Comparing Scores Intended for Ranking

Narayan L. Bhamidipati and Sankar K. Pal, *Fellow, IEEE*

Abstract—Often, ranking is performed on the basis of some scores available for each item. The existing practice for comparing scoring functions is to compare the induced rankings by one of the multitude of rank comparison methods available in the literature. We suggest that it may be better to compare the underlying scores themselves. To this end, a generalized Kendall distance is defined, which takes into consideration not only the final ordering that the two schemes produce but also at the spacing between pairs of scores. This is shown to be equivalent to comparing the scores after fusing with another set of scores, making it theoretically interesting. A top k version of the score comparison methodology is also provided. Experimental results clearly show the advantages score comparison has over rank comparison.

Index Terms—Score comparison, rank comparison, Kendall distance, top k lists.

1 Introduction

Ranking a set of items is a fairly frequent task and involves pairwise comparison of the given items. This comparison may be performed by inquiring an oracle for each pair of items, in which case, the ranking procedure is known as comparison-based ranking. On the other hand, one may assign scores to each item, thus producing a total ordering on the set of items. Each item is assigned a score that denotes how early the item appears in the list, and thus, comparing each pair is now performed by comparing the corresponding scores. The present work is concerned with score-based rankings only, and each ranking is assumed to be induced by a *scoring scheme* or *function*.

Several scoring schemes may compete with each other for ranking the same set of items, and the items may be ranked differently by each of them. Given two such scoring schemes, two questions arise:

- Which scheme is better?
- How different are the two schemes?

In this paper, we are concerned with the answers to the second question. One may note that it is not sufficient to ask the first question alone as, instead of just declaring one of them to be better than the other, it is imperative to measure how much better one scoring is over the other.

Comparing such scoring schemes is generally performed by comparing the induced ranking on the set of items. However, several different scoring systems may lead to the same ranking of the items, and a rank-based comparison cannot discriminate between such schemes. In this paper, we propose a more general approach, whereby the scoring schemes may be perceived to be different even if they induce identical rankings.

Our approach is based on the idea that similar scoring schemes discriminate between two items in a similar manner. If the scores assigned to items i and j by one scoring scheme are far apart, while those by another are very close to each other, it indicates that the two schemes are dissimilar. It is interesting to note that this also corresponds to a fusion-based approach for measuring similarity/dissimilarity of scoring schemes. Often, these scoring schemes are used in combination with some other scoring method, say T, to produce the final ranking [1], [2], [3], [4]. This process of combining scores is referred to as *score fusion* [2], [3]. If two sets of scores are exactly the same, their rankings remain the same even after score fusion. Also, differences in the scores assigned by two methods may lead to different rankings, depending on the scores used for fusion. If we know T beforehand, then we may rank the items after fusing their scores and compute a dissimilarity value on the basis of the induced rankings.

In this paper, we look at the case where T remains unknown. Based on certain simple assumptions about this unknown scoring scheme T, we compare two scoring schemes on the basis of how likely they are to produce a discordant pair. We provide a metric in this regard, which relies on the margins separating the scores. Even if two items receive almost equal scores, they might be ranked differently depending on T. The present investigation is about studying how likely it is for them to be ranked differently upon score fusion.

This paper is organized as follows: We introduce the notation and background for comparing scores and rankings, and rank fusion in Section 2. The proposed methodology is described in Section 3, which includes motivational examples and a discussion on the characteristics of our method. Section 4 deals with extending the proposed metric for comparing top k scorings, and applications of the metric are discussed in Section 5. We report some preliminary experimental results in Section 6, before drawing our conclusions and mentioning future work in Section 7.

- N.L. Bhamidipati is with the Data Mining and Research Group, Yahoo! Software Development India Pvt. Ltd., "Torrey Pines," Embassy Golf Links Business Park, Off Indiranagar-Koramangala Intermediate Ring Road, Bangalore 560 071, India. E-mail: narayan_bl@yahoo.com.
- S.K. Pal is with the Indian Statistical Institute, 203 B.T. Road, Kolkata 700108, India. E-mail: sankar@isical.ac.in.

2 Comparing Scoring Functions: Background

2.1 Notation

Our universe consists of a set of objects or items indexed by $\Omega = \{1, 2, 3, \ldots, n\}$, and each of them shall be referred to by its index. Unless otherwise stated, i and j refer to two elements of Ω, and $i < j$. These objects or items may be documents in corpus, states in a country, students in a university, and so on. A scoring scheme or function assigns a real number $s_i \in [0, 1]$, called a score, to i, for each $i \in \Omega$. In the present work, only normalized scoring schemes shall be considered, i.e., $\max_i s_i = 1$ and $\min_i s_i = 0$. The kth scoring scheme, or equivalently, the kth score vector is denoted by $S_k = (s_{k1}, s_{k2}, \ldots, s_{kn})$. We use $R(s_i)$ to denote the rank of i, and $R(S)$ is an abbreviation for $(R(s_1), R(s_2), \ldots, R(s_n))$. Objects with larger scores appear earlier in the ranked list, so that $s_i > s_j \Rightarrow R(s_i) < R(s_j)$.

Let S_1 and S_2 be two scoring schemes. A pair (i, j) is called discordant with respect to S_1 and S_2, if $(s_{1i} - s_{1j})(s_{2i} - s_{2j}) < 0$, i.e., the two schemes order i and j in different ways. If $(s_{1i} - s_{1j})(s_{2i} - s_{2j}) > 0$, then i and j are said to be concordant. The comparison is called a tie if $(s_{1i} - s_{1j})(s_{2i} - s_{2j}) = 0$. A tie may occur in one of three ways: $s_{1i} = s_{1j}$, $s_{2i} = s_{2j}$, or both, and it is called a 1-tie, a 2-tie, or a double tie, respectively. Without loss of generality, we assume that the first set of scores is sorted.

Assumption 1 (Monotonicity of S_1). $1 = s_{11} \geq s_{12} \geq \cdots \geq s_{1n} = 0$, *for, otherwise, we may sort S_1 and S_2 in descending order with S_1 as the primary key.*

Let S and $T = (t_1, t_2, \ldots, t_n)$ be two score vectors. $\alpha S + \beta T$ means that the score vectors S and T are *fused together*, with the fusing proportions $0 < \alpha < 1$ and $\beta = 1 - \alpha$. So, the ith element of the fused vector is $\alpha s_i + \beta t_i$. When $\alpha = \beta = 0.5$, we shall simply write $S + T$, instead of the technically correct $0.5S + 0.5T$, noting that the ranking remains the same in both cases.

We now provide some background on comparing rankings and fusion.

2.2 Background on Rank Comparison

Comparison of rankings is a fairly well-studied problem, and we mention the most popular rank comparison methods here. Comparing two different rankings has been studied in various fields. In each case, a measure has been provided that takes into account how much the positions of each item differ in the two ordered lists. The measure is zero when the two rankings are exactly the same, whereas it is maximum when the rankings are completely opposite to each other. Some very useful and widely used measures for comparing two rankings are Spearman's footrule, Spearman's rank correlation, and Kendall's τ.

Spearman's footrule is defined as

$$\rho_1 := \sum_{i=1}^{n} |R(s_{1i}) - R(s_{2i})|. \tag{1}$$

Spearman's rank correlation [5] is defined as

$$\rho_2 := \left(\sum_{i=1}^{n} (R(s_{1i}) - R(s_{2i}))^2 \right)^{\frac{1}{2}}. \tag{2}$$

Both ρ_1 and ρ_2 are 0 if both the rankings are the same, and they attain their maximum values when $R(S_{1i}) = n + 1 - R(S_{2i})$, $\forall i \in \Omega$.

Kendall's Tau (or τ) [5] is defined as the difference of the proportions of concordant and discordant pairs according to S_1 and S_2:

$$\tau(S_1, S_2) = \frac{2}{n(n-1)} \sum_{i<j} sign(s_{1i} - s_{1j})(s_{2i} - s_{2j}), \tag{3}$$

and may be rewritten in terms of only the number of discordant pairs as

$$\tau(S_1, S_2) = 1 - \frac{4}{n(n-1)} \sum_{i<j} I_{[(s_{1i}-s_{1j})(s_{2i}-s_{2j}) < 0]}. \tag{4}$$

The summation in (4) is the number of discordant pairs with respect to S_1 and S_2 and is referred to as the Kendall distance between them [6]. Formally, the Kendall distance between each pair $\{i, j\}$ with respect to S_1 and S_2 is defined as

$$K(S_1, S_2; i, j) = \begin{cases} 1 & \text{if} \quad i \text{ and } j \text{ are discordant,} \\ \frac{1}{2} & \text{if} \quad i \text{ and } j \text{ have a single tie,} \\ 0 & \text{o.w.,} \quad \text{i.e., if } i \text{ and } j \text{ are concordant,} \\ & \qquad\quad \text{or have a double tie.} \end{cases} \tag{5}$$

When this quantity is summed over all the pairs of items, then it is called the Kendall distance between S_1 and S_2, or equivalently, between $R(S_1)$ and $R(S_2)$, and is denoted by $K(S_1, S_2)$ or $D(R(S_1), R(S_2))$. This is also referred to as Kendall (Tau) distance, Kemeny distance, or bubblesort distance between $R(S_1)$ and $R(S_2)$ when interpreted as the number of pairwise adjacent transpositions needed to transform from one ranked list to the other.

More recently, Bar-Ilan et al. proposed that differences in ranking in the initial part of the lists should be given more weight than those toward the end of the lists [7]. The dissimilarity between the two rankings is computed as

$$\mu = \sum_{i=1}^{n} \left| \frac{1}{R(S_1(i))} - \frac{1}{R(S_2(i))} \right|. \tag{6}$$

2.3 Background on Fusion

Fusion is the process of combining multiple sets of ranks or scores available for the given items. Rank fusion [8], [9], also known as rank aggregation [6], [10], obtains a consensus ranking from the available ranked lists. These lists need not be full lists, making rank fusion a very challenging problem.

Score fusion, on the other hand, combines the scores directly, in order to produce a consensus score vector, on which the final ranking may be based upon. Such fusion may be performed by taking an average of the scores assigned to an item. Two of the standard score fusion techniques are CombSUM (a simple average) and CombMNZ (a weighted average) [11], [4].

Several studies have compared the effectiveness of rank and score fusion. Scores contain more information than ranks but may be prone to noise. It is suggested in [6] that only the induced ranks should be considered for fusion, whereas in [12], it is found that score fusion is advantageous, provided that normalization is performed properly. A detailed discussion on ranks versus scores is available in [13].

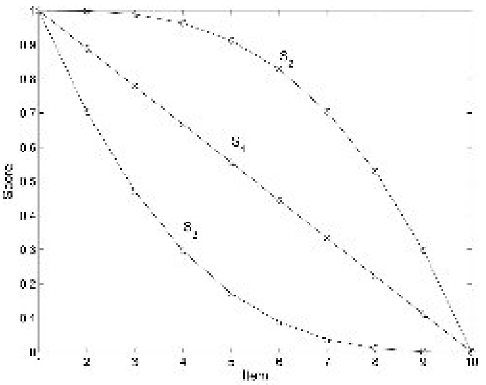


Fig. 1. Same ranks, different scores.

3 Comparing Underlying Scores Directly

One may compare two score vectors directly using a measure like Pearson's correlation coefficient. However, the interpretation of the coefficient in terms of the resultant ranking is lost. Also, the correlation coefficient is not a metric and hence cannot be interpreted directly as a distance between two scoring systems. The correlation coefficient may be transformed into a metric, but it still does not reflect the rank-specific differences between two scorings and is not always useful for comparing rank-inducing scoring functions.

An alternative is to compare the scores assigned to the available objects on the basis of the rankings they produce.

3.1 Motivation

We shall now emphasize upon the significance of comparing scorings directly. We start by asking the following question: "Is it sufficient to use only $R(S_1)$ and $R(S_2)$ for comparing S_1 and S_2?" We look at the following examples to gain some insight in this regard.

Example 1. Suppose that items i and j receive identical scores in each scoring scheme, i.e., $s_{1i} = s_{1j}$ and $s_{2i} = s_{2j}$. By definition, there is a double tie between i and j with respect to S_1 and S_2. Now, if there is a measurement error, due to which the scores are slightly perturbed, i and j would be declared to be either concordant or discordant (with probability of 1), though, in reality, they are neither of the two. This is a consequence of the fact that discordance is a hard concept, and a pair may be either discordant or not but nothing in between.

Example 2. Let $n = 10$. The objects $1, 2, \ldots, 10$ are scored in three different ways, as shown in Fig. 1. Here, $s_{1i} = \frac{n-i}{n-1}$, $s_{2i} = 1 - \frac{(i-1)^2}{(n-1)^2}$, $s_{3i} = \frac{(n-i)^2}{(n-1)^2}$. It may be noted that, while the ranks are identical in all three cases, the scores differ significantly. For example, items 1, 2, and 3 have barely distinguishable scores with respect to S_2, whereas the scores are quite varied in the cases of S_1 and S_3. If ranking by themselves is the sole objective of the three scoring functions, then they may be deemed identical. Otherwise, that the resolving power of the three scoring functions is different implies some amount of dissimilarity between them.

Example 3. Let $(s_{1i}, s_{1j}, s_{2i}, s_{2j}, s_{3i}, s_{3j}) = (0.4, 0.5, 0.5, 0.4, 0.9, 0.1)$. So, the items i and j are discordant with respect to S_1 and S_2, as well as, with respect to S_1 and S_3. However, are they "more" discordant in the second case? Again, this question cannot be answered without the notion of a degree of discordance.

Example 4. Let $s_{12} = 0.7$, $s_{13} = 0.4$, $s_{22} = 0.6$, and $s_{23} = 0.5$. Assume $s_{1k} = s_{2k} \forall k \in \Omega \setminus \{2, 3\}$. Note that items 2 and 3 are concordant with respect to S_1 and S_2 and also that both S_1 and S_2 induce the same rankings, i.e., $R(S_1) = R(S_2)$. Now, suppose that the objects are to be ranked after fusing their scores with T. So, the two rankings obtained are $R(S_1 + T)$ and $R(S_2 + T)$. The question we are concerned with is whether these two rankings are identical. It is obvious that the answer depends on the values of t_2 and t_3. For example, if $t_2 = 0.3$ and $t_3 = 0.5$, then the fused scores are given by $s_{12} + t_2 = 1.0 > s_{13} + t_3 = 0.9$ and $s_{22} + t_2 = 0.9 < s_{23} + t_3 = 1.0$, and hence (2, 3) forms a discordant pair according to $S_1 + T$ and $S_2 + T$. One may easily observe that (2, 3) would form a discordant pair whenever $t_3 - t_2 \in (0.1, 0.3)$.

Example 5. Now, if $s_{12} = 0.9$ and $s_{13} = 0.1$, while s_{22} and s_{23} remain the same as in Example 4, 2 and 3 again form a concordant pair with respect to S_1 and S_2. However, (2, 3) forms a discordant pair with respect to $S_1 + T$ and $S_2 + T$ whenever $t_3 - t_2 \in (0.1, 0.8)$. In a sense, it is more likely for (2, 3) to be discordant in this case than in the earlier one.

The essence of these examples was to demonstrate that even though S_1 and S_2 may appear identical or similar on the basis of the rankings they produce by themselves, the likelihood of a discordant pair being produced after score fusion depends both on the distribution of $t_j - t_i$ (for all pairs $i < j$) and the spacing between the scores assigned to the objects of the universe.

Another compelling reason for comparing scores directly is that given just S_1 and S_2, rank comparison methods have no way of distinguishing between the cases when the fusing parameter, α, is big (say, 0.9) or small (say, 0.1). In isolation, as long as both the vectors are multiplied by the same scalar, rank comparison measures come up with the same value each time. Of course, if all one needs to do is to rank the items on the basis of just S_1 and S_2, then $R(S_1)$ and $R(S_2)$ should suffice for comparing the scores, i.e., score comparison methods provide no additional advantage.

3.2 Comparing Scores Directly

The objective in the present investigation is to discern between two scoring functions directly without performing the additional task of computing the induced ranks. Taking a cue from Examples 1 and 3, we propose the concept of a degree of discordance for a pair of items, which subsumes the usual definition of discordance as a special case. As discussed in Example 2, the dissimilarity of two scoring functions with respect to a pair $\{i, j\}$ may be inferred from the differences in the separation of i and j by the scoring functions. Thus, a measure of dissimilarity between S_1 and S_2 may be based on the separations $d_{ij}^{(1)} = s_{1i} - s_{1j}$ and $d_{ij}^{(2)} = s_{2i} - s_{2j}$. The more $d_{ij}^{(1)}$ and $d_{ij}^{(2)}$ are apart, the higher the dissimilarity.

We now look at an alternative approach, which leads to the same notion of discordance once again. In particular, we would like to study how likely it is for a discordant pair to appear during score fusion.

In this regard, let us formalize our observations in Examples 4 and 5 in Section 3.1. The fused scores of i with respect to S_1+T and S_2+T are $s_{1i}+t_i$ and $s_{2i}+t_i$, respectively, for each $1 \le i \le n$. The pair (i,j) forms a discordant pair with respect to S_1+T and S_2+T, if and only if

$$\big((s_{1i}+t_i)-(s_{1j}+t_j)\big)\big((s_{2i}+t_i)-(s_{2j}+t_j)\big)<0,$$

or equivalently, if and only if

$$\big((t_j-t_i)-(s_{1i}-s_{1j})\big)\big((t_j-t_i)-(s_{2i}-s_{2j})\big)<0. \quad (7)$$

That the quantity (7) is negative is equivalent to having (t_j-t_i) in the interval

$$\left(\min\left\{d_{ij}^{(1)},d_{ij}^{(2)}\right\},\max\left\{d_{ij}^{(1)},d_{ij}^{(2)}\right\}\right),$$

where $d_{ij}^{(1)}$ and $d_{ij}^{(2)}$ denote the differences $s_{1i}-s_{1j}$ and $s_{2i}-s_{2j}$, respectively. Thus, once again, the dissimilarity is proportional to the difference of $d_{ij}^{(1)}$ and $d_{ij}^{(2)}$. Note that $d_{ij}^{(1)}$ is positive, by Assumption 1, in Section 2.1.

Similarly, it may be easily seen that the pair (i,j) forms a discordant pair with respect to $\alpha S_1+\beta T$ and $\alpha S_2+\beta T$, if and only if t_j-t_i belongs to the interval

$$\left(\min\left\{\frac{\alpha}{\beta}d_{ij}^{(1)},\frac{\alpha}{\beta}d_{ij}^{(2)}\right\},\max\left\{\frac{\alpha}{\beta}d_{ij}^{(1)},\frac{\alpha}{\beta}d_{ij}^{(2)}\right\}\right).$$

Let γ denote the ratio $\frac{\alpha}{\beta}$, and let, for each pair $i<j$,

$$\left.\begin{aligned} a_{ij}^{S_1,S_2} &= \min\{s_{1i}-s_{1j},s_{2i}-s_{2j}\}, \text{ and} \\ b_{ij}^{S_1,S_2} &= \max\{s_{1i}-s_{1j},s_{2i}-s_{2j}\}. \end{aligned}\right\} \quad (8)$$

We note that, for each pair (i,j), there are associated real numbers $a_{ij}^{S_1,S_2}$ and $b_{ij}^{S_1,S_2}$, such that (i,j) is a discordant pair according to $\alpha S_1+\beta T$ and $\alpha S_2+\beta T$ whenever t_j-t_i is in the interval $(\gamma a_{ij}^{S_1,S_2},\gamma b_{ij}^{S_1,S_2})$. For ease of notation, we drop the superscripts S_1 and S_2 when they are clear from the context. As seen earlier, the interval $[a_{ij},b_{ij}]$ holds the key to the likelihood of a discordant pair being produced. For this reason, we propose

$$D_\gamma(S_1,S_2)=\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}D_\gamma(S_1,S_2;i,j), \quad (9)$$

as a measure of discordance between S_1 and S_2, where $D_\gamma(S_1,S_2;i,j)$ is a suitably chosen measure on the interval (a_{ij},b_{ij}). We shall write D as a shorthand for D_1. In this paper, we make the choice of $D_\gamma(S_1,S_2;i,j)$ as

$$D_\gamma(S_1,S_2;i,j)=\int_{\gamma a_{ij}}^{\gamma b_{ij}} f(x)dx=\left|\int_{\gamma(s_{1i}-s_{1j})}^{\gamma(s_{2i}-s_{2j})} f(x)dx\right|, \quad (10)$$

where $f(x)$ is a continuous probability density function with $(-1,1)$ as its support. As a particular case, in the present work, we choose f to be the following triangular density function:

$$f(x)=\begin{cases}1+x & \text{if} \quad -1<x\le 0\\ 1-x & \text{if} \quad 0<x\le 1\\ 0 & \text{o.w.}\end{cases} \quad (11)$$

$D_\gamma(S_1,S_2;i,j)$ ranges between 0 and 1 and shall be called the degree of discordance between i and j with respect to S_1 and S_2. A value of 0 denotes either a double tie or perfect concordance (i.e., $d_{ij}^{(1)}=d_{ij}^{(2)}$), whereas 1 implies perfect (or extreme) discordance (i.e., $d_{ij}^{(1)}d_{ij}^{(2)}=-1$).

The significance of choosing this particular function f is as explained below. The relevance scores lie in the interval [0, 1], and thereby, $t_j-t_i\in[-1,1]$. We make a simplistic assumption that t_i and t_j are iid U(0, 1), in which case the density of t_j-t_i is f. Another reason for choosing the uniform distribution is that it is the least biased prior distribution and corresponds to the fact that nothing else is known about T (that is, there is no particular T on the basis of which S_1 and S_2 are being compared). Now, let I_1 be defined as follows:

$$\begin{aligned} I_1(a,\gamma) &= \int_0^{\gamma a}(1-x)dx \\ &= \begin{cases}\int_0^1(1-x)dx & \text{if} \quad \gamma a\ge 1\\ \int_0^{\gamma a}(1-x)dx & \text{o.w.}\end{cases} \\ &= \begin{cases}\frac{1}{2} & \text{if} \quad \gamma\ge\frac{1}{a}\\ \gamma a-\frac{1}{2}\gamma^2a^2 & \text{o.w.}\end{cases}\end{aligned} \quad (12)$$

Also, one may note that

$$\int_{\gamma a}^{0}(1+x)dx=I_1(-a,\gamma).$$

Thus, $D_\gamma(S_1,S_2;i,j)$ may be evaluated as

$$\begin{aligned} &\int_{\gamma a_{ij}}^{\gamma b_{ij}} f(x)dx \\ &= \begin{cases}\int_0^{\gamma b_{ij}}(1-x)dx-\int_0^{\gamma a_{ij}}(1-x)dx & \text{if } 0\le a_{ij}\le b_{ij}\\ \int_0^{\gamma b_{ij}}(1-x)dx+\int_{\gamma a_{ij}}^{0}(1+x)dx & \text{if } a_{ij}\le 0<b_{ij}\\ \int_{\gamma a_{ij}}^{0}(1+x)dx-\int_{\gamma b_{ij}}^{0}(1+x)dx & \text{if } a_{ij}\le b_{ij}\le 0\end{cases} \\ &= \begin{cases}I_1(b_{ij},\gamma)-I_1(a_{ij},\gamma) & \text{if } 0\le a_{ij}\le b_{ij}\\ I_1(b_{ij},\gamma)+I_1(-a_{ij},\gamma) & \text{if } a_{ij}\le 0\le b_{ij}\\ I_1(-a_{ij},\gamma)-I_1(-b_{ij},\gamma) & \text{if } a_{ij}\le b_{ij}\le 0.\end{cases}\end{aligned} \quad (13)$$

It may be noted that we do not need to consider the case $a_{ij}\le b_{ij}\le 0$, since $b=\max\{d_{ij}^{(1)},d_{ij}^{(2)}\}\ge d_{ij}^{(1)}\ge 0$, by Assumption 1.

3.3 Characteristics and Discussion

It may be easily verified that the proposed measure for comparing two scoring systems, $D_\gamma(S_1, S_2)$, is a pseudometric. It is a metric when $\gamma \leq 1$. The following theorem provides a more general proof, where f is chosen to be any continuous probability density function on $(-1, 1)$ (i.e., taking strictly positive values on $(-1, 1)$ and 0 elsewhere) and is not restricted to the triangular density function.

Theorem 1 (Metric properties of $D_\gamma(S_1, S_2)$). *Let S_1 and S_2 be two normalized score vectors of length n, and let γ be a positive real number. Let the discordance for the pair (i, j) with respect to S_1 and S_2, $D_\gamma(S_1, S_2; i, j)$, be defined as in (10), where a_{ij} and b_{ij} are as defined in (8) and let $D_\gamma(S_1, S_2)$ be defined as in (9). Then, $D_\gamma(S_1, S_2)$ is a metric if $\gamma \leq 1$ and a pseudometric otherwise.*

Proof. To prove the theorem, it needs to be shown that $D_\gamma(S_1, S_2)$ satisfies the following properties:

$$D_\gamma(S_1, S_2) \geq 0 \ \forall\, S_1, S_2, \tag{14}$$

$$D_\gamma(S_1, S_1) = 0, \tag{15}$$

$$\text{If } \gamma \leq 1, D_\gamma(S_1, S_2) = 0 \Rightarrow S_1 = S_2, \tag{16}$$

$$D_\gamma(S_1, S_2) = D_\gamma(S_2, S_1), \tag{17}$$

$$D_\gamma(S_1, S_2) + D_\gamma(S_2, S_3) \geq D_\gamma(S_1, S_3). \tag{18}$$

Properties (14), (15), (16), and (17) may be easily verified from definitions (9), (10), and (11). Property (14) (nonnegativity) follows immediately from the fact that each $D_\gamma(S_1, S_2; i, j)$ is the probability that a random variable with density f takes a value in a subinterval (a_{ij}, b_{ij}) of $(-1, 1)$. The proof of Property (15) is trivial, as each of the subintervals (a_{ij}, b_{ij}) are now of length 0.

To prove Property (16), we note that $\gamma \leq 1 \Rightarrow (\gamma a_{ij}, \gamma b_{ij}) \subseteq (-1, 1)\ \forall i < j$. Since $f(x) > 0$ if $x \in (-1, 1)$, we have

$$D_\gamma(S_1, S_2; i, j) = 0 \Leftrightarrow a_{ij} = b_{ij}.$$

So, $D_\gamma(S_1, S_2) = 0$ implies that

$$s_{1i} - s_{1j} = s_{2i} - s_{2j}\ \forall\ i < j \in \Omega.$$

Note that $s_{11} = 1$. Also,

$$\begin{aligned} s_{21} - s_{2n} &= \sum_{i=1}^{n-1} s_{2i} - s_{2,i+1} \\ &= \sum_{i=1}^{n-1} s_{1i} - s_{1,i+1} \\ &= s_{11} - s_{1n} \\ &= 1. \end{aligned}$$

Thus, $s_{21} - s_{2n} = 1$, and so the normalization constraint implies that $s_{21} = 1$ and $s_{2n} = 0$. Setting $j = i + 1$ and varying i from 1 to $n - 1$, we observe that $s_{1j} = s_{2j}\ \forall\, 2 \leq j \leq n$, and hence, $S_1 = S_2$. Thus, Property (16) is proved. This property need not hold for $\gamma > 1$ because the integration interval may have no intersection with $(-1, 1)$, in which case, f would be 0 throughout the interval.

From the definition in (8), and the symmetry of $\max$ and $\min$,

$$a_{ij}^{S_1,S_2} = a_{ij}^{S_2,S_1}$$

and

$$b_{ij}^{S_1,S_2} = b_{ij}^{S_2,S_1},\ \forall\ i < j \in \Omega,$$

and hence,

$$D_\gamma(S_1, S_2; i, j) = D_\gamma(S_2, S_1; i, j),$$

thereby confirming Property (17).

To prove Property (18), we make use of the following facts about $\max$ and $\min$:

$$\min\{a, b\} \leq \max\{b, c\}, \tag{19}$$

$$\min\{a, c\} \geq \min\{\min\{a, b\}, \min\{b, c\}\}, \tag{20}$$

and

$$\max\{a, c\} \leq \max\{\max\{a, b\}, \max\{b, c\}\}. \tag{21}$$

Fact (19) implies that

$$a_{ij}^{S_1,S_2} \leq b_{ij}^{S_2,S_3},$$

and

$$a_{ij}^{S_2,S_3} \leq b_{ij}^{S_1,S_2}, \tag{22}$$

and hence, the following inequalities hold:

$$\begin{aligned} &\left(a_{ij}^{S_1,S_2}, b_{ij}^{S_1,S_2}\right) \cup \left(a_{ij}^{S_2,S_3}, b_{ij}^{S_2,S_3}\right) \\ &= \left(\min\left\{a_{ij}^{S_1,S_2}, a_{ij}^{S_2,S_3}\right\}, \max\left\{b_{ij}^{S_1,S_2}, b_{ij}^{S_2,S_3}\right\}\right) \\ &\supseteq \left(a_{ij}^{S_1,S_3}, b_{ij}^{S_1,S_3}\right). \end{aligned} \tag{23}$$

Here, the first equality is a consequence of (22), which ensures that the union of the given intervals is indeed an interval. The second inequality is a consequence of (20) and (21). It may be noted that the same inequalities hold even when the a_{ij}'s and b_{ij}'s are multiplied by a positive constant γ.

Integrating the nonnegative function f over the above intervals (scaled by the constant γ), we have the following set of inequalities:

$$\begin{aligned} &\int_{\gamma a_{ij}^{S_1,S_2}}^{\gamma b_{ij}^{S_1,S_2}} f(x)dx + \int_{\gamma a_{ij}^{S_2,S_3}}^{\gamma b_{ij}^{S_2,S_3}} f(x)dx \\ &\geq \int_{\gamma \min\left\{a_{ij}^{S_1,S_2},\, a_{ij}^{S_2,S_3}\right\}}^{\gamma \max\left\{b_{ij}^{S_1,S_2},\, b_{ij}^{S_2,S_3}\right\}} f(x)dx \\ &\geq \int_{\gamma a_{ij}^{S_1,S_3}}^{\gamma b_{ij}^{S_1,S_3}} f(x)dx, \end{aligned} \tag{24}$$

which is the same as the triangular inequality

$$D_\gamma(S_1, S_2; i, j) + D_\gamma(S_2, S_3; i, j) \geq D_\gamma(S_1, S_3; i, j).$$

Summing over all pairs $(i < j)$, we have Property (18).

Thus, $D_\gamma(S_1, S_2)$ is a metric if $\gamma \leq 1$, and a pseudometric otherwise. □

Kendall distance corresponds to a special case of the proposed metric by choosing f to have equal mass on $(-1, 0)$ and (0, 1) (a weaker condition than symmetry), and β to be very close to zero, or equivalently, γ very large. Intuitively, this means that there is no fusion, and S_1 and S_2 are being compared directly to each other. It may be observed that, in such a case, the interval $(\gamma a_{ij}, \gamma b_{ij})$ either contains the whole of $(-1,1)$ (when $a_{ij} < 0$) or does not have any intersection with $(-1,1)$ (when $a_{ij} > 0$), which is the support of f, and thus, the degree of discordance is either 1 or 0, respectively. When there is a tie (and it is not a double tie), one limit of the integral in (10) becomes zero, and the other limit is either larger than 1 or smaller than -1, and hence, the degree of discordance is $\frac{1}{2}$. For a double tie, the degree of discordance is 0 for any γ (indicating perfect concordance). Thus, when γ is very large, $D_\gamma(S_1, S_2; i, j)$ assumes the value of 1 whenever (i, j) is a discordant pair with respect to S_1 and S_2, 0 when it is a concordant pair, $\frac{1}{2}$ in case of a single tie, and 0 for a double tie, and hence, $D_\gamma(S_1, S_2)$ is the Kendall distance between S_1 and S_2. Formally, this may be defined as

$$K(S_1, S_2; i, j) = \lim_{\gamma \to \infty} D_\gamma(S_1, S_2; i, j) \; \forall \; i < j. \tag{25}$$

For each pair, $\{i, j\}$, as γ increases, the value of $D_\gamma(S_1, S_2; i, j)$ monotonically decreases to 0 if $\{i, j\}$ is concordant and increases to 1 otherwise. However, this does not imply that $D_\gamma(S_1, S_2)$ either monotonically increases or decreases with γ, the reason being that some of the individual components may increase while others decrease, and the rates may not balance each other. It may be easily verified in the case of $n = 3$ that $D_\gamma(S_1, S_2)$ may first increase and then decrease with γ.

$D_\gamma(S_1, S_2)$ is bounded above by $\frac{n(n-1)}{2}$. This is obvious because $D_\gamma(S_1, S_2; i, j)$ is bounded above by 1 for each pair $\{i, j\}$. Also, $D_\gamma(S_1, S_2)$ and $K(S_1, S_2)$ do not dominate each other. For example, when $K(S_1, S_2) = \frac{n(n-1)}{2}$, $D_\gamma(S_1, S_2)$ may be smaller (say, when $\gamma = 1$), and $D_\gamma(S_1, S_2)$ may be positive when $K(S_1, S_2)$ is zero (when $S_1 \neq S_2$ but $R(S_1) = R(S_2)$). Thus, $D_\gamma(S_1, S_2) - K(S_1, S_2)$ may be positive for certain choices of S_1 and S_2 and negative for some others.

Though the Kendall distance may be computed naïvely in $O(n^2)$ time, Knight's algorithm [14] based on mergesort achieves the same in $O(n \log n)$ time by taking advantage of the redundancy involved in computing $K(S_1, S_2)$. No such algorithm is known, as yet, for computing $D_\gamma(S_1, S_2)$. Preliminary ongoing research in this direction is promising and suggests the existence of linear time algorithms to compute a "reasonable" approximation to $D_\gamma(S_1, S_2)$. Alternatively, when the number of items, n, is large, one may resort to a method like the one suggested by Fagin et al. [15], where only the k top ranked items of both lists are considered for comparison. A "top k" version of the proposed metric is presented in Section 4.

4 Comparing Top k Scores

A top k list is the set of items with the largest scores. Top k lists differ from full lists because two lists need not have the same set of items. If C is the set of items common to both lists, then there are a total of $2k - |C|$ items in the two lists combined together, and thus, there is a total of $\frac{(2k-|C|)(2k-|C|-1)}{2}$ pairs. To compute the degree of discordance of a pair $\{i, j\}$, the four score values, namely, s_{1i}, s_{1j}, s_{2i}, and s_{2j} need to be known. However, all four scores are known for only $\frac{|C|(|C|-1)}{2}$ pairs, and for the remaining pairs, either one or two of the scores are unknown, and hence, some sort of estimation needs to be performed for determining their degree of discordance.

We extend our procedure to comparing the top k scores of two scoring functions by mimicking the work of Fagin et al. [15]. Fagin et al. compared the top k lists obtained by two different rankings [15]. When dealing with items that appear in only one list, the definitions of ranks and discordance are appropriately modified, resulting in, among many others, a Kendall distance for top k lists.

4.1 Comparing Top k Lists

We first study the approach of Fagin et al. [15] for generalizing the definition of discordance to the case of top k lists. We reproduce the text from [15] and, simultaneously, make a note of how the same extension for computing the degree of discordance would differ in each case. Let τ_1 and τ_2 be two top k lists. The generalized discordance between i and j, with respect to two lists τ_1 and τ_2, is denoted as $K^{(p)}(\tau_1, \tau_2; i, j)$.

- Case 1 (i and j appear in both top k lists). If i and j are in the same order (such as i being ahead of j in both top k lists), then let $K^{(p)}(\tau_1, \tau_2; i, j) = 0$; this corresponds to "no penalty" for $\{i, j\}$. If i and j are in the opposite order (such as i being ahead of j in τ_1 and j being ahead of i in τ_2), then let the penalty $K^{(p)}(\tau_1, \tau_2; i, j) = 1$.

In this case, the usual definitions of discordance and degree of discordance are applicable.

- Case 2 (i and j both appear in one top k list (say τ_1), and exactly one of i or j, say i, appears in the other top k list (τ_2)). If i is ahead of j in τ_1, then let the penalty $K^{(p)}(\tau_1, \tau_2; i, j) = 0$; otherwise, let $K^{(p)}(\tau_1, \tau_2; i, j) = 1$. Intuitively, we know that i is ahead of j as far as τ_2 is concerned, since i appears in τ_2 but j does not.

Here, there is no confusion regarding what the discordance should be as it is clear that i is ahead of j in τ_2. However, the degree of discordance needs the information regarding the separation between i and j. If i appears higher in τ_2, then j is far below i as compared to when i is toward the bottom of τ_2. Also, since Fagin et al. [15] consider only lists of items, there are no ties, whereas in our case, the scores may be tied.

- Case 3 (i, but not j, appears in one top k list (say τ_1), and j, but not i, appears in the other top k list (τ_2)). Then, let the penalty $K^{(p)}(\tau_1, \tau_2; i, j) = 1$. Intuitively, we know that i is ahead of j as far as τ_1 is concerned and j is ahead of i as far as τ_2 is concerned.

Again, though one is sure of discordance in this case, the degree of discordance may be partially inferred from the positions of i and j in their respective lists.

- Case 4 (i and j both appear in one top k list (say τ_1), but neither i nor j appears in the other top k list (τ_2)). This is the interesting case (the only case where there is really an option as to what the penalty should be). Such pairs $\{i, j\}$ are called special pairs. In this case, we let the penalty $K^{(p)}(\tau_1, \tau_2; i, j) = p$.

This is the most difficult case, since the order of i and j in τ_2 is not known. However, the positions of i and j in τ_1 carries some information regarding what the degree of discordance may now be.

4.2 Degree of Discordance for Top k Scores

We shall now extend the definition of the degree of discordance to the case of the top k scores by making the maximum use of the available information and averaging out the unknown part. To compute the average over the unknown score values, we assume (as in Section 3.2) that they are uniformly distributed and independent of each other, and take the expectation. We assume that the top k scores of two scoring functions S_1 and S_2, say S_1^k and S_2^k, are given, along with the corresponding lists of items, τ_1 and τ_2. By Assumption 1, $\tau_1 = \{1, 2, \ldots, k\}$. Let $D_\gamma^k(S_1^k, S_2^k; i, j)$ denote the degree of discordance between i and j with respect to S_1^k and S_2^k (though not mentioned explicitly, $D_\gamma^k(S_1^k, S_2^k; i, j)$ involves τ_1 and τ_2 also).

- Case 1 ($i, j \in \tau_1 \cap \tau_2$). Since, s_{1i}, s_{1j}, s_{2i}, s_{2j} are all known, the earlier definition is applied straightaway and $D_\gamma^k(S_1^k, S_2^k; i, j) = D_\gamma(S_1, S_2; i, j)$.
- Case 2 ($i, j \in \tau_1$, but $i \in \tau_2$ and $j \notin \tau_2$). So, s_{1i}, s_{1j}, s_{2i} are known but s_{2j} is unknown. All that is known about $y = s_{2j}$ is that $0 \leq s_{2j} \leq s_{2k} \leq s_{2i}$. Let a denote $s_{1i} - s_{1j}$. Therefore, we have

$$D_\gamma(S_1, S_2; i, j) = \int_{\gamma \min\{a, s_{2i} - y\}}^{\gamma \max\{a, s_{2i} - y\}} f(x) dx.$$

- Since, y is unknown, we average the degree of discordance over all possible values of y by taking the expectation as follows:

$$\begin{aligned} D_\gamma^k(S_1^k, S_2^k; i, j) &= E\big[D_\gamma(S_1, S_2; i, j)\big] \\ &= \frac{1}{s_{2k}} \int_0^{s_{2k}} \int_{\gamma \min\{a, s_{2i} - y\}}^{\gamma \max\{a, s_{2i} - y\}} f(x) dx dy. \end{aligned} \tag{26}$$

- It may be noted that, as it is already given that $0 \leq y \leq s_{2k}$, the above is a conditional expectation, where y is assumed to be from $U(0, s_{2k})$ distribution. Also, (26) corresponds to the definition of $K^{(p)}(\tau_1, \tau_2; i, j)$ in Case 2 in Section 4.1. This may be seen by noting that when γ is large, the integral in (26) is 0, $\frac{1}{2}$, or 1 accordingly as $a > 0$, $a = 0$, or $a < 0$.
- Case 3 ($i \in \tau_1$, $j \in \tau_2$, and $i \notin \tau_2$, $j \notin \tau_1$). Letting $y = s_{1j}$ and $z = s_{2i}$, and noting that $z - s_{2j} \leq 0 \leq s_{1i} - y$, the expected value of the degree of discordance may once again be computed as

$$\begin{aligned} D_\gamma^k(S_1^k, S_2^k; i, j) &= E\big[D_\gamma(S_1, S_2; i, j)\big[\\ &= \frac{1}{s_{1k} s_{2k}} \int_0^{s_{1k}} \int_0^{s_{2k}} \int_{\gamma(z - s_{2j})}^{\gamma(s_{1i} - y)} f(x) dx\, dz\, dy. \end{aligned} \tag{27}$$

- Case 4 ($i, j \in \tau_1, i, j \notin \tau_2$). Let us denote $s_{1i} - s_{1j}$, s_{2i}, and s_{2j} by a, y, and z, respectively, and without loss of generality, assume that $a \geq 0$. Thereby, the expected degree of discordance is given by

$$\begin{aligned} D_\gamma^k(S_1^k, S_2^k; i, j) &= E\big[D_\gamma(S_1, S_2; i, j)\big[\\ &= \frac{1}{s_{2k}^2} \int_0^{s_{2k}} \int_0^{s_{2k}} \int_{\min\{\gamma a, \gamma(y - z)\}}^{\max\{\gamma a, \gamma(y - z)\}} f(x) dx\, dz\, dy. \end{aligned} \tag{28}$$

5 Applications

Scores contain more information than ranks, especially because the ranks may themselves be derived from the scores. So, comparing scorings finds applications in any field where rankings need to be compared. We describe two such application areas related to page ranking. In addition, we elaborate on how the scores may also be used to measure how representative the ranks are.

5.1 Comparing Web Page Rankings

Ranking web pages has attracted the attention of several researchers, mainly due to the challenges it poses in terms of scalability and the imprecise and subjective nature of the task. Given the wide variety of ranking methods available, it is natural to compare them to decide which one is better. A more fundamental task is to decide whether the two given rankings are indeed different, and if so, how well separated they are.

Though the task is to *rank* Web documents, page ranking algorithms assign scores to pages. These scores are called page ranks. Existing works [16] compare the rankings by converting the scores into ranks and then computing the distance between these ranks. As discussed earlier in this paper, and also, as is evident from the literature, these scores are seldom used in isolation for producing the final rankings. In such a case, the proposed methodology is more appropriate for comparing the page ranks, as it takes maximum advantage of the available information regarding the intended use of the page ranks. For example, if the fusing proportions (and thereby, γ) are known beforehand, the distance $D_\gamma(S_1, S_2)$ may be computed appropriately. On the other hand, if the algorithms produce the final ranking without fusing the scores, then, γ may be set to a very high value, which results in the computation of the Kendall distance.

It is common in the case of ranking web pages that the number of items is very large, and under such circumstances, $D_\gamma^k(S_1^k, S_2^k)$ should be used to compare S_1 and S_2 in terms of their top k scores, with k set to a few hundred or thousand.

5.2 Stopping Criterion for the Power Method

Another application of the proposed metric is in deciding when to stop the iterations in the *power method* [17]. The power method is used to obtain the dominant (or principal) eigenvector of matrix A, starting with an arbitrary vector $\mathrm{x}^{(0)}$. It is an iterative procedure, whereby successive vectors $\mathrm{x}^{(i+1)}$

are produced by multiplying A with $\mathbf{x}^{(i)}$, and is guaranteed to converge as the number of iterations tends to infinity.

In several studies like [18] and [19], the page rank vectors correspond to the principal eigenvectors of some transition probability matrix and are computed iteratively by the power method. Once again, as in Section 5.1, the size of the document collection under consideration may be huge, in which case, each iteration is very costly, sometimes taking several hours to a few days [20], and hence, early stopping is desirable. The page rank vectors produced by consecutive iterations are compared to see if near convergence is attained.

In some instances, the computation is performed for a fixed number of iterations, say 50 or 100 iterations [18]. Though convergence may not be attained (in the L_1 or L_2 sense) by the time the computation is stopped, the resultant vector is declared to be the final page rank vector. The justification provided for such a behavior is that this vector serves its purpose in terms of ranking the documents, which is the final objective. In other words, even if the iterations are allowed to run for longer, the ranking would not change by much, as determined by the Kendall distance. Alternatively, one may base the stopping criterion in terms of the L_1 distance between the consecutive page rank vectors. However, since the ultimate objective is to rank the pages, it is preferable to use a rank comparison method for determining the stopping time [18].

While such a justification is acceptable if ranking is the sole objective, there might also be other objectives. In most cases, the page rank vector is considered as a set of importance scores and is combined with other entities, such as relevance scores, before the final ranking is produced [21]. Thus, it is natural to ask if it is sufficient to stop the computation after a certain number of iterations. The proposed distance measure may be used to check if (near) convergence has been attained. This convergence would be in a sense that considers the purpose of computing the eigenvector.

5.3 Quantitative Measurement of the Representation of Scores by Ranks

A system ranking items on the basis of scores assigned to them may choose to reveal only the ranks of those items, usually, by returning them in a particular order. While it is true that the items would be ordered in exactly the same manner on the basis of their scores too, the scores are only partially revealed. For example, let there be four items (1, 2, 3, 4), each of which is assigned a score s_i, $i = 1, 2, 3, 4$, and the items are ordered in descending order of scores. If it is known that the ordered list is 1, 2, 3, 4, then all that is revealed about the scores is that $1 = s_1 \geq s_2 \geq s_3 \geq s_4 = 0$. However, there is a general (human) tendency to perceive that the scores are uniformly distributed. So, the scores are implicitly assumed to be 1, $\frac{2}{3}$, $\frac{1}{3}$, and 0, respectively (or something similar). This also corresponds to the average case, where one may observe that though s_3 may be either less than or greater than $\frac{1}{3}$, it is expected (under the assumption of uniformity) to be close to $\frac{1}{3}$. Such uniform scores, implicitly assumed on the basis of the ranks, shall be called the *uniformly perceived scores*, and the uniformly perceived score vector shall be denoted by $R(S)$ (or simply by R if S is clear from the context). For the sake of notational simplicity, we shall sometimes refer to uniformly perceived scores as just perceived scores in the rest of this paper.

In reality, the underlying scores may not be reflected properly by the rankings available. For instance, in the above example, the actual scores could have been (1, 0, 0, 0), or (1, 1, 1, 0) or (1, 0.5, 0.5, 0), or any such 4-tuple satisfying the ordering criterion. So, the perceived scores may or may not reflect the underlying scores. Thereby, one should be able to measure if at all the perceived scores are similar to the actual scores.

The present work provides a methodology to quantify the separation between the actual scores and the perceived scores. Let there be n items $1, 2, \ldots, n$, and let their perceived and actual scores be denoted by $r_1, r_2, \ldots, r_n$ and $s_1, s_2, \ldots, s_n$, respectively. So, the perceived score of the ith item is given by $r_i = 1 - \frac{i-1}{n-1} = \frac{n-i}{n-1}$. It may be noted that Kendall distance between the scores and the perceived scores is zero, which is due to the fact that both of them result in the same rankings.

The degree of discordance of a pair $(i < j)$ is thereby given by

$$D_\gamma(R, S; i, j) = \int_{\min\left(s_i - s_j, \frac{j-i}{n-1}\right)}^{\max\left(s_i - s_j, \frac{j-i}{n-1}\right)} (1 - x) dx,$$

since it is already known that $s_i \geq s_j$ and $r_j - r_i = \frac{j-i}{n-1} > 0$. If $s_i - s_j = \frac{j-i}{n-1}$, then the degree of discordance is zero. The other extreme is the case when the degree of discordance for the pair (i, j) is maximum. This happens, when $s_i - s_j$ is either 0 or 1, which of the two is determined by $j - i$. If $j - i$ is close to $n - 1$, then $D_\gamma(R, S; i, j)$ is maximized at $s_j = s_i$, whereas if $j - i$ is near 1, $s_i = 1$ and $s_j = 0$ maximizes the value of $D_\gamma(R, S; i, j)$.

As earlier, we would also be concerned with the total score-based discordance between the two scorings. This is obtained by summing the degree of discordance over all possible pairs and is given by

$$D_\gamma(R, S) = \sum_{i=0}^{n-1} \sum_{j=i+1}^{n} D_\gamma(R, S; i, j).$$

Again, if $s_i = 1 - \frac{i-1}{n-1}$, for each $i = 1, 2, \ldots, n$, the s_i's coincide with the r_i's, and hence, $D_\gamma(R, S)$ turns out to be zero.

Since each $D_\gamma(R, S; i, j)$ is bounded above by $\frac{1}{2}$, it may be trivially seen that $D_\gamma(R, S) \leq \frac{n(n-1)}{4}$. However, finding R and S such that $D_\gamma(R, S)$ attains a maximum is not as simple as maximizing $D_\gamma(R, S; i, j)$ for each pair (i, j), the reason being that the pairs are not independent of each other, due to the monotonicity and normalization constraints. For example, if $n = 5$, $D_\gamma(R, S; 1, 2)$ is maximized when $s_2 = 0$, whereas $D_\gamma(R, S; 1, 3)$ is maximized when $s_3 = 1$; however, both cannot happen simultaneously (because $s_2 \geq s_3$).

How well R represents a particular score vector S_0 may be measured by the notion of the p-value of $D_\gamma(R, S_0)$. Consider the set, $\mathcal{S}_0$, of all S vectors such that $R(S) = R(S_0)$. The p-value is the probability of having an S vector ($S \in \mathcal{S}_0$) such that $D_\gamma(R, S) \geq D_\gamma(R, S_0)$. In other words, the p-value of $D_\gamma(R, S)$ is the proportion of S vectors in $\mathcal{S}_0$, which are at an equal or higher distance from R than S_0 is from R. So, when most of the S vectors are such that $D_\gamma(R, S) \geq D_\gamma(R, S_0)$, then $D_\gamma(R, S_0)$ may be considered to be small, and vice versa. Thus, when the p-value corresponding to $D_\gamma(R, S)$ is very small, it may be declared that $R(S)$ does not represent S well enough.

6 Experimental Results

To understand the significance of the present work and to validate the claims made in this paper, several experiments of the following kinds were conducted:

- Study of the behavior of $D_\gamma(S_1, S_2)$ for various values of γ.
- Study of the behavior of $D_\gamma^k(S_1, S_2)$ for various values of n and k, and testing the dependence on the assumption of uniformity.
- Determining the number of iterations for eigenvector computation.
- Predicting the discordance in a pair of vectors after score fusion.
- Computing the distance between uniformly perceived and actual scores.

We now describe each of these experiments and their results along with our observations and analysis.

6.1 Behavior of $D_\gamma(S_1, S_2)$

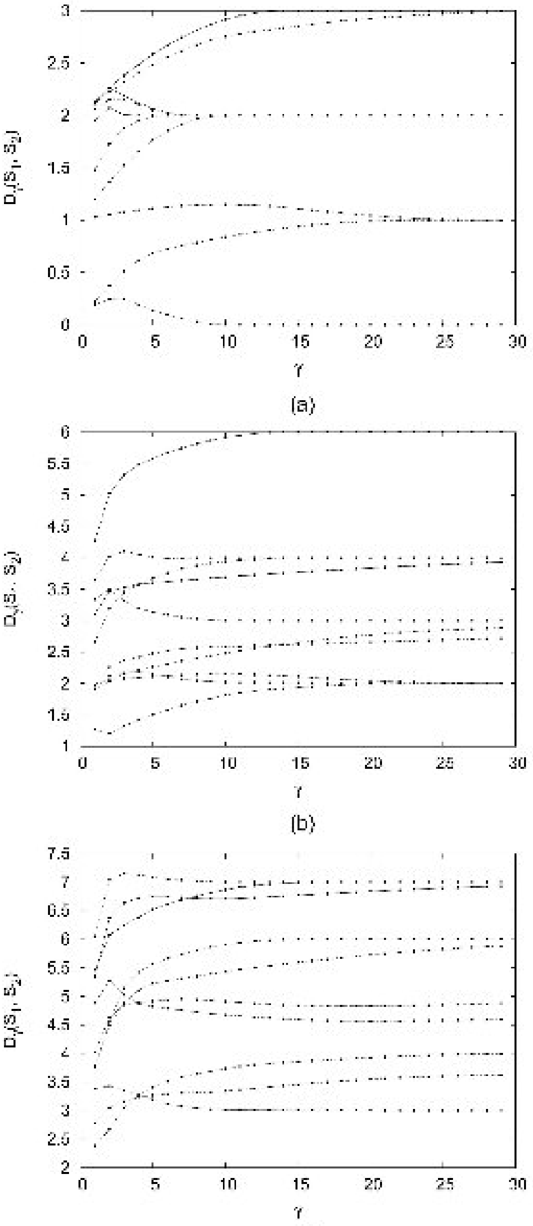
It is theoretically assured that $D_\gamma(S_1, S_2) \leq \frac{n(n-1)}{2}$ and also that $\lim_{\gamma\to\infty} D_\gamma(S_1, S_2) = K(S_1, S_2)$. Moreover, our remarks in Section 3.3 make it amply clear that $D_\gamma(S_1, S_2)$ is not a monotone function of γ although each $D_\gamma(S_1, S_2; i, j)$ is so.

The aforementioned properties are graphically depicted in Figs. 2a, 2b, and 2c. Here, n is set to 3, 4, and 5, respectively, and 10 S_1, S_2 pairs are randomly generated, and $D_\gamma(S_1, S_2)$ is computed for values of γ varying from 1 to 30. The Kendall distance, $K(S_1, S_2)$, may take values only from $\{0, 1, \ldots, \frac{n(n-1)}{2}\}$, and the $D_\gamma(S_1, S_2)$ values are seen to be converging to the respective Kendall distances. The rates of convergence, however, are different in each case. Also, in some of the cases, $D_\gamma(S_1, S_2)$ varies monotonely with γ, whereas in the remaining cases, it is not so. However, eventually (beyond some value of γ), monotonicity is restored in each of the cases.

6.2 Behavior of $D_\gamma^k(S_1, S_2)$

We now study the properties of $D_\gamma^k(S_1, S_2)$ as k varies from 1 to n. A pair of score vectors is generated randomly and min-max normalization [22] is applied. The top k items (according to the scores) are selected from each vector, while the scores of the remaining items are assumed to be unknown, and $D_\gamma^k(S_1, S_2)$ is computed as described in Section 4.2. Since $D_\gamma^k(S_1, S_2)$ is an expected value, with the expectation being taken over all the unknown score values, it is imperative to know how good this approximation is. In the present case, all the score values are known, and therefore, the exact value $E_\gamma^k(S_1, S_2)$ may also be computed by summing up the (exact) degree of discordance of all the pairs appearing in the union of the two top k lists.

The computed values of $D_\gamma^k(S_1, S_2)$, $E_\gamma^k(S_1, S_2)$, and $D_\gamma^k(S_1, S_2) - E_\gamma^k(S_1, S_2)$ are shown graphically in Fig. 3, from which it may be seen that the $D_\gamma^k(S_1, S_2)$ approximates $E_\gamma^k(S_1, S_2)$ very well. It may be noted that the computation of $D_\gamma^k(S_1, S_2)$ is based on the assumption that S_1 and S_2 arise from the uniform $(U(0,1))$ distribution. In order to test the dependence of the approximation on the distributional assumptions, two more experiments were conducted, with S_1 and S_2 arising from the Gaussian $(N(0,1))$ distribution in one and the exponential $(E(1))$ distribution in the other. The results are presented in Figs. 4 and 5, respectively, and it may be observed that there is a consistent overestimation in the case of the Gaussian distribution, whereas the approximation is very good in the case of the exponential distribution.



(a)

(b)

(c)

Fig. 2. Plots of $D_\gamma(S_1, S_2)$ versus γ, for 10 randomly chosen (S_1, S_2) pairs, with (a) $n = 3$, (b) $n = 4$, and (c) $n = 5$.

6.3 Determining the Number of Iterations for Computing Page Ranks

As discussed in Section 5.2, one would like to know when to terminate the power iterations based on the amount of change in the rankings in the consecutive iterations. We chose two data sets from Stanford's WebBase [23] and

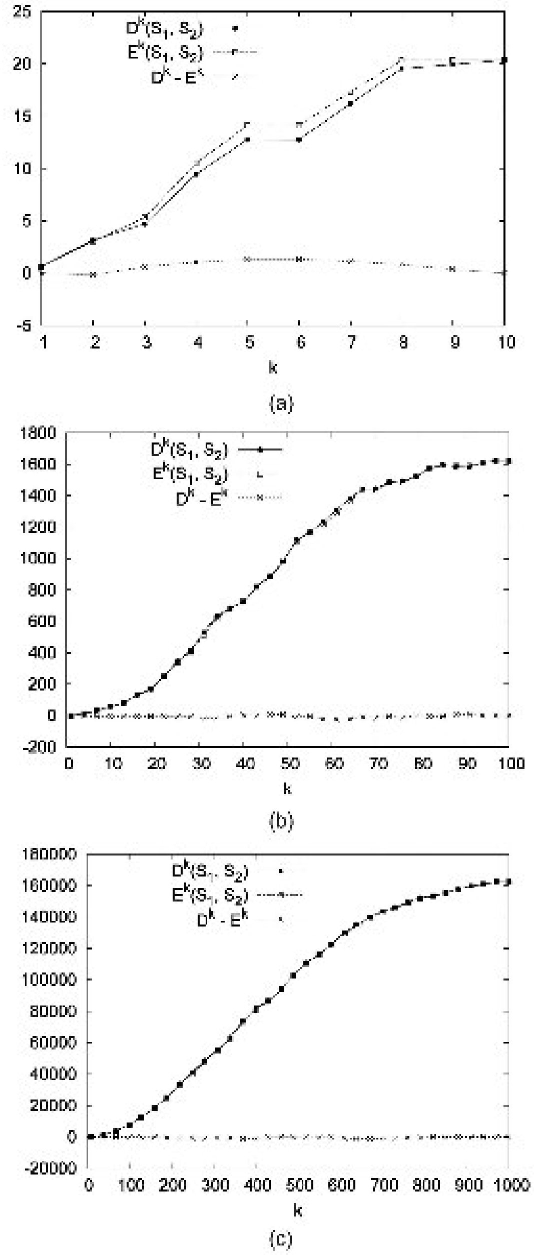


Fig. 3. Plots of $D^k(S_1, S_2)$, $E^k(S_1, S_2)$, and $D^k(S_1, S_2) - E^k(S_1, S_2)$ versus k, for (S_1, S_2) generated from uniform distribution, with (a) $n = 10$, (b) $n = 100$, and (c) $n = 1{,}000$.

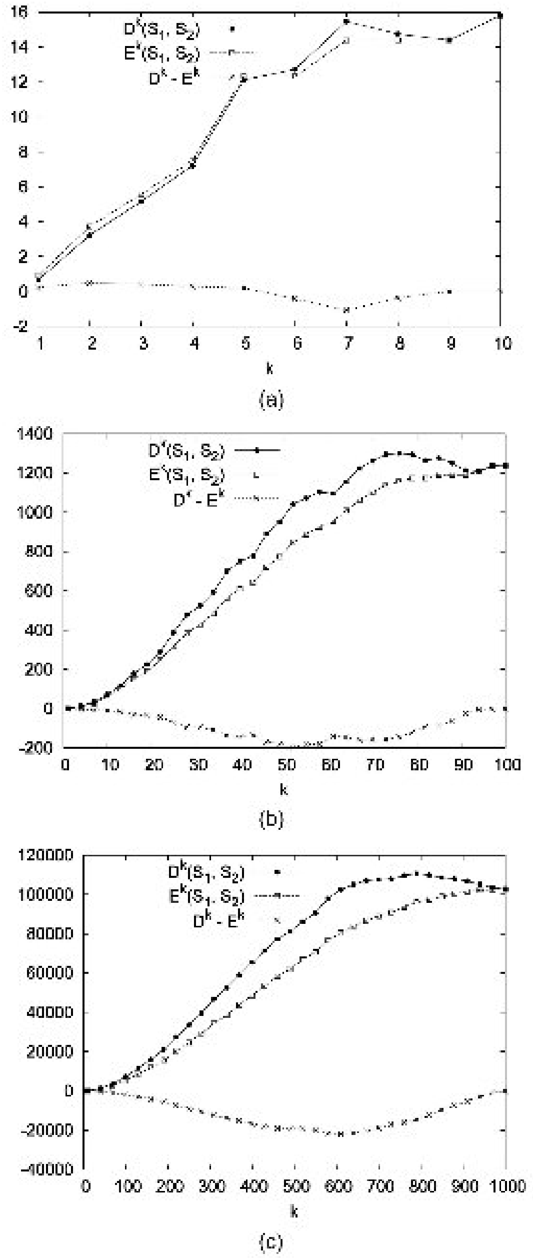


Fig. 4. Plots of $D^k(S_1, S_2)$, $E^k(S_1, S_2)$, and $D^k(S_1, S_2) - E^k(S_1, S_2)$ versus k, for (S_1, S_2) generated from Gaussian distribution, with (a) $n = 10$, (b) $n = 100$, and (c) $n = 1{,}000$.

named them WB1_7440 and WB4_7060, after the host and port numbers from which they are available. The former is a crawl of a part of the berkeley.edu domain, and there are about 140,000 (140K) pages with over 1.6 million (1.6M) links to pages within the same data set. WB4_7060, which is a crawl of a part of the stanford.edu domain, consists of about 40,000 (40K) pages and over 260,000 (260K) links to pages within itself. The PageRank [24] algorithm was run for 100 iterations on both the chosen data sets. We then computed $K^{(0.5)}(S_i, S_{i+1})$ and $D^k(S_i, S_{i+1})$ (γ set to 1), which are the top k versions of Kendall distance and the proposed distance, respectively. Here, S_i is the page rank vector at the end of the ith iteration, and k was chosen to be 100, 1,000, and 5,000. We have also computed $K^{(p)}(S_i, S_{100})$ and $D^k(S_i, S_{100})$, though these quantities would not be available *during* the page rank computation. These values are presented in the plots in Figs. 6 and 7.

It may be noted from Figs. 6 and 7 that if $D^k(S_i, S_{i+1})$ is to be used instead of $K^{(0.5)}(S_i, S_{i+1})$, (near) convergence is declared much earlier. For example, in Fig. 6a, $D^k(S_i, S_{i+1})$ would have recommended the termination of the procedure after 20 iterations, whereas $K^{(0.5)}(S_i, S_{i+1})$ would have led to at least 26 iterations. This indicates that once it is decided that the obtained ranks would be fused together with some other score vector (in equal proportions, since we have set γ to be 1), there would be no significant improvement by continuing beyond 20 iterations.

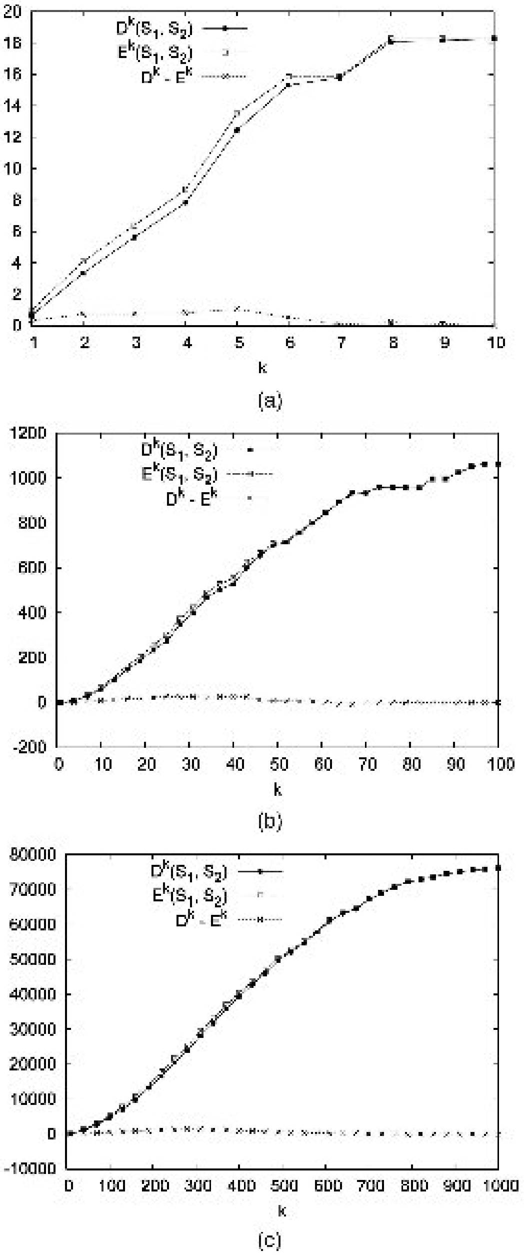


Fig. 5. Plots of $D^k(S_1, S_2)$, $E^k(S_1, S_2)$, and $D^k(S_1, S_2) - E^k(S_1, S_2)$ versus k, for (S_1, S_2) generated from exponential distribution, with (a) $n = 10$, (b) $n = 100$, and (c) $n = 1{,}000$.

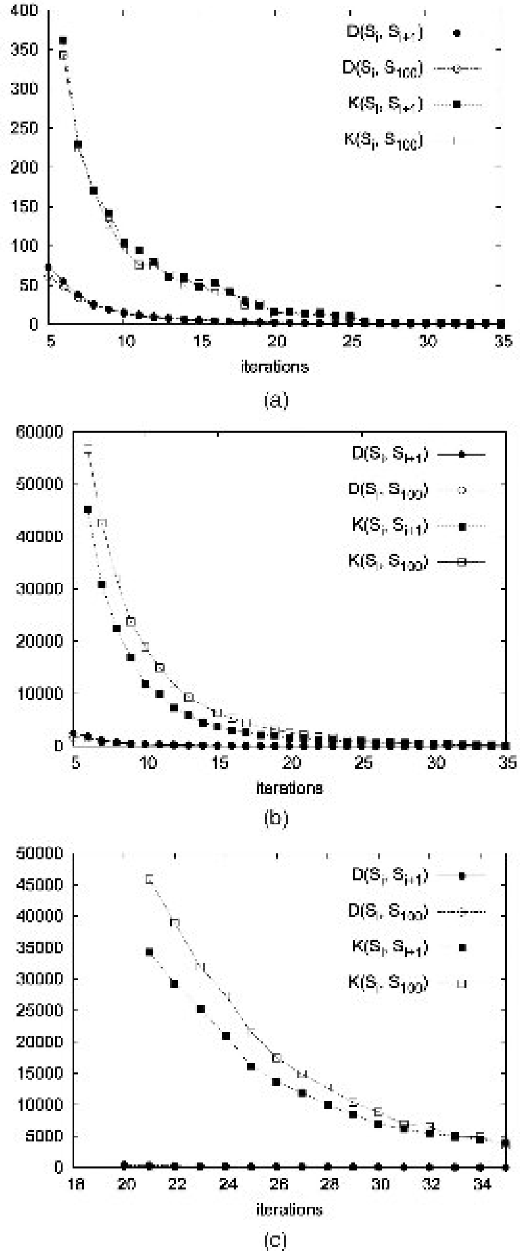


Fig. 6. Plots of $D^k(S_i, S_{i+1})$, $D^k(S_i, S_{100})$, $K^{(0.5)}(S_i, S_{i+1})$, and $K^{(0.5)}(S_i, S_{100})$ versus i, for the WB1_7440 data set with (a) $k = 100$, (b) $k = 1{,}000$, and (c) $k = 5{,}000$.

6.4 Predicting Discordance after Score Fusion

This set of experiments is aimed at predicting the discordance after score fusion, without actually performing the fusion. Such requirements arise often in the Web domain, say, for comparing page rank vectors. The traditional method of comparing page rank vectors is to use each of them to retrieve the top pages for a set of queries and computing the discordance between them (see, for example, [21]). This involves doing the following for each query. The set of documents matching (or containing) the query is identified, and the query relevance scores for each document are computed. These relevance scores are combined with the ranks of the documents, and the documents are ordered according to the fused scores. A rank comparison measure is then computed between the top k lists arising from each of the page rank vectors fused with the relevance vectors.

Ideally, the set of queries should be very large so that a comprehensive comparison between the page rank vectors may be made. In addition, the number of documents in the corpus may also be huge. Under such circumstances, it is computationally prohibitive to compare the page rank vectors for various choices of the weights (or fusing parameters). To this end, the proposed measure may be employed to circumvent the actual computation of discordance measures between the fused score vectors by having a reasonable approximation as demonstrated by the following experiment.

We consider the WB1_7440 and WB4_7060 corpora once more. Two page rank vectors, labeled S_1 and S_2, over these

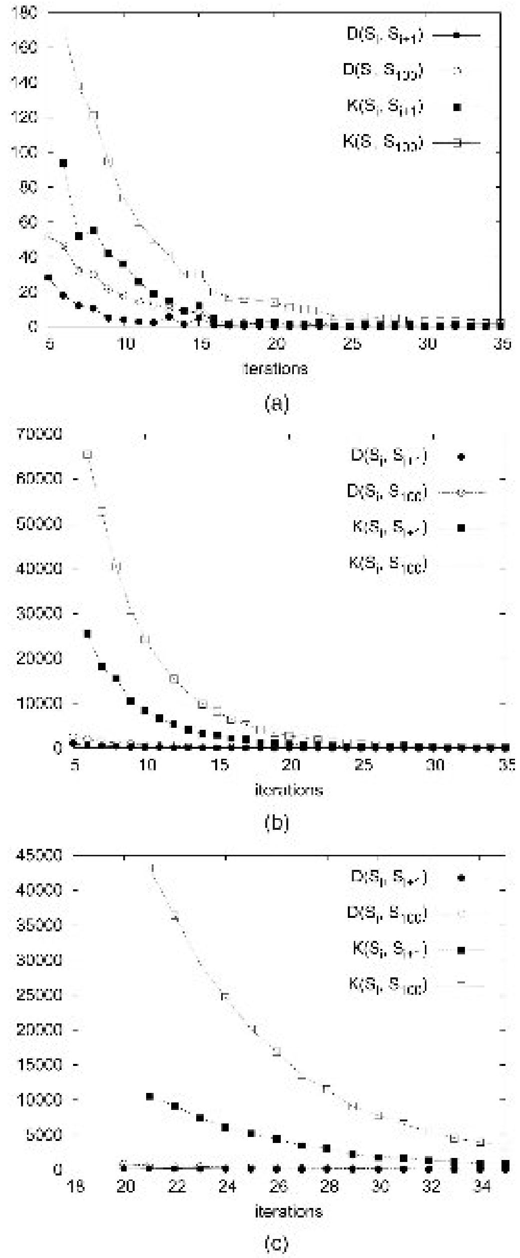


Fig. 7. Plots of $D^k(S_i, S_{i+1})$, $D^k(S_i, S_{100})$, $K^{(0.5)}(S_i, S_{i+1})$, and $K^{(0.5)}(S_i, S_{100})$ versus i, for the WB4_7060 data set with (a) $k = 100$, (b) $k = 1{,}000$, and (c) $k = 5{,}000$.

data sets are obtained by considering the vectors after 3 and 50 iterations, respectively, of the eigenvector computation mentioned earlier. Note that the exact method of obtaining these vectors is not relevant to the present experiment. The fusing proportion $\beta\ (= 1 - \alpha)$ is varied over the values 0.25, 0.50, and 0.75.

We first compare the two vectors in the following naive manner. We choose each word in the corpus as a single-term query and stem them using Porter's algorithm [25]. An inverted index is created for each data set. Now, for each stem word w in the corpus, the list of documents containing that stem is extracted, and the TFIDF [26] vector, T_w, is computed. Let the corresponding page rank vectors, of the same length as T_w, be called S_{1w} and S_{2w}, respectively. There are over 44,000 (44K) and 35,000 (35K) distinct stems in WB1_7440 and WB4_7060 data sets, respectively. When only the 5,000 top ranked documents are considered in these data sets, there are over 14K and 16K distinct stems, respectively.

TABLE 1
Kendall and Degree of Discordance Values Averaged over All Words

WB1_7440				WB1_7060			
Pages chosen(k)	$K\tau^k_{0.5}$	γ	$D\tau^k_\gamma$	Pages chosen(k)	$K\tau^k_{0.5}$	γ	$D\tau^k_\gamma$
All (140K)	0.0160	3	0.0049	All (40K)	0.0222	3	0.0096
All (140K)	0.0160	1	0.0023	All (40K)	0.0222	1	0.0049
All (140K)	0.0160	$\frac{1}{3}$	0.0012	All (40K)	0.0222	$\frac{1}{3}$	0.0020
Top 5000	0.1220	3	0.0279	Top 5000	0.1130	3	0.0409
Top 5000	0.1220	1	0.0137	Top 5000	0.1130	1	0.0200
Top 5000	0.1220	$\frac{1}{3}$	0.0067	Top 5000	0.1130	$\frac{1}{3}$	0.0081

Now, for each word, we compute $K^{(0.5)}(S_{1w}, S_{2w})$ and $K^{(0.5)}(\alpha S_{1w} + \beta T_w, \alpha S_{2w} + \beta T_w)$, which are the top k discordance values between S_{1w} and S_{2w} and between the fused vectors $\alpha S_{1w} + \beta T_w$ and $\alpha S_{2w} + \beta T_w$, respectively. These quantities are averaged over all the words by dividing their sum by the sum of all possible pairs for each word, and the averages are denoted $K\tau^k_{0.5}(S_1, S_2)$ and $D\tau^k_\gamma(S_1, S_2)$ (where, as earlier, $\gamma = \frac{1-\beta}{\beta}$), respectively, reflecting the quantities they estimate. Mathematically, if df_w is the number of documents in which w appears, the average is computed as

$$K\tau^k_{0.5}(S_1, S_2) = \frac{2\sum_w K^{(0.5)}(S_{1w}, S_{2w})}{\sum_w df_w(df_w - 1)}, \tag{29}$$

and

$$D\tau^k_\gamma(S_1, S_2) = \frac{2\sum_w K^{(0.5)}(\alpha S_{1w} + \beta T_w, \alpha S_{2w} + \beta T_w)}{\sum_w df_w(df_w - 1)}. \tag{30}$$

The above procedure is repeated using just the top 5,000 documents in each data set. These averages, which would have been the measurements from the traditional rank comparison technique described earlier, are tabulated in Table 1.

Next, we employ the proposed metric to obtain an approximation of the values in Table 1. The discordance measures $K\tau^k_{0.5}(S_1, S_2)$ and $D\tau^k_\gamma(S_1, S_2)$ are shown in Table 2. Here, $K\tau^k_{0.5}(S_1, S_2)$ is the top k Kendall distance $K^{(0.5)}(S_1, S_2)$ normalized by all possible pairs of pages in the union. Similarly, $D\tau^k_\gamma(S_1, S_2)$ is the normalized version of $D^k_\gamma(S_1, S_2)$. These computations are repeated for the top 5,000 pages of each data set and are also tabulated in

TABLE 2
Normalized Kendall and Degree of Discordance Values

WB1_7440				WB1_7060			
Pages chosen(k)	$K\tau^k_{0.5}$	γ	$D\tau^k_\gamma$	Pages chosen(k)	$K\tau^k_{0.5}$	γ	$D\tau^k_\gamma$
All (140K)	0.0139	3	0.0032	All (40K)	0.0253	3	0.0067
All (140K)	0.0139	1	0.0017	All (40K)	0.0253	1	0.0029
All (140K)	0.0139	$\frac{1}{3}$	0.0006	All (40K)	0.0253	$\frac{1}{3}$	0.0011
Top 5000	0.1214	3	0.0133	Top 5000	0.1090	3	0.0323
Top 5000	0.1214	1	0.0078	Top 5000	0.1090	1	0.0110
Top 5000	0.1214	$\frac{1}{3}$	0.0045	Top 5000	0.1090	$\frac{1}{3}$	0.0054

TABLE 3
EFI for 20 States of India

Rank	State	EFI	Rank	State	EFI
1	Gujarat	0.40	11	Orissa	0.32
2	Andhra Pradesh	0.38	12	Karnataka	0.31
3	Kerala	0.37	13	Uttar Pradesh	0.30
4	Chhattisgarh	0.37	14	West Bengal	0.30
5	Tamil Nadu	0.37	15	Himachal Pradesh	0.30
6	Maharashtra	0.37	16	Jharkhand	0.29
7	Rajasthan	0.35	17	Punjab	0.29
8	Haryana	0.35	18	Uttaranchal	0.28
9	Madhya Pradesh	0.33	19	Bihar	0.26
10	Jammu & Kashmir	0.33	20	Assam	0.22

Table 2. Note that the values in Table 2 are independent of the content of the web pages, or to rephrase, the relevance of pages to queries is not considered during ranking.

We observe from Tables 1 and 2 that the corresponding values in Tables 1 and 2 are very similar and show similar patterns and trends. For example, on the WB1_7440 data set, when γ is set to 1, the discordance value as found from the naive experiment is 0.0023, and its estimate turns out to be 0.0017. When the possibility of fusion is not taken into consideration, the corresponding estimate is 0.0189, which is well away from 0.0023. The $D\tau_{\gamma}^{k}$ values in Table 2, however, are all underestimates of the corresponding values in Table 1, possibly due to the TFIDF scores not being uniformly distributed. Nevertheless, these values differ significantly from Kendall's τ values and shrink as β is increased. Moreover, it was also observed that, as β increases, the shrinking happens for each word (and not just on average).

In conclusion, the proposed metric helps us directly predict the discordance between the two page rank vectors instead of taking recourse to the traditional rank-based comparisons, which involve comparing fused vectors for a large number of queries that, in the present case, run into several thousands. In cases where more relevance factors are involved (for example, search engines like Yahoo! and Google weigh hundreds of factors to compute relevance [27]), for comparing two variants of a particular factor, all that is needed to be known is the fusion parameter (α) for the factor under consideration. That way, without knowing the weights for the remaining factors involved, or even what the exact factors are, one may estimate the amount of discordance that could result from differences in the two variants of this single factor.

6.5 Uniformly Perceived versus Actual Scores

This experiment is aimed at measuring the similarity between the uniformly perceived and actual scores by computing the distance $D(R, S)$. In other words, this is an attempt to quantify how much information is lost by converting the scores to ranks. For this purpose, the Economic Freedom Index (EFI) data set is chosen, in which 20 states of India were assigned a set of composite scores and were ranked accordingly [28]. The data set is presented in Table 3.

For the EFI data set, the distance between ranks and scores is 11.36. In order to be able to judge how good or how bad it is, we generated several score vectors (of size 20) uniformly and counted the number of times a score vector had a distance of 11.36 or more from the perceived score vector. The p-value turns out to be 0.12, which indicates that only about 12 percent of score vectors are more separated from the perceived score vectors. This, in turn, signifies that the scores in the EFI data set are not very well represented by the corresponding ranks (through the perceived scores). The authors in [28] had expressed their opinion that the ranks did not represent the scores well. The results of our experiments now provide a quantitative evidence for the same.

The individual degrees of discordance for each pair of states are displayed in Table 4. We have also computed the p-values for each of the individual cells. For each pair of states, we count the number of randomly generated score vectors with a higher degree of discordance for the same pair of states. All entries with the corresponding p-value less than 0.05 are shown in bold, and there are 31 such values, which is about 16 percent of the total of 190 entries. It may be noted that the 31 entries in bold are not the largest entries of Table 4. For example, the entry at position (3, 8) (0.12) is larger than that at (3, 19) (0.06), however, the former is more likely to occur than the latter.

TABLE 4
Degree of Discordance between the Actual and Uniformly Perceived Scores for Each Pair of States in the EFI Data Set

	State																			
State	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Total
1	0.05	0.05	0.01	0.04	0.03	0.03	0.06	0.02	0.05	0.04	0.04	0.03	0.05	0.06	**0.05**	**0.06**	**0.05**	**0.02**	0	0.79
2		0	0.05	0.09	0.13	0.08	0.11	0.06	0.09	0.08	0.07	0.07	**0.09**	**0.10**	**0.09**	**0.10**	**0.09**	**0.05**	0	1.42
3			0.05	0.10	**0.15**	0.08	0.12	0.07	0.10	0.09	0.08	0.07	**0.10**	**0.12**	**0.10**	**0.12**	**0.10**	**0.06**	0.01	1.60
4				0.05	0.10	0.04	0.08	0.03	0.07	0.06	0.05	0.05	0.07	**0.10**	**0.09**	**0.10**	**0.09**	**0.05**	0	1.15
5					0.05	0.01	0.04	0.01	0.03	0.03	0.02	0.02	0.05	0.07	0.07	**0.09**	**0.08**	0.04	0.01	0.88
6						0.05	0.01	0.05	0.01	0.01	0.01	0.01	0.02	0.05	0.04	0.07	0.06	0.03	0.02	0.94
7							0.05	0.01	0.04	0.04	0.03	0.03	0.06	0.09	0.08	**0.11**	**0.10**	0.06	0.01	0.99
8								0.05	0.01	0.01	0.01	0.01	0.03	0.06	0.05	0.08	0.07	0.04	0.03	0.93
9									0.05	0.05	0.04	0.04	0.08	0.11	0.10	**0.13**	**0.12**	0.07	0.01	1.11
10										0	0.01	0.01	0.04	0.08	0.07	0.10	0.09	0.05	0.04	0.93
11											0	0.01	0.04	0.08	0.08	0.11	0.10	0.05	0.04	0.93
12												0	0.05	0.09	0.08	0.12	0.11	0.06	0.04	0.94
13													0.05	0.10	0.09	0.13	0.12	0.07	0.05	0.95
14														0.05	0.05	0.09	0.08	0.03	0.08	1.09
15															0	0.05	0.04	0.01	0.12	1.39
16																0.05	0.05	0.01	0.12	1.28
17																	0	0.05	**0.17**	1.76
18																		0.05	**0.18**	1.59
19																			**0.15**	0.96
20																				1.08

7 Conclusions and Future Work

This paper has dealt with generalizing measures of discordance for the case when the underlying scores are known. A metric has been provided to compare score vectors directly. This metric turns out to be the Kendall distance when a parameter γ, denoting the ratio of fusing proportions, is large. Experiments of various kinds demonstrate the wide range of theory and applications of the metric introduced in the present work.

There is a tremendous scope for future work, including studying the cases where T is assumed to arise from specific distributions, obtaining the properties such as maximum and minimum of $D_\gamma(S_1, S_2)$ and $D^k_\gamma(S_1, S_2)$ for particular values of γ, and speeding up the computation of the proposed metric.

Acknowledgments

The first author's doctoral fellowship was funded by INSEAD, France. Narayan L. Bhamidipati was with the Machine Intelligence Unit, Indian Statistical Institute (ISI), Kolkata, India, when this work was conducted.

References

[1] F. Crestani, "Combination of Similarity Measures for Effective Spoken Document Retrieval," *J. Information Science,* vol. 29, no. 2, pp. 87-96, 2003.

[2] K.M. Donald and A.F. Smeaton, "A Comparison of Score, Rank and Probability-Based Fusion Methods for Video Shot Retrieval," *Proc. Int'l Conf. Image and Video Retrieval (CIVR '03),* pp. 61-70, 2003.

[3] R. Nuray and F. Can, "Automatic Ranking of Information Retrieval Systems Using Data Fusion," *Information Processing and Management,* vol. 42, no. 3, pp. 595-614, 2006.

[4] M.E. Renda and U. Straccia, "Web Metasearch: Rank versus Score Based Rank Aggregation Methods," *Proc. 18th Ann. ACM Symp. Applied Computing (SAC '03),* pp. 841-846, 2003.

[5] W.J. Conover, *Practical Nonparametric Statistics,* third ed. John Wiley & Sons, 1999.

[6] C. Dwork, R. Kumar, M. Naor, and D. Sivakumar, "Rank Aggregation Methods for the Web," *Proc. 10th Int'l World Wide Web Conf. (WWW '01),* pp. 613-622, 2001.

[7] J. Bar-Ilan, M. Mat-Hassan, and M. Levene, "Methods for Comparing Rankings of Search Engine Results," *Computer Networks,* vol. 50, pp. 1448-1463, 2006.

[8] A.F. Smeaton, "Independence of Contributing Retrieval Strategies in Data Fusion for Effective Information Retrieval," *Proc. 20th BCS-IRSG Colloquium,* 1998.

[9] S.A. Mounir, N. Goharian, M. Mahoney, A. Salem, and O. Frieder, "Fusion of Information Retrieval Engines (FIRE)," *Proc. Int'l Conf. Parallel and Distributed Processing Techniques and Applications (PDPTA),* 1998.

[10] H.P. Young, "An Axiomatization of Borda's Rule," *J. Economic Theory,* vol. 9, no. 1, pp. 1-91, 1974.

[11] J.H. Lee, "Analyses of Multiple Evidence Combination," *Proc. 20th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR '95),* pp. 267-276, 1995.

[12] M. Montague and J.A. Aslam, "Relevance Score Normalization for Metasearch," *Proc. 10th Int'l Conf. Information and Knowledge Management (CIKM '01),* pp. 427-433, 2001.

[13] M. Montague, "Metasearch: Data Fusion for Document Retrieval," PhD dissertation, Dartmouth College, 2002.

[14] W.R. Knight, "A Computer Method for Calculating Kendall's Tau with Ungrouped Data," *J. Am. Statistical Assoc.,* vol. 61, no. 314, pp. 436-439, 1966.

[15] R. Fagin, R. Kumar, and D. Sivakumar, "Comparing Top k Lists," *Siam J. Discrete Math.,* vol. 17, no. 1, pp. 134-160, 2003.

[16] A. Borodin, G.O. Roberts, J.S. Rosenthal, and P. Tsaparas, "Link Analysis Ranking: Algorithms, Theory, and Experiments," *ACM Trans. Internet Technology,* vol. 5, no. 1, pp. 231-297, 2005.

[17] A.R. Rao and P. Bhimasankaram, *Linear Algebra.* Tata-McGraw Hill, 1992.

[18] P. Berkhin, "A Survey on PageRank Computing," *Internet Math.,* vol. 2, no. 1, pp. 73-120, 2005.

[19] A.N. Langville and C.D. Meyer, "A Survey of Eigenvector Methods for Web Information Retrieval," *SIAM Rev.,* vol. 47, no. 1, pp. 135-161, 2005.

[20] S.D. Kamvar, T.H. Haveliwala, C.D. Manning, and G.H. Golub, "Extrapolation Methods for Accelerating Pagerank Computations," *Proc. 12th Int'l World Wide Web Conf. (WWW '03),* pp. 261-270, 2003.

[21] M. Richardson and P. Domingos, "The Intelligent Surfer: Probabilistic Combination of Link and Content Information in Pagerank," *Advances in Neural Information Processing Systems,* vol. 14, pp. 1441-1448, 2002.

[22] J.H. Lee, "Combining Multiple Evidence from Different Properties of Weighting Schemes," *Proc. 18th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR '95),* pp. 180-188, 1995.

[23] J. Hirai, S. Raghavan, A. Paepcke, and H. Garcia-Molina, "Webbase: A Repository of Web Pages," *Proc. 10th Int'l World Wide Web Conf. (WWW),* 2000.

[24] T.H. Haveliwala, "Efficient Computation of Pagerank," technical report, Stanford Univ., 1999.

[25] M.F. Porter, "An Algorithm for Suffix Stripping," *Program,* vol. 14, pp. 130-137, 1980.

[26] G. Salton, A. Wong, and C.S. Yang, "A Vector Space Model for Automatic Indexing," *Comm. ACM,* vol. 18, no. 11, pp. 613-620, 1975.

[27] S. Dominich, "PageRank: Quantitative Model of Interaction Information Retrieval," *Proc. 12th Int'l World Wide Web Conf. (WWW '03),* pp. 13-18, 2003.

[28] B. Debroy and L. Bhandari, "Economic Freedom for the States of India," technical report, Rajiv Gandhi Inst. Contemporary Studies, 2005.

[29] V. Ha and P. Haddawy, "Similarity of Personal Preferences: Theoretical Foundations and Empirical Analysis," *Artificial Intelligence,* vol. 146, no. 2, pp. 149-173, 2003.

[30] D. Hawking, N. Craswell, P. Bailey, and K. Griffiths, "Measuring Search Engine Quality," *Information Retrieval,* vol. 4, pp. 33-59, 2001.

Narayan L. Bhamidipati received the BS (with honors) and MS degrees in statistics from the Indian Statistical Institute, Kolkata. He is awaiting the PhD degree in computer science from the same university, after having submitted his thesis. He is currently with the Data Mining and Research Group, Yahoo! Software Development India Pvt., Bangalore. His interests span data mining, machine learning, and various related topics.

Sankar K. Pal received the PhD degree in radio physics and electronics from the University of Calcutta in 1979 and the PhD degree in electrical engineering along with DIC from Imperial College, University of London in 1982. He is the director and a distinguished scientist at the Indian Statistical Institute. He founded the Machine Intelligence Unit and the Center for Soft Computing Research, a national facility in the Institute in Calcutta. He is a coauthor of 14 books and more than 300 research publications in the areas of pattern recognition and machine learning, image processing, data mining and Web intelligence, soft computing, neural nets, genetic algorithms, fuzzy sets, rough sets, and bioinformatics. He has received many prestigious awards in India and abroad including the 1990 S.S. Bhatnagar Prize. He has also served on the editorial boards of numerous journals including the *IEEE Transactions on Pattern Analysis and Machine Intelligence* and *IEEE Transactions on Neural Networks*. He is a fellow of the IEEE.