Reconstructing Hidden Permutations Using the Average-Precision (AP) Correlation Statistic

Lorenzo De Stefani, Alessandro Epasto, Eli Upfal
Brown University
Providence, RI 02906, United States
{lorenzo, epasto, eli}@cs.brown.edu

Fabio Vandin
University of Padova
Padova, PD 35131, Italy
vadinfa@dei.unipd.it

Abstract

We study the problem of learning probabilistic models for permutations, where the order between highly ranked items in the observed permutations is more reliable (i.e., consistent in different rankings) than the order between lower ranked items, a typical phenomena observed in many applications such as web search results and product ranking. We introduce and study a variant of the Mallows model where the distribution is a function of the widely used Average-Precision (AP) Correlation statistic, instead of the standard Kendall's tau distance. We present a generative model for constructing samples from this distribution and prove useful properties of that distribution. Using these properties we develop an efficient algorithm that provably computes an asymptotically unbiased estimate of the center permutation, and a faster algorithm that learns with high probability the hidden central permutation for a wide range of the parameters of the model. We complement our theoretical analysis with extensive experiments showing that unsupervised methods based on our model can precisely identify ground-truth clusters of rankings in real-world data. In particular, when compared to the Kendall's tau based methods, our methods are less affected by noise in low-rank items.

1 Introduction

Probabilistic models of ranking data have been studied extensively in statistics (Mallows 1957), machine learning (Awasthi et al. 2014) and theoretical computer science (Braverman and Mossel 2008; Chierichetti et al. 2014). Applications of ranking models include understanding user preferences in electoral systems, ordering web search results, aggregating crowd-sourcing data, and optimizing recommendation systems results (Vigna 2015; Saari 2006; Dwork et al. 2001; Sorz et al. 2015; Brian 2008; Yilmaz, Aslam, and Robertson 2008; Melucci 2007).

Most of the analytic work in this area has focused on the Mallows model (1957) which defines a probability distribution over a set of permutations of n elements given a fixed *center* permutation π (the *ground truth*), and a *dispersion* parameter $\beta > 0$. The probability of a permutation σ in Mallows model is $\Pr_{\mathcal{M}(\beta,\pi)}(\sigma) = N_\beta^{-1} \exp(-\beta d_K(\pi,\sigma))$, where $d_K(\pi,\sigma)$ is the Kendall's tau distance between π and σ, and N_β is a normalization factor independent of σ (see Section 3 for more details). Since the Kendall's tau distance simply counts the number of pairs of items whose order is inverted in σ with respect to their order in π, a permutation with inversions of elements occupying positions towards the end of the ranking has the same probability as a permutation with the same number of inverted pairs near the top of the ranking.

In many practical applications, such as web search ranking, voter preferences surveys, consumer evaluation systems and recommendation systems we expect the input samples to provide a more reliable order for items at the top of the ranking than for "less important" items at the bottom of the ranking (Melucci 2007; Vigna 2015; Yilmaz, Aslam, and Robertson 2008). The limitation of the Kendall's tau distance and the associated Mallows model, when applied to such data has been pointed out in a number of recent works (Vigna 2015; Yilmaz, Aslam, and Robertson 2008; Kumar and Vassilvitskii 2010). Note that the Spearman's footrule distance (Spearman 1904), another widely used statistic, suffers from the same problem.

In this work we address the limitation of the standard Mallows distribution by proposing an alternative exponential distribution of permutations, $\mathcal{M}_{\mathcal{AP}}(\beta,\pi)$, in which the Kendall's tau distance of the Mallows distribution is replaced by a new distance based on the widely used Average Precision (AP) correlation statistic (Yilmaz, Aslam, and Robertson 2008). The AP statistic (and hence the AP model) takes into account not only the number of miss-ranked pairs but also the locations (and hence the importance) of the miss-placed pairs in the ranking. Let $[n]$ be the set of natural numbers $\{1,\ldots,n\}$, and let S_n be the set of all permutations over $[n]$. Given a fixed *center* permutation $\pi \in S_n$ and a *dispersion* parameter $\beta > 0$, the probability associated with a permutation $\sigma \in S_n$ according to the new model is $\Pr_{\mathcal{M}_{\mathcal{AP}}(\beta,\pi)}(\sigma) = Z_\beta^{-1} \exp(-\beta d_{\mathcal{AP}}(\pi,\sigma))$, where $d_{\mathcal{AP}}(\pi,\sigma)$ is the AP-distance between π and σ and Z_β is a normalization factor independent of σ (see Section 3 for more details). To the best of our knowledge, no work has addressed the problem of modeling permutations with provable guarantees subject to such properties.

We now summarize the main contributions of this work:

- We introduce a novel variant of the standard Mallows

model based on the more nuanced AP statistic.

- We introduce a generative process for this probability distribution. This process, besides defining an efficient algorithm for sampling permutations from the distribution, is a useful tool in establishing formal properties of our model.
- We provide bounds on the probability of swapping elements and use these bounds to quantify the number of samples required for learning the parameters of the model.
- We design efficient algorithms that given $\mathcal{O}(\log n/(n\beta))$ samples learn the central permutation with high probability for $\beta = \Omega(1/n)$. We also show an alternative algorithm that efficiently computes an asymptotically unbiased and consistent estimator of the central permutation for any $\beta > 0$.
- We experimentally evaluate our model and algorithms with both synthetic and real-world data. We show experimentally the accuracy of the estimators in the reconstruction of the hidden permutation with a limited number of samples. We also show with real data that unsupervised methods based on our model can precisely (and efficiently) identify ground-truth clusters of rankings with tens of thousand of elements. Moreover we show that our method is less affected by noise in the lowest positions of the ranking, which is more likely to occur (Vigna 2015), than the Kendall-based ones. Finally we show that simple supervised classification algorithms based on the AP statistic outperform classifications based on the Kendall-based distance measure.

The new $\mathcal{M}_{\mathcal{AP}}(\beta, \pi)$ model is significantly more challenging for mathematical analysis than the Mallows model. The $\mathcal{M}_{\mathcal{AP}}(\beta, \pi)$ model does not have many of the simplifying features used in the analysis of Mallows model, such as translational invariant of the distribution (Chierichetti et al. 2014). Nevertheless, as we show in this work, the model remains sufficiently mathematically tractable to derive non-trivial theoretical results. Furthermore, while many of the results for the Mallows model require a constant β, our analysis applies to $\beta = \beta(n)$ that is decreasing in n.

The paper is structured as follows. Section 2 reviews part of the literature which is relevant for our problem. Section 3 formally defines the $\mathcal{M}_{\mathcal{AP}}(\beta, \pi)$ model and provides some intuition on the behavior of the AP model. Section 4 defines a generative process for the distribution that will be pivotal in the analysis of our model. Section 5 presents the in depth study the mathematical properties of our model and the learning algorithms designed on this study. Section 6 presents an experimental evaluation of our method on real and synthetic data. Finally, Section 7 draws conclusions and hints at possible future directions stemming from our work.

For lack of space, most proofs are omitted from this paper. All proofs are available in the full version of the paper (De Stefani et al. 2015).

2 Related Work

Ranking problems have been studied extensively since at least the 18th century works of Borda and Condorcet on social choice theory (Brian 2008; Saari 2006). We discuss here only the most relevant works derived from Mallow's original work.

Properties of the Mallows model (1957) were studied in (Diaconis and Ram 2000; Fligner and Verducci 1988). Tail bounds on the displacement of an element were studied in (Braverman and Mossel 2009; Bhatnagar and Peled 2014). Reconstructing the hidden central permutation was studied in (Braverman and Mossel 2009; Meila et al. 2012; Chierichetti et al. 2014; Lu and Boutilier 2011). Finding the maximum likelihood permutation is equivalent to the well known rank aggregation problem which is NP-hard (Bartholdi III, Tovey, and Trick 1989) and has a polynomial-time approximation scheme (Kenyon-Mathieu and Schudy 2007). Several heuristics were introduced in order to solve this problem (Meila et al. 2012; Schapire and Singer 1998) with no provable guarantees. The problem of learning a mixture of Mallows models was studied in (Awasthi et al. 2014; Chierichetti et al. 2014). The Mallows model has also been extended in various ways by generalizing the Kendall's tau distance in order to weigh the number of inversions with respect to previously defined parameters for each element (Fligner and Verducci 1986; 1988). In the Recursive Inversion Model (Meek and Meila 2014) different weights are assigned to the inversions of elements, requiring a priori specification of $n-1$ weights. A priori assigning such weights is not easy. In contrast, the AP model we consider gradually and coherently weighs the elements based on their positions.

Recent work (Qin, Geng, and Liu 2010) generalizes the Mallows model to other distances in which the maximum likelihood can be efficiently computed. Shieh (1998) proposed a weighting scheme for Kendall's tau distance where the weight of errors depends on the ranks of the inverted items. Kumar and Vassilvitskii (2010) later extended Shieh's model defining an even more sophisticated scheme that takes into account positions of the items, weights, and their similarity. Various works (Fagin, Kumar, and Sivakumar 2003; Haveliwala et al. 2002) addressed the issue of assigning different weights to various elements by restricting the computation of some statistics only on specific parts of the ranking.

The AP correlation statistic, which is the focus of this work, was introduced by Yilmaz et al. (Yilmaz, Aslam, and Robertson 2008; Yilmaz and Aslam 2006). Since its introduction, AP correlation has been widely used in the literature (for instance (Vigna 2015; Sakai 2012; Schubert et al. 2012) and see the references of these papers). Recently (Vigna 2015) generalized the AP correlation and other metrics to handle ties and other ways to weighting inversions. We believe that our approach could be adapted to the corresponding extension of the AP-distance. To the best of our knowledge no prior work has addressed the problem of provably reconstructing the center permutation of a generalized Mallows model using AP distance or similar more nuanced measures.

3 The AP model

For $\pi \in S_n$ let π_i be the i-th element in the order defined by π, and $\pi(i)$ is the position of $i \in [n]$ in π. We use the notation $i <_\pi j$ to indicate that π ranks i before j, or

equivalently, that $\pi(i) < \pi(j)$. Finally we use $\pi[t]$ to denote the prefix of the t highest ranked elements in π, i.e. $\pi[t] = (\pi_1, \ldots, \pi_t)$.

Our new model uses the following *AP distance* of a permutation σ from permutation π:

$$d_{\mathcal{AP}}(\pi, \sigma) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} E_{ij} \frac{n}{2(j-1)},$$

where $E_{ij} = 1$ iff item π_i is ranked after item π_j in permutation σ, and $E_{ij} = 0$ otherwise. Note that the AP-distance $d_{\mathcal{AP}}(\pi, \sigma)$ is defined with respect to a ground truth π and it is therefore not symmetric. Notice that the Kendall's tau distance $d_K(\pi, \sigma)$ can be expressed in the same form by assigning cost 1 (instead of $\frac{n}{2(j-1)}$) for each inversion, i.e. $d_K(\pi, \sigma) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} E_{ij}$.

The AP-distance has the same range and extreme points as the Kendall's tau distance: $0 \leq d_{\mathcal{AP}}(\pi, \sigma) \leq \binom{n}{2}$, $d_{\mathcal{AP}}(\pi, \sigma) = 0$ iff $\pi = \sigma$, and $d_{\mathcal{AP}}(\pi, \sigma) = \binom{n}{2}$ iff σ is obtained by reversing π. However, the AP-distance assigns weights to inverted pairs in σ which depend on the locations of the items in π. More specifically, the cost assigned to inversion of elements (π_i, π_j) for $i < j$ is $\frac{n}{2(j-1)}$. The cost assigned by AP is strictly higher than the one given by tau distance (always 1) for elements in the first half of the ranking, i.e. for $j < \frac{n}{2}+1$ and strictly lower than 1 for $j > \frac{n}{2}+1$.

Given this distance function we define the AP model $\mathcal{M}_{\mathcal{AP}}(\beta, \pi)$:

Definition 1 (AP model)**.** *The* AP model *is a probability distribution over S_n. Given a fixed* center *permutation π and a* dispersion *parameter $\beta > 0$, the probability of a permutation $\sigma \in S_n$ is*

$$\Pr_{\mathcal{M}_{\mathcal{AP}}(\beta,\pi)}(\sigma) = Z_\beta^{-1} \exp(-\beta d_{\mathcal{AP}}(\pi, \sigma)), \qquad (1)$$

where $Z_\beta = \sum_{\sigma \in S_n} \exp(-\beta d_{\mathcal{AP}}(\pi, \sigma))$ is a normalization coefficient independent of σ.

Notice that our model differs from the traditional Mallows model $\mathcal{M}(\beta, \pi)$ only in the use of the AP distance $d_{\mathcal{AP}}(\pi, \sigma)$ instead of the Kendall's tau distance $d_K(\pi, \sigma)$.

Before studying the details of the model we show that, as in the traditional Mallows model, finding the maximum likelihood center permutation for an *arbitrary* multi-set of permutations is NP-hard.

More precisely, we define the AP-ML problem as follows. Given an arbitrary multiset P of elements of S_n (and $\beta > 0$) find the permutation $\pi_{AP} \in S_n$ such that:

$$\pi_{AP} = \arg\max_{\pi \in S_n} \prod_{\sigma \in P} \Pr_{\mathcal{M}_{\mathcal{AP}}(\beta,\pi)}(\sigma)$$

Theorem 1. *The AP-ML problem is NP-hard*

Proof available in the extended version. The result follows by reduction from the Kendall-ML problem, which is NP-Hard (Braverman and Mossel 2008).

Note that for any $\beta > 0$ we have that a permutation π which is a solution of the AP-ML problem for P is the one that minimizes the average AP distance to the multi-set P, that is also NP-hard to find.

We stress that the previous NP-hardness result holds for arbitrary multi-sets (i.e. not generated by the AP model distribution). As we will see in the rest of the paper, when the permutations are generated according to the AP model distribution, it is possible to learn important properties of the model in polynomial time.

4 Generative Process for $\mathcal{M}_{\mathcal{AP}}(\beta, \pi)$

We give a randomized algorithm for constructing a permutation $\sigma \in S_n$ according to the $\mathcal{M}_{\mathcal{AP}}(\beta, \pi)$ model. The algorithm is based on the one presented in (Diaconis and Ram 2000) and provides useful insight on the probability distribution for the AP model. Recall that $\pi[i]$ refers to the prefix of length i of the permutation π.

The $\mathcal{M}_{\mathcal{AP}}(\beta, \pi)$ Generative Process: Let $\pi = \pi_1, \pi_2, \ldots, \pi_n$ be the central permutation. We generate a random permutation σ in n steps. At step i we produce a permutation σ^i in the following way:

1. $\sigma^1 = (\pi_1)$;
2. for $i = 2$ to n do:
   (a) choose a random value $0 \leq r \leq i-1$ according to the probability

$$\Pr(r = j) = \frac{1}{B_i} \exp\left(-\frac{\beta n}{2} \frac{j}{i-1}\right),$$

   where

$$B_i = \sum_{j=0}^{i-1} \exp\left(-\frac{\beta n}{2} \frac{j}{i-1}\right).$$

   (b) Generate a permutation σ^i from σ^{i-1} by placing item π_i after item σ^{i-1}_{i-1-r}. i.e., for $1 \leq t \leq i-1-r$ we have $\sigma^i_t = \sigma^{i-1}_t$, then $\sigma^i_{i-r} = \pi_i$, and for $i-r+1 \leq t \leq i$ we have $\sigma^i_t = \sigma^{i-1}_{t-1}$. If $r = i-1$, we have $\sigma^i_1 = \pi_i$.
   (c) Finally output $\sigma = \sigma^n$.

The proof of the next result is in the extended version:

Theorem 2. *For $i = 1, \ldots, n$, the permutation σ^i generated by the $\mathcal{M}_{\mathcal{AP}}(\beta, \pi)$ Generative Process has distribution $\mathcal{M}_{\mathcal{AP}}\left(\frac{n}{i}\beta, \pi[i]\right)$. In particular the output permutation $\sigma = \sigma^n$ has distribution $\mathcal{M}_{\mathcal{AP}}(\beta, \pi)$.*

The normalization coefficient Z_β of the model is given by:

Lemma 1. *For $n = 1$, $Z_\beta = 1$. Otherwise, for $n > 1$,*

$$Z_\beta = \prod_{i=2}^{n} \frac{\exp\left(-\frac{\beta n}{2} \frac{i}{i-1}\right) - 1}{\exp\left(-\frac{\beta n}{2} \frac{1}{i-1}\right) - 1}.$$

5 Reconstructing the Center Permutation in the AP model

In this section we study the problem of reconstructing the hidden center permutation given a set of samples drawn from the AP model.

Problem 1 (Permutation Reconstruction in the AP model)**.** *Given T i.i.d. samples obtained from $\mathcal{M}_{\mathcal{AP}}(\beta, \pi)$ with unknown parameters $\beta > 0$ and $\pi \in \mathcal{S}_n$, compute a best estimate for the center permutation π.*

A minimum requirement for solving the problem is $\beta > 0$, since for $\beta = 0$, the distribution is uniform over S_n and all centers define the same distribution.

Our first goal is to show that with a sufficient number of samples we can reconstruct the center permutation for any $\beta > 0$, even for $\beta = \beta(n) \overset{n\to\infty}{\longrightarrow} 0$. To achieve this goal we design a rank-based algorithm that for any set of samples outputs a permutation that is an unbiased estimate of the central permutation. As we increase the number of samples, the sequence of estimates converges; thus the estimate must be consistent and therefore converges to the central permutation.

For a more practical result we present a comparison-based algorithm that for any $\beta > \frac{2\ln(2)}{n}$, and given $O(\log(n))$ independent samples from the distribution, computes w.h.p.[1] the correct center permutation.

5.1 Rank-based algorithm for $\beta > 0$

Theorem 3 (Probability of a swap between elements adjacent in π)**.** *For any $\beta \geq 0$ and $1 \leq i \leq n-1$,*

$$\Pr\left(\sigma(\pi_{i+1}) < \sigma(\pi_i)\right) \leq 1/2$$

with equality iff $\beta = 0$.

Note that the above result applies only for adjacent items in the permutation π, and only with probability $\leq 1/2$. Thus, the result does not imply a total order. Recall that in general, $Pr(X \geq Y) > 1/2$ does not imply $E[X] \geq E[Y]$ or vice versa. Nevertheless, in this model we can prove a total order on the expected position of items in the observed permutations. In the following, let $\sigma(\pi_i)$ denote the position of the element π_i in a permutation $\sigma \sim M_{AP}(\pi, \beta)$. Note the distinction with σ_i which denotes the element at the i-th position in σ instead.

Theorem 4. *For any $\beta > 0$ and $1 \leq i \leq n-1$,*

$$E[\sigma(\pi_i)] < E[\sigma(\pi_{i+1})].$$

We will now describe a rank-based algorithm for the permutation reconstruction problem which takes as input a sample $\sigma^1, \ldots, \sigma^s$ of s independent random permutations from $\mathcal{M}_{\mathcal{AP}}(\beta, \pi)$ and computes an estimate π_{Rank} for the central permutation π.

For each element $\pi_1, \ldots, \pi_n$ we compute the respective average rank $\overline{\sigma(\pi_i)} = \frac{1}{s}\sum_{j=1}^{s} \sigma^j(\pi_i)$. We then build the permutation π_{Rank} by ordering the elements π_i according to their average ranks $\overline{\sigma(\pi_i)}$. The running time of the proposed algorithm is $O(sn + n\log n)$ where $O(sn)$ time is needed to computed the average rank for each element, and $O(n \log n)$ time is required by the sorting algorithm. Since the average ranks converge to their expectations we have:

[1] We say that an event happens with high probability (w.h.p.) if it happens with probability $\geq 1 - \frac{1}{n}$.

Theorem 5. *The ranking π_{Rank} is an asymptotically unbiased and consistent estimator.*

Since the ranks are integers we can conclude that:

Corollary 1. *For any value of $\beta > 0$, and with a sufficient number of samples, the reconstructed permutation π_{Rank} equals the central ranking π .*

5.2 Comparison-Based algorithm for $\beta \in \Omega(\frac{1}{n})$

We now present an efficient algorithm for $\beta > \frac{c\ln(2)}{n}$, $c > 2$ independent of n, which relies on the following result:

Theorem 6. *Let σ be obtained from $\mathcal{M}_{\mathcal{AP}}(\beta, \pi)$, with $\pi \in \mathcal{S}_n$ and $\beta > 0$. Given a pair (i, k) with $1 \leq i < n$ and $0 \leq k < n - i$, let $\Pr(S_{i,i+k+1})$ be the probability of two elements occupying the positions i and $i + k + 1$ in π are swapped in σ.*

$$\Pr\left(S_{i,i+k+1}\right) \leq \exp\left(-\frac{\beta n}{2}\right)\exp\left(-\frac{\beta n}{2}\frac{i}{i+k}\right).$$

For our choice of β, for any pair (i, k), $\Pr(S_{i,i+k+1}) \leq q < 1/2$.

With $O(\lceil \log n/(\beta n)\rceil)$ samples we can run any $O(n \log n)$ sorting algorithm, repeating every comparisons $O(\lceil \log n/(\beta n)\rceil)$ times to guarantee that with probability $1 - O(1/n)$ all pairs are placed in the right order. The total running time of this algorithm is $O(n \log n \log n/(\beta n))$. Note that with a careful analysis of the dependencies between the comparisons one can apply the noisy sorting algorithm in (Feige et al. 1994) to obtain an $O(n \log n)$ algorithm.

6 Experimental evaluation

We used both synthetic and real-world data to evaluate our algorithms and to explore the use of the AP distance as an alternative to the standard Kendall's tau distance in learning a true order from noisy sample data. We used two real-world datasets. First, we obtained a new dataset with long permutations and ground truth information, by automatically ranking Wikipedia pages with multiple ground-truth classes (a similar approach was used in (Pilehvar, Jurgens, and Navigli 2013) for linguistic graphs). We believe that our practical and efficient approach for generating classified permutations datasets will be useful in other studies. Second, we used publicly available data from human cancer studies with permutations with binary classifications.

We implemented the algorithms in C++. Each run used a single core and less than 4GB of RAM.[2]

6.1 Reconstructing the Hidden Permutation

In this experiment, we generated, using the algorithm defined in Section 4, a set P of i.i.d. permutations of size $n = 100$ from the $\mathcal{M}_{\mathcal{AP}}(\beta, \pi)$ model with different settings of β and size of $|P|$.[3] We obtained an estimate π^* of

[2] All our algorithms are easily parallelizable, if needed, but we did not pursue this direction.

[3] W.l.o.g., we set π to the identity permutation $(1, \ldots, 100)$.

the central permutation π by applying our estimation algorithm to the set P. We then tested the quality of the estimate π^* using two measures: Correctness (Corr.), the fraction of correctly ranked pairs $i, j \in [n]$ s.t. $\pi^*(\pi_i) < \pi^*(\pi_j)$ for $i < j$; Precision at 10 (P. at 10), the fraction of elements in the first 10 positions of π^* that are in $\{\pi_1, ..\pi_{10}\}$. Both of the measures range from 0 (π^* is the inverse of π) to 1 ($\pi^* = \pi$).

The results are shown in Figure 1(a) for the algorithm in Section 5.2 and are consistent with our theoretical analysis (we report averages over 300 runs of the experiment). Results for the Rank-based algorithm in Section 5.1 are very close and thus omitted. We observe that for both algorithms fewer samples are required to correctly reconstruct π as β grows, as expected. Notice that the algorithms achieve high precision even with few samples for $\beta > \frac{1}{n}$. For the more challenging settings (e.g., $\beta = 0.01$) we observe fairly high precision with more samples. Notice also the high precision in the reconstruction of the first (and most important) positions of the central ranking even in the higher variance experiment. This confirms our theoretical insight discussed in Section 5 according to which the AP model has lower variance then the Mallows model using Kendall's tau distance in the first positions of the ranking.

6.2 Experiments with real-data

Clustering We now assess the ability of a simple unsupervised clustering algorithm based on k-means, and equipped with our AP model estimators, to correctly cluster permutations from real-data for which we possess ground truth labeling. We address this problem in the context of web-pages rankings. In this context ranking correctly web-pages as well classifying pages in ground-truth categories is key to provide high quality recommendations.

Numerous algorithms, like the PageRank, are known to induce rankings over web-pages. In this experiment we use the well known Personalized PageRank (Haveliwala 2002) random walk with different seed pages to produce random permutations with ground-truth classes. Note that the AP model finds a natural application in this context as the first results in the permutations are more meaningful than the last ones.

More precisely, we used a snapshot of the entire English Wikipedia graph (Boldi and Vigna 2004) with approximately 4.2 million nodes and 101 million edges; where nodes represent pages and edges represent hyperlinks. Each page in our dataset is classified with ground-truth categories (≈ 30 thousand classes) in the `YAGO2` (Hoffart et al. 2011) knowledge-base taxonomy. We select a set C of `YAGO2` categories. Then for, each category $c \in C$, we create t rankings of nodes with ground-truth category c as follows: for t times, we select a page u in category c uniformly at random (u.a.r.), then we compute the the Personalized PageRank (PPR) ranking using u as seed of the walk. We set $\alpha = 0.15$ as the jump-back probability in all the walks. Notice that each of these rankings is biased towards the origin node and hence, toward the nodes belonging to category c of the seed node. This produces a set P of $|P| = t|C|$ permutations over nodes in the graph with ground-truth classes.[4]

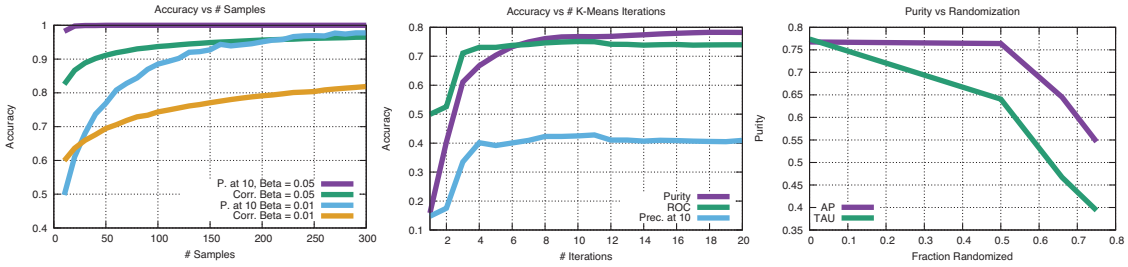
Given rankings P as input (without their ground-truth category) our aim is to cluster them as homogeneously as possibly by category, and to derive a representative ranking from each cluster. To do so we apply a heuristic based on the k-means algorithm as follows. Given the permutations P and a value k, we first assign the permutations to k clusters C_i for $i \in [k]$ u.a.r.. Then, we apply the following two procedures in order for a certain number of iterations (or until convergence). **Step 1:** For each C_i we compute the center permutation $\bar{\pi}^i$ using the comparison-based algorithm in Section 5.2 over the permutations in C_i. **Step 2:** For each permutation $\sigma^j \in P$ we assign σ^j to the cluster C_i whose center $\bar{\pi}^i$ minimizes the AP cost function $d_{AP}(\bar{\pi}^i, \sigma^j)$. Finally, the output of the algorithm is a partition of the rankings in k clusters as well as a center permutation $\bar{\pi}^i$ for each cluster.

For our quantitative analysis, we selected $t = 200$ u.a.r. nodes for each of the eight well-represented categories in the Wikipedia taxonomy.[5] In this experiment $|P| = 1600$ and each permutation has $n \approx 26{,}000$ elements. To evaluate the obtained clusters, we used the purity score– i.e., the fraction of correctly classified rankings if each ranking in cluster C_i is assigned to a majority ground-truth class in C_i (purity 1 is obtained only for the perfect partitioning). Figure 1(b) shows the evolution of purity over the iterations setting $k = 8$ and averaging over 20 runs. Notice that purity converges to a high level (close to 80% percent) in a few iterations, showing that the obtained clusters are indeed close to the ground truth classes. We then evaluate the correctness of the obtained $k = 8$ centers' rankings. To do so we use the standard ROC and Precision at 10 measure obtained by considering each position of the central ranking of a cluster as correct if it belongs to a node of the majority class of that cluster. The results are shown in Figure 1(b). Notice that the algorithm converges quickly to good quality central rankings (ROC significantly higher than 50%, precision close to 40%).

Finally we tested the ability of our algorithms (and of the AP distance) to overcome noise in the lowest position of the permutations — which are more noisy in practice (Vigna 2015). Consider a ranking σ in P and let c be the category to which it is assigned. We randomized the lowest γn positions, for $\gamma \in (0, 1)$, as follows: we kept the first $(1 - \gamma)n$ positions intact and then we added the remaining elements in the order of a u.a.r. ranking chosen from a *different* category. This randomization was applied to every ranking. As γ grows, the described randomization increases the amount of noise in the lower part of the rankings making them less related to the original category (and more biased towards a random different category). We then applied the k-means algorithm as before. In this experiment we compared also the result obtained by the same algorithm using the Kendall's

[4] We restrict these rankings to contain only the nodes associated with some category $c \in C$.

[5] We used scientific fields (computer scientists, neuroscientists, mathematicians), art-related (American actors, singers, English writers) and other categories (American politicians, Olympic athletes). For this test we discarded from the sample nodes belonging to more than one category in the list.



(a) Accuracy on synthetic data

(b) Accuracy vs. # Iterations in K-means with Real Data

(c) Purity vs. Level of Randomization in K-means with Real Data

Figure 1: Experimental Results for K-Means

tau distance instead of the AP distance (in Step 2) to assign permutations to centers (all the rest left equal). Figure 1(c) shows the purity reached after 10 iterations of k-means using both measures depending on the amount of randomization. We noticed that the results without randomization are qualitatively similar using the two measures. However, while the AP-based algorithm is almost unaffected even when the lowest 50% of rankings are randomized, the Kendall's tau one degrades its performance quickly. The AP maintains good purity in the clusters found even for a 75% randomization of the rankings. This confirms experimentally the intuition that the AP distance is less affected by the noise occurring in the lowest positions with respect to the Kendall's tau distance.

Dataset	Prec. Tau	Prec. AP
BC1	0.662	**0.674**
BC2	**0.621**	0.601
CT	0.848	**0.868**
LA1	0.666	**0.685**
LC2	0.986	**0.993**
MB	0.613	**0.648**
OV	**0.836**	0.817
PC1	0.657	**0.667**
PC2	**0.499**	0.493
Average	0.709	**0.716**

Table 1: Average precision of the classification algorithm in the human cancer dataset. The best result for each dataset is highlighted in bold.

Classification We now turn our attention to a supervised learning algorithm for classification based on our model. We used 10 publicly available datasets from human cancer research. From each dataset, we have a series of real-valued feature vectors measuring gene expressions or proteomic spectra — see (Jiao and Vert 2015) for details on the dataset. Each feature vector is assigned to one of two binary classes (e.g., "Normal vs. Tumor", "Non-relapse vs. Relapse").

It has been observed that in the context of high-dimensional gene expressions, the relative order of the features is more important than their absolute magnitude (Jiao and Vert 2015; Geman et al. 2004), since the relative order is less affected by pre-processing designed to overcome technological biases. Hence, as in (Jiao and Vert 2015), for each vector in our dataset we obtain a permutation over the set of the features by sorting them in decreasing order of value (ties broken randomly). We split each dataset in training and testing sets using 5-fold cross-validation.

We applied the following simple classification algorithm. For each class we reconstructed the center permutation using our comparison-based algorithm (in Section 5.2) on the set of permutations belonging to that class in the training set. We then classified each permutation in the testing set with the class of the center permutation having minimum distance to that permutation. We used as distance either AP or Kendall's tau. The results are reported in Table 1 where we show the average classification precision over 10 independent runs of the 5-fold cross-validation (a total of 50 independent test/train experiments). The results show that the AP distance improves over the Kendall's tau distance in most datasets, and the average precision improves as well.

These experiments confirm our original motivation for the study of the AP model suggesting that items in the beginning of the permutations are likely to have higher importance and less noise in both supervised and unsupervised learning tasks.

7 Conclusion

In this work we have introduced a novel stochastic permutation model based on a more nuanced permutation statistic measure that is widely used in practice. Despite its sophistication is still amenable to formal analysis. This allows us to define provably correct algorithms to learn the parameters of this model efficiently.

We believe that this is a first step towards defining even more sophisticated (and arguably more realistic) probabilistic models of rankings for which many of the results achieved in the traditional Mallows model literature could be extended.

As a future work we would like to define models that allow both ties in the rankings—which are very frequent in many practical applications (Vigna 2015)—and more general cost functions as those recently defined in (Vigna 2015).

Acknowledgments Supported in part by NSF award IIS-1247581 and by the University of Padova under project CPDA121378/12.

References

Awasthi, P.; Blum, A.; Sheffet, O.; and Vijayaraghavan, A. 2014. Learning mixtures of ranking models. In *Advances in Neural Information Processing Systems*, 2609–2617.

Bartholdi III, J.; Tovey, C. A.; and Trick, M. A. 1989. Voting schemes for which it can be difficult to tell who won the election. *Social Choice and welfare* 6(2):157–165.

Bhatnagar, N., and Peled, R. 2014. Lengths of monotone subsequences in a Mallows permutation. *Probability Theory and Related Fields* 1–62.

Boldi, P., and Vigna, S. 2004. The webgraph framework i: compression techniques. In *Proceedings of the 13th conference on World Wide Web*, 595–602. ACM.

Braverman, M., and Mossel, E. 2008. Noisy sorting without resampling. In *Proceedings of the nineteenth ACM-SIAM symposium on Discrete algorithms*, 268–276. Society for Industrial and Applied Mathematics.

Braverman, M., and Mossel, E. 2009. Sorting from noisy information. *arXiv preprint*.

Brian, E. 2008. Condorcet and Borda in 1784. Misfits and documents. *Journal Electronique d'Histoire des Probabilts et de la Statistique* 4(1).

Chierichetti, F.; Dasgupta, A.; Kumar, R.; and Lattanzi, S. 2014. On reconstructing a hidden permutation. *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques* 28:604–617.

De Stefani, L.; Epasto, A.; Upfal, E.; and Vandin, F. 2015. Reconstructing Hidden Permutations Using the Average-Precision (AP) Correlation Statistic. Extended version, available from http://bigdata.cs.brown.edu/AAAI16DeStefani.pdf.

Diaconis, P., and Ram, A. 2000. *Analysis of systematic scan Metropolis algorithms using Iwahori-Hecke algebra techniques*. Department of Statistics, Stanford University.

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

Fagin, R.; Kumar, R.; and Sivakumar, D. 2003. Comparing top k lists. *SIAM Journal on Discrete Mathematics* 17(1):134–160.

Feige, U.; Raghavan, P.; Peleg, D.; and Upfal, E. 1994. Computing with noisy information. *SIAM Journal on Computing* 23(5):1001–1018.

Fligner, M. A., and Verducci, J. S. 1986. Distance Based Ranking Models. *Journal of the Royal Statistical Society. Series B (Methodological)* 48(3).

Fligner, M. A., and Verducci, J. S. 1988. Multistage ranking models. *Journal of the American Statistical Association* 83(403).

Geman, D.; d'Avignon, C.; Naiman, D. Q.; and Winslow, R. L. 2004. Classifying gene expression profiles from pairwise mRNA comparisons. *Statistical applications in genetics and molecular biology*.

Haveliwala, T. H.; Gionis, A.; Klein, D.; and Indyk, P. 2002. Evaluating strategies for similarity search on the web. In *Proceedings of the 11th conference on World Wide Web*, 432–442. ACM.

Haveliwala, T. H. 2002. Topic-sensitive pagerank. In *Proceedings of the 11th conference on World Wide Web*, 517–526. ACM.

Hoffart, J.; Suchanek, F. M.; Berberich, K.; Lewis-Kelham, E.; de Melo, G.; and Weikum, G. 2011. YAGO2: Exploring and querying world knowledge in time, space, context, and many languages. In *Proceedings of the 20th conference companion on World wide web*, 229–232. ACM.

Jiao, Y., and Vert, J.-P. 2015. The Kendall and Mallows kernels for permutations. In *Proceedings of the 32nd International Conference on Machine Learning (ICML-15)*, 1935–1944.

Kenyon-Mathieu, C., and Schudy, W. 2007. How to rank with few errors. In *Proceedings of the thirty-ninth ACM symposium on Theory of computing*, 95–103. ACM.

Kumar, R., and Vassilvitskii, S. 2010. Generalized distances between rankings. In *Proceedings of the 19th conference on World wide web*, 571–580. ACM.

Lu, T., and Boutilier, C. 2011. Learning Mallows models with pairwise preferences. In *Proceedings of the 28th International Conference on Machine Learning*, 145–152.

Mallows, C. L. 1957. Non-null ranking models. i. *Biometrika* 44(1/2):pp. 114–130.

Meek, C., and Meila, M. 2014. Recursive inversion models for permutations. In *Advances in Neural Information Processing Systems*, 631–639.

Meila, M.; Phadnis, K.; Patterson, A.; and Bilmes, J. A. 2012. Consensus ranking under the exponential model. *arXiv preprint*.

Melucci, M. 2007. On rank correlation in information retrieval evaluation. In *ACM SIGIR Forum*, volume 41, 18–33. ACM.

Pilehvar, M. T.; Jurgens, D.; and Navigli, R. 2013. Align, disambiguate and walk: A unified approach for measuring semantic similarity. In *ACL (1)*, 1341–1351.

Qin, T.; Geng, X.; and Liu, T.-Y. 2010. A new probabilistic model for rank aggregation. In *Advances in neural information processing systems*, 1948–1956.

Saari, D. G. 2006. Which is better: the Condorcet or Borda winner? *Social Choice and Welfare* 26(1):107–129.

Sakai, T. 2012. Evaluation with informational and navigational intents. In *Proceedings of the 21st conference on World Wide Web*, 499–508. ACM.

Schapire, W. W. C. R. E., and Singer, Y. 1998. Learning to order things. In *Advances in Neural Information Processing Systems*, volume 10, 451. MIT Press.

Schubert, E.; Wojdanowski, R.; Zimek, A.; and Kriegel, H. 2012. On Evaluation of outlier rankings and outlier scores. In *SDM*, 1047–1058.

Shieh, G. S. 1998. A weighted Kendall's tau statistic. *Statistics & probability letters* 39(1):17–24.

Sorz, J.; Fieder, M.; Wallner, B.; and Seidler, H. 2015. High statistical noise limits conclusiveness of ranking results as a benchmarking tool for university management. *PeerJ PrePrints* 3:e1162.

Spearman, C. 1904. The proof and measurement of association between two things. *American Journal of Psychology* 15:88–103.

Vigna, S. 2015. A weighted correlation index for rankings with ties. In *Proceedings of the 24th Conference on World Wide Web*, 1166–1176.

Yilmaz, E., and Aslam, J. A. 2006. Estimating average precision with incomplete and imperfect judgments. In *Proceedings of the 15th ACM conference on Information and knowledge management*, 102–111. ACM.

Yilmaz, E.; Aslam, J. A.; and Robertson, S. 2008. A new rank correlation coefficient for information retrieval. In *Proceedings of the 31st annual ACM SIGIR conference*, 587–594. ACM.