COMPARING PARTIAL RANKINGS

Laura Wilke
Mathematics Honors Thesis
Advisor: Professor David Meyer, Ph.D
UCSD
May 13, 2014

Abstract. A partial ranking is a partially known total ordering of a set of objects, i.e., elements in some subsets of the set may be assigned the same rank. A typical example is a top k list, in which all the remaining objects are, as far as list tells us, tied at rank $k + 1$. We derive a metric on the set of partial rankings that, unlike previous metrics, incorporates the intuition that there should be greater significance to changes at the top of the ranking than lower down. We apply this metric to the world ranking of professional men's tennis players to illustrate how it captures the desired phenomena.

Acknowledgements. I would like to thank my advisor, David Meyer, who has offered me guidance and support for my undergraduate thesis. He has believed in me over the past year and a half and has encouraged my academic growth. I am very fortunate that I have had the opportunity to work with him.

Contents

COMPARING PARTIAL RANKINGS

1. Introduction

Metrics on permutations have been well studied and include classical constructions like Kendall's tau [6] and Spearman's Footrule [8]. Over the past decade, the study of metrics on permutations (i.e., full rankings) has focussed on top k lists and partial rankings. Top k lists are exemplified by top 10 lists, which can be the results of a search engine query. Metrics have been developed to compare top k lists with the goal of finding a top k list that is a "good" consolidation of multiple top k lists [2].

Partial rankings arise in several applications such as sports rankings, political rankings and commerce rankings. Due to the way these rankings are formed, there is the possibility of ties. For example, the Chinese Communist Party produces a ranked list of the Central Committee members, however, there are cases in which several members are tied since the Party does not disclose complete information. By comparing the partial rankings from National Congress to National Congress, we are able to gain information about succession within the Party [7]. Another application is online rating systems, such as for restaurants or hotels, in which even with multiple criteria ties can arise.

We are specifically interested in timeseries of partial rankings, in which we will quantify changes from one time step to the next. In the context of tennis, for example, changes arise from individual players moving up and down in the rankings each week. There is a connotation that when an element is in position 1, the top position of the partial ranking, that is the best position, so moving up into that position is signifiant, as is moving out of it. An ideal comparison of partial rankings would incorporate this real world significance of different ranks. Current measures treat each rank change equally, so real world significance from one change to the next is not accurately described.

In this paper, we will focus on improving upon an existing family of metrics on partial rankings. This is obtained by using the generalization of the L_1-norm on permutations as described in a recent paper of Fagin, Kumar, Mahdian, Sivakumar and Vee [3]. In order to improve upon this metric, we introduce a Rank Transform Function (RTF) that places a weight on each rank, and therefore on each change in rank. This function weights changes at the top of a ranking more than changes at the bottom. We prove that the L_p-norm remains a metric even when the ranks are transformed by the RTF. We use real data to compare our metric with Fagin, et al.'s, and show that ours is better at accounting for the real-world significance of top ranks.

1.1. Metrics on Permutations.

Metrics on permutations quantify how permutations differ from each other. A *permutation*, ρ, is a bijection from a set D onto $[N] = \{1, \ldots, N\}$, where $N = |D|$. For a permutation, $\rho(i)$ is interpreted as the position or rank of the element i.

Definition 1. *A metric on a set X is a function $d : X \times X \to \mathbb{R}$ satisfying the following conditions:*

(1) $d(x,x) = 0$ $\forall x \in X$

(2) $d(x,y) > 0$ $\forall x \neq y \in X$

(3) $d(x,y) = d(y,x)$ $\forall x, y \in X$

(4) $d(x,y) \leq d(x,z) + d(y,z)$ $\forall x, y, z \in X.$

The classical metrics that have been well-studied for full rankings are Kendall's tau [6] and the L_p-norm.

Definition 2. *The Kendall tau distance between two permutations ρ_1 and ρ_2 is*

$$\tau(\rho_1,\rho_2) = \left|\left\{(i,j) : i<j, \left(\rho_1(j) < \rho_1(j) \wedge \rho_2(i) > \rho_2(j)\right) \vee \left(\rho_1(i) > \rho_1(j) \wedge \rho_2(i) < \rho_2(j)\right)\right\}\right|.$$

The Kendall tau distance between two full rankings is the number of pairs that are in a different order in the rankings. It is the minimum number of pairwise adjacent transpositions needed to transform ρ_1 into ρ_2 [1]. For example, if given the following permutations ρ_1=(3,1,6,2,5,4) and ρ_2=(1,3,6,4,2,5), the Kendall tau distance between ρ_1 and ρ_2 is 3 ($\tau(\rho_1,\rho_2) = 3$).

Definition 3. *The L_1-norm, Spearman's Footrule, is defined by*

$$\ell_1(\rho_1,\rho_2) = \sum_i |\rho_1(i) - \rho_2(i)|.$$

The L_1-norm is the sum of the absolute values of the differences in ranks between two permutations. Using the example above, $\ell_1(\rho_1,\rho_2) = 10$. The maximum value occurs when ρ_1 and ρ_2 are in reverse orders.

1.2. **Hausdorff Metric.** The Hausdorff metric measures how far two subsets of a metric space are from each other [4]. Top k lists and partial rankings are viewed as subsets of consistent full rankings so the Hausdorff distance between such subsets gives a distance between partial rankings.

Definition 4. *Let X and Y be two non-empty subsets of a metric space (M,d). The Hausdorff distance is defined by*

$$d_H(X,Y) := \max\{\max_{x\in X} d(x,Y), \max_{y\in Y} d(y,X)\},$$

where

$$d(x,Y) := \min_{y\in Y} d(x,y).$$

The quantity $\min_{y\in Y} d(x,y)$ is the distance between x and the set Y. Therefore, the quantity $\max_{x\in X} d(x,Y)$ is the maximal distance of a member of X from the set Y. Similarly, the quantity $\max_{y\in Y} d(y,X)$ is the maximal distance of a member of Y from the set X. The Hausdorff distance between X and Y is the maximal distance of a member of X or Y from the other set.

1.3. **Metrics on Top k Lists.** A top k list is the top k of a full ranking so metrics on permutations can be applied to top k lists through the Hausdorff metric.

Definition 5. *A top k list, τ, is a bijection from a domain D_τ to $[k]$* [2].

The element i appears in the top k list τ if $i \in D_\tau$ where $\tau(i)$ is the rank of i in τ. If τ is a top k list and ρ is a permutation on $D \supseteq D_\tau$, then we say that ρ is an extension of τ, if $\rho(i) = \tau(i)\ \forall i \in D_\tau$.

When we are comparing the top k, the elements might not be the same in each list. Let $D = D_{\tau_1} \cup D_{\tau_2}$; $Z = D_{\tau_1} \cap D_{\tau_2}$; $S = D_{\tau_1} \backslash D_{\tau_2}$; $T = D_{\tau_2} \backslash D_{\tau_1}$ [2].

Crtichlow introduced "induced Hausdorff metrics" in order to determine the distance between two top k lists [1]. The distance between two top k lists using Kendall's tau is

$$\tau_{\text{Haus}}(\rho_1, \rho_2) = \frac{1}{2}(k-z)(5k-z+1) + \sum_{i,j \in Z} \tau_{i,j}(\rho_1, \rho_2) - \sum_{i \in S} \rho_1(i) - \sum_{i \in T} \rho_2(i),$$

where $z = |Z|$ and $\tau_{i,j}(\rho_i, \rho_2) = 1$ if ρ_1 and ρ_2 rank i and j oppositely, and vanishes otherwise [2]. Just as Kendall's tau can be "induced" with the Hausdorff metric, so can the L_1-norm. Similarly, Fagin et al. found [2]

$$F_{\text{Haus}}(\rho_1, \rho_2) = (k-z)(3k-z+1) + \sum_{i \in Z} |\rho_1(i) - \rho_2(i)| - \sum_{i \in S} \rho_1(i) - \sum_{i \in T} \rho_2(i)$$

.

1.4. **Metrics on Partial Rankings.** A partial ranking is a ranking in which the ordering is not complete [1]. The formulas that have been used to determine the distance between top k lists can not be applied to partial rankings because those metrics do not completely deal with the ties.

Definition 6. *A partial ranking on D, with $|D| = N$, is a map $\sigma : D \to [N]$ such that $|\sigma^{-1}([n-1])| > n-1 \Rightarrow \sigma^{-1}(n) = \emptyset$.*

Definition 7. *A permutation $\rho : D \to [N]$ is an extension of a partial ranking σ on D for all x, $y \in D$, $\sigma(x) < \sigma(y) \Rightarrow \rho(x) < \rho(y)$. Let $S_D \supset R_\sigma := \{\rho \in S_D | \rho$ is an extension of $\sigma\}$.*

A partial ranking, σ is a permutation in which ties occur. A tie is defined when two or more elements in the partial ranking have the same rank or position. For example, if $\sigma(x) = \sigma(y) = \sigma(z)$, x, y, z are tied at the same position. A top k list is a very specific type of partial ranking. There are k elements in the partial ranking with known positions. The remaining elements in the domain are not listed in the top k are all tied in the position $k+1$.

The reverse of a partial ranking with domain D is defined as $\sigma^R(d) = |D| + 1 - \sigma(d)$ for all $d \in D$. For two partial rankings α and β with domain D, it is said that α is a refinement of β if for $i, j \in D$, $\alpha(i) < \alpha(j)$ whenever $\beta(i) < \beta(j)$. A τ refinement of σ, $(\tau * \sigma)$, is the refinement of σ with the following properties: for $i, j \in D$ if

$$\sigma(i) = \sigma(j) \text{ and } \tau(i) < \tau(j) \quad \Rightarrow \quad (\tau * \sigma)(i) < (\tau * \sigma)(j); \tag{5}$$

$$\sigma(i) = \sigma(j) \text{ and } \tau(i) = \tau(j) \quad \Rightarrow \quad (\tau * \sigma)(i) = (\tau * \sigma)(j); \tag{6}$$

$$\sigma(i) < \sigma(j) \quad \Rightarrow \quad (\tau * \sigma)(i) < (\tau * \sigma)(j). \tag{7}$$

With this notation, the Hausdorff distance induced by the L_1-norm for partial rankings is

$$F_{Haus}(\sigma, \tau) = \max\{F(\omega * \tau^R * \sigma, \omega * \sigma * \tau), F(\omega * \tau * \sigma, \omega * \sigma^R * \tau)\},$$

where ω is any full ranking, σ, τ are partial rankings and F is Spearman's Footrule distance, i.e., the L_1-norm [3].

We will be using the following global notation throughout the paper.

Notation 1. *When two partial rankings σ_1 and σ_2 are understood, $D = D_{\sigma_1} \cup D_{\sigma_2}$; $Z = D_{\sigma_1} \cap D_{\sigma_2}$; $S = D_{\sigma_1} \backslash D_{\sigma_2}$; $T = D_{\sigma_2} \backslash D_{\sigma_1}$ [2].*

2. Results

We will use the metric on S_D defined by the L_p-norm for $p \geq 1$,

$$\ell^p(\rho_1, \rho_2) = (\sum_d |\rho_1(d) - \rho_2(d)|^p)^{1/p}.$$

The value of p determines the relative importance of small versus large differences in position; the larger p is, the more important are larger changes. When ℓ^p is extended to a metric on partial rankings, we will necessarily compute distances between permutations containing some maximized differences in positions.

2.1. Rank Transform Function.

In practice, one views a difference at a smaller rank to be more significant than the same difference at a larger rank. In order to add this significance into a metric, we introduce a Rank Transform Function. The classical metrics for comparing partial rankings treat each change in rank as the same so changes at the larger ranks are given the same significance as those at a smaller rank. In order for changes at the top of the partial ranking to make a larger contribution than the same changes at the bottom, we use a Rank Transform Function.

Definition 8. *Let $f : \mathbb{R} \to \mathbb{R}$, which we will use as a Rank Transform Function (RTF): Define*

$$\ell_f^p(\rho_1, \rho_2) := \ell^p(f \circ \rho_1, f \circ \rho_2).$$

Then ℓ_f^p is symmetric and positive semidefinite for any f. Also, the triangle inequality holds:

Proof.

$$\ell_f^p(\rho_1, \rho_2) + \ell_f^p(\rho_2, \rho_3) = \ell^p(f \circ \rho_1, f \circ \rho_2) + \ell^p(f \circ \rho_2, f \circ \rho_3) \geq \ell^p(f \circ \rho_1, f \circ \rho_3) = \ell_f^p(\rho_1, \rho_3).$$

□

Proposition 1. *If f is a bijection then ℓ_f^p is a metric on S_D.*

Proof. It remains to be shown that if f is bijective the $\ell_f^p(\rho_1, \rho_2) = 0 \Leftrightarrow \rho_1 = \rho_2$. Certainly if $\rho_1 = \rho_2, \ell_f^p(\rho_1, \rho_2) = 0$. And if $\rho_1 \neq \rho_2, f \circ \rho_1 \neq f \circ \rho_2$ since f is bijective, whence $\ell_f^p(\rho_1, \rho_2) = \ell^p(f \circ \rho_1, f \circ \rho_2) \neq 0$. □

If f is bijective and continuous, then it is necessarily monotone and we will assume that each RTF is monotone increasing such that $f'(r) > 0$, where r is the rank. We are interested in permutations ρ where the value of $\rho(i)$ is interpreted as the rank of i. Then $f(r)$ determines how important a difference at rank r is compared with differences at other ranks.

In most of the applications we consider, if not all, it is natural to think of a difference at a smaller rank as being more significant than the same difference at a larger rank. This implies that f' should be decreasing, i.e., $f'' \leq 0$, so f is concave.

Lemma 1. *For any monotone increasing $f : \mathbb{R} \to \mathbb{R}$ and $\rho \in S_D$, if $\rho(x) < \rho(y)$ and $1 \le r < s \le N$, then*

$$|f(\rho(x)) - f(r)|^p + |f(\rho(y)) - f(s)|^p \le |f(\rho(x)) - f(s)|^p + |f(\rho(y)) - f(r)|^p.$$

Proof. Since f is monotone increasing, we have $a := f(\rho(x)) < f(\rho(y)) =: b$ and $c := f(r) < f(s) =: d$. Hence, we want to show

$$|a - c|^p + |b - d|^p \le |a - d|^p + |b - c|^p.$$

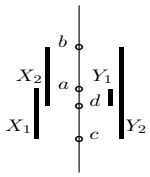
Let the convex function $g(x) = |x|^p$ for $p \ge 1$. There are three cases which arise as follows: i) $c < d < a < b$; ii) $c < a < d < b$; iii)$a < c < d < b$.

Case 1. *Let $X_1 = a - c$, $X_2 = b - d$, $Y_1 = b - c$, and $Y_2 = a - d$. The sequence $Y = (Y_1, Y_2)$ majorizes $X = (X_1, X_2)$. $Y_1 \ge X_1$ and $Y_1 + Y_2 = X_1 + X_2$. Since $g(x)$ is convex, using Karamata's inequality* [5]*, one obtains*

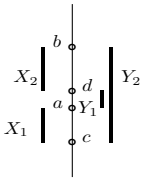
$$g(Y_1) + g(Y_2) \ge g(X_1) + g(X_2).$$

This can be rewritten as:

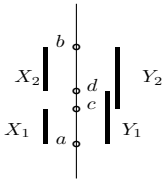
$$|a - c|^p + |b - d|^p \le |a - d|^p + |b - c|^p.$$



Case 2. *The inequality holds term by term.*



Case 3. *The inequality holds term by term.*



□

2.2. **Rank Transformed L_p Metric.** Given any metric on S_D, the Hausdorff metric [4] on subsets of S_D defines a metric on partial rankings:

$$d(\sigma_1, \sigma_2) := d_H(R_{\sigma_1}, R_{\sigma_2}) := \max\{\max_{\rho_1 \in R_{\sigma_1}} d(\rho_1, R_{\sigma_2}), \max_{\rho_2 \in R_{\sigma_2}} d(\rho_2, R_{\sigma_1})\},$$

where

$$d(\rho_1, R_{\sigma_2}) := \min_{\rho_2 \in R_{\sigma_2}} d(\rho_1, \rho_2),$$

and similarly for $d(\rho_2, R_{\sigma_1})$. To compute this distance between partial rankings efficiently, we apply Lemma 1.

In order to break ties in a partial ranking σ_1, we will use the refinement described earlier. The permutations that resolve the ties in σ_1 do so in such a way as to maximize the distance to the closest extension of σ_2. Lemma 1 implies that any set of elements tied in σ_1 should be placed into the reverse order in which they are ranked in σ_2. This permutation, which is an extension of σ_1 is defined as: $\kappa_1 := \omega * \sigma_2^R * \sigma_1$. Similarly for κ_2 which is an extension of σ_2. Now Lemma 1 implies that any set of elements tied in σ_2 should be put into the order of which they are ranked in σ_1. The permutation which is an extension of σ_2 is defined as: $\mu_2 := \omega * \sigma_1 * \sigma_2$. Similarly for μ_1 which is an extension of σ_1. This proves the following theorem:

Theorem 1. *With the definitions in the preceding paragraph,*

$$d(\sigma_1, \sigma_2) = \max\{d(\kappa_1, \mu_2), d(\mu_1, \kappa_2)\}.$$

Still using the definitions in the paragraph preceding Theorem 1, when all the elements are tied beneath the last rank of the partial ranking, $d(\kappa_1, \mu_2)$ becomes:

(8)

$$d(\kappa_1, \mu_2) = (\sum_{t_i \in T} |f(\mu(t_i)) - f(N+1-i)|^p + \sum_{s_i \in S} |(f(\kappa(s_i)) - f(|D_{\sigma_2}|) + i)|^p,$$

(9)

$$+ \sum^{N-|S|-|T|-|Z|} |f(|D_{\sigma_1}| + i) - f(|S| + |T| + |Z| + i)|^p + \sum_{z \in Z} |f(\kappa_1(z)) - f(\mu_2(z))|^p)^{1/p};$$

and similarly for $d(\mu_1, \kappa_2)$.

3. Application to Data

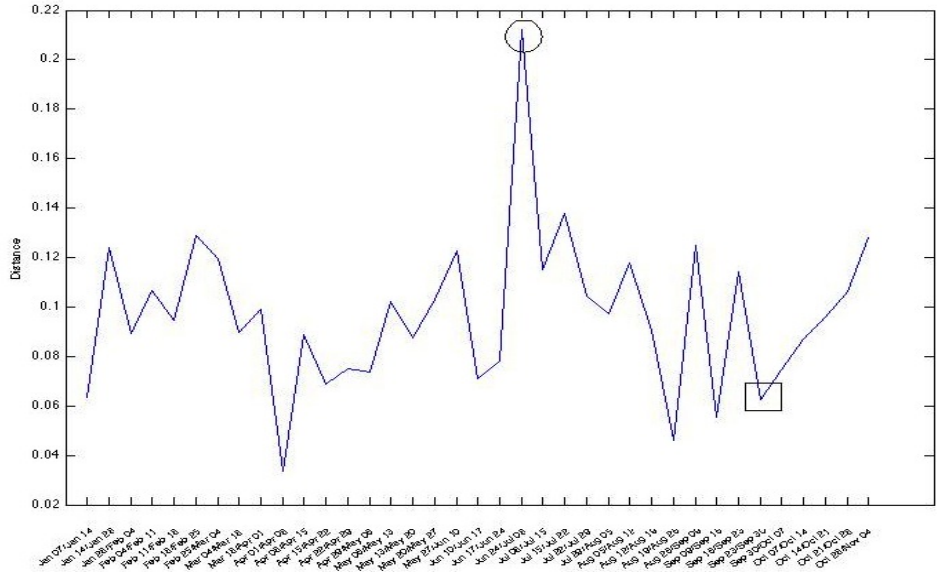
In order to determine if the Rank Transformed L_p metric gives more information about how partial rankings change, we apply F_{Haus} metric and Rank Transformed L_p metric with $p = 1$. If $p > 1$, then the relative contribution of large distances is not minimized. Therefore the only difference between the metrics is the Rank Transform Function. The RTF that is used is $f(r) = -1/r$. We will examine the Association of Tennis Professional's partial rankings of Men's Singles Tennis Players from January to early November of 2013.

Before we apply the metrics, it is useful to understand how tennis players are ranked. Points are assigned to the losers of specific rounds in tournaments. Hence if a player wins in the third round, that player gets the number of points specified for that round. If players have to qualify to be in the main draw of the tournament,

they receive qualifying points as well. Different tournaments are worth different numbers of points, i.e., Grand Slams such as Wimbledon, have the most points [9].

For the application, we will be looking at the Top 100 Men's Singles players where the value of $N = 130$. The value of N is chosen based on what the last rank is for a player enters the top 100 from one week to the next.

FIGURE 1. Men's Top 100 Players January - November 4 2013 Partial Ranking Distances with Fagin et al.'s Metric

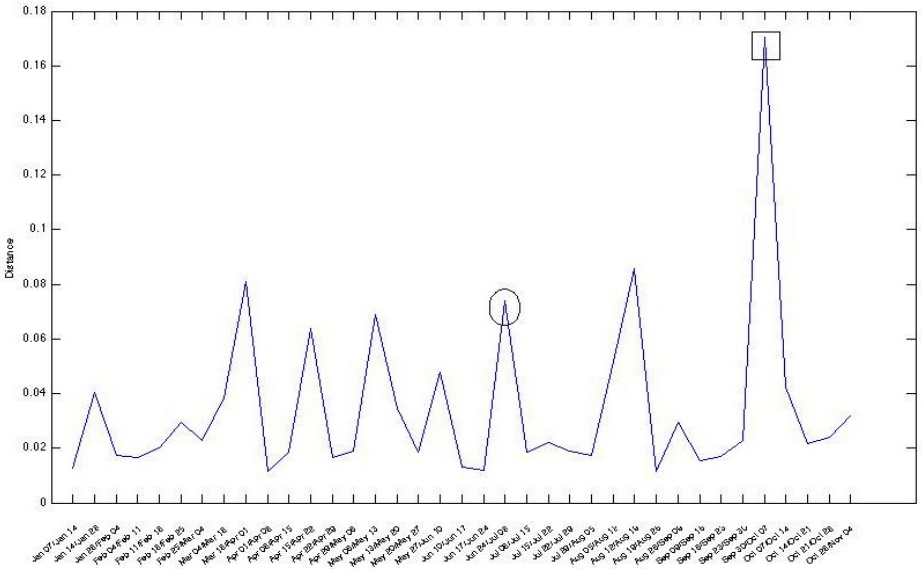


First, we will look at the maximum in the graph of Fagin et al. results (Figure 1), which occurs at $d(\text{June24, July8})$. The distance at this point is the maximum so one would suspect that a lot of change occurred, which is an accurate interpretation of the large distance. The changes that occur however, are at the bottom half of the partial ranking. A lot of changes occurred in the bottom half because Wimbledon took place between June 24 and July 7. Several players' ranks changed due to participating in qualifying tournaments so the bottom half of the ranking was more unstable compared to the top half. This metric is unable to determine what changes are significant. Since changes in the Top 25 are the most significant in the tennis rankings, when the distance is dominated by the changes at the bottom of the ranking, the metric is not capturing what it should be.

When we look at $d(\text{June24, July8})$, in the graph of the Rank Transformed L_p metric results (Figure 2), we notice that it is not a maximum. This a more accurate representation of what is happening in the partial ranking. Since the majority of changes that occur are at the bottom half of the ranking, they contribute less to the total distance because of the RTF. The RTF gives less importance to the changes at the bottom of the ranking compared to the top.

Lastly, we will look at $d(\text{September30, October7})$, when the most significant event of the year took place, when Rafael Nadal overtook Novak Djokovic as the best player in the world, ranked number 1. The F_{Haus} metric does not produce a

FIGURE 2. Men's Top 100 Players January - November 4 2013
Partial Ranking Distances with Rank Transformed L_p Metric



maxima between these partial rankings. There are not many large changes in the partial rankings which is why the distance is not a maxima. Fagin et al.'s metric treats each change in a ranking as the same hence, changes at the top of the ranking are viewed the same as changes at the bottom of the ranking, which is not how changes are viewed in application.

However, when we examine the Rank Transformed L_p metric results for d(September30, October7), we see that it is indeed the maximum. This is an accurate representation of the significance of the changes that have occurred. The RTF allows the change in position one to contribute the most distance to the total distance. The RTF is essential in distinguishing what rank changes are significant and which ones are of less significance.

4. Conclusion

In this paper, we developed a metric that accurately describes the changes in partial rankings. With the use of the Rank Transform Function, rank changes have a weight that accurately describes how rank changes are viewed in the world, changes at the top carrying more weight than changes at the bottom. We have proved that with the inclusion of the RTF, we still have a metric. In application, the Rank Transformed L_p metric describes the rank changes that are occurring better than the existing metrics.

References

[1] Critchlow, *Metric Methods for Analyzing Partially Ranked Data*, Lecture Notes in Statistics. 1-77, Springer-Verlag, Berlin, 1980.

[2] Fagin, Kumar, Sivakumar, "Comparing Top k Lists", SIAM J. Discrete Math (2003) Volume 17, 134-145.

[3] Fagin, Kumar, Mahdian, Sivakumar, Vee, "Comparing Partial Rankings", SIAM J. Discrete Math (2006) Volume 20, 628-636.
[4] F. Hausdorff, Grundzüge der Mengenlehre (Leipzig: Veit and company 1914).
[5] J. Karamata, "Sur une inégalité rélative aux fonctions convexes", Publ. Math. Univ. Belgrade 1 (1932) 145-148.
[6] M. Kendall, "A new measure of rank correlation", Biometrika, 30 (1938).
[7] D. Meyer, M. Ram, G. Shaw, L. Wilke, "Quantifying Political Transformation in China", in preparation (UCSD 2014)
[8] C. Spearman, "The Proof and Measurement of Association Between Two Things", The American Journal of Psychology(1904), Vol 15, No. 1,72-101.
[9] 2014 ATP World Tour - Rule Book, Chapter IX: Emirates ATP Ranking