web. In *Proc. 10th WWW*, pages 613–622, 2001.
[11] R. Fagin, R. Kumar, M. Mahdian, D. Sivakumar, and E. Vee. Comparing partial rankings. *SIAM J. Discrete Math.*, 20(3):628–648, 2006.
[12] R. Fagin, R. Kumar, and D. Sivakumar. Comparing top k lists. In *Proc. 14th SODA*, pages 28–36, 2003.
[13] R. Fagin, R. Kumar, and D. Sivakumar. Efficient similarity search and classification via rank aggregation. In *Proc. SIGMOD*, pages 301–312, 2003.
[14] K. Järvelin and J. Kekäläinen. IR evaluation methods for retrieving highly relevant documents. In *Proc. 23rd SIGIR*, pages 41–48, 2000.
[15] M. G. Kendall. A new measure of rank correlation. *Biometrika*, 30(1/2):81–93, 1938.
[16] C. Kenyon-Mathieu and W. Schudy. How to rank with few errors. In *Proc. 39th STOC*, pages 95–103, 2007.
[17] D. Sculley. Rank aggregation for similar items. In *SDM*, 2007.
[18] G. Shieh, Z. Bai, and W. Tsai. Rank tests for independence—with a weighted contamination alternative. *Statistica Sinica*, 10:577–593, 2000.
[19] G. S. Shieh. A weighted Kendall's tau statistic. *Statistics & Probability Letters*, 39(1):17–24, 1998.
[20] C. Spearman. The proof and measurement of association between two things. *The American Journal of Psychology*, 15(1):72–101, 1904.
[21] E. Yilmaz, J. A. Aslam, and S. Robertson. A new rank correlation coefficient for information retrieval. In *Proc. 31st SIGIR*, pages 587–594, 2008.