



Fewer Axioms for a More Flexible Distance between Rankings

Farzad Farnoud
a joint work with with O. Milenkovic
and Touri, Su, Raisali.

Many Applications of Rankings...

We often encounter rankings of:

- politicians, celebrities, performers, job candidates
- schools, teams in professional sports
- movies, products
- emotions, pain levels, quality of drug treatments, ...

and use ranking theory in:

- Computer science (search engines, etc).
- Recommender systems, marketing.
- General social sciences: competitions, voting.
- Management and decision making.

Rank Aggregation

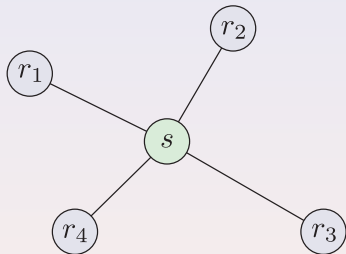
Rank Aggregation: Combining a set of rankings such that the result is a ranking “representative” of the set.

Expert 1	Expert 2	Expert 3	Aggregate
Caltech	UIUC	UCB	?
UIUC	UCB	UIUC	?
Stanford	Caltech	MIT	?
MIT	MIT	Stanford	?
UCB	Stanford	Caltech	?

Mathematically, rankings are represented by permutations, i.e., **arrangements** of a set of objects.

Example: (b, c, a) – a permutation of the set $\{a, b, c\}$

Distance-Based Rank Aggregation



Given expert rankings $\sigma_1, \sigma_2, \dots, \sigma_m$, the rank aggregation problem can be stated as

$$\pi^* = \arg \min_{\pi} \sum_{i=1}^m d(\pi, \sigma_i).$$

Equivalently, want the median of permutations.

But how do we choose the distance?

Rank aggregation requires a **distance function** on the space of permutations

- Kemeny 59 **Kemeny's axiomatic approach to determine appropriate distance function** – Kendall τ .
- Dwork 01 Finding Kemeny aggregate is NP-hard, bipartite matching and Markov chain methods for aggregation [Dwork et al].
- Sculley 07 Aggregation with similarity score [Sculley et.al.].
- Kumar 10 Generalizing Kendall τ and Spearman's footrule [Kumar et al].

Kemeny's Axioms

Kemeny's axiomatic approach for determining a distance function:

- 1 $d(\cdot, \cdot)$ is a metric.
- 2 Relabeling of objects does not change the distance.
- 3 $d(\pi, \sigma) = d(\pi, \omega) + d(\omega, \sigma)$ iff ω is “between” π and σ .
Betweenness: for $a, b \in [n]$, if π and σ both rank a before b , then ω also ranks a before b
- 4 If two rankings agree except on a “segment,” position of segment within ranking is not important:
 $d(abcd e, abdc e) = d(cdabe, dcabe)$.



“What’s in a name? That which we call a rose, by any other name would smell as sweet.”

The unique distance that satisfies Kemeny's axioms is Kendall τ

Kendall τ = minimum number of swaps of adjacent elements needed to transform one into the other = number of disagreements between two rankings.

A swap of two elements is called a **transposition**. Transposition of elements in positions i and j is denoted by (ij)

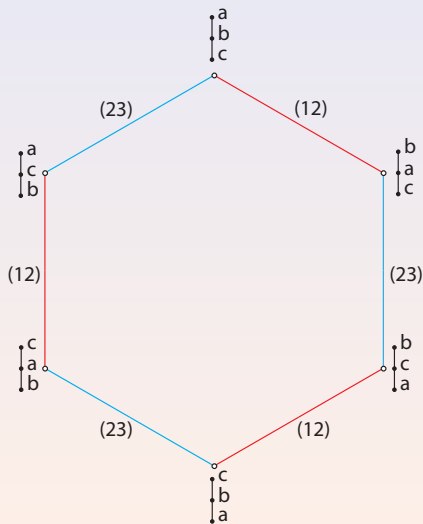
Example: $K(abcde, cabde) = 2 : abcde \xrightarrow{(23)} acbde \xrightarrow{(12)} cabde$

Kendall τ

Kendall τ can be represented by a graph with $n!$ vertices.

Neighboring vertices differ by an adjacent transposition.

Distance is the length of the shortest path.



Kemeny's method is the only rule that is [Young, Levenglick, 1978]:

- **Consistent:** If two committees meeting separately arrive at the same ranking, their joint meeting will still give the same ranking.
- **Condorcet:** If a candidate exists that wins against all other in pairwise comparison, that candidate will be ranked first.
- **Neutral:** Treats all candidates the same.

Rank Aggregation: When Kendall Is Not Suitable

- Kendall τ treats all positions in a ranking similarly
- For voters, top portion of rankings may be more important than the bottom
- A voter with vote σ is likely to prefer π_1 to π_2
- But: $K(\sigma, \pi_1) = K(\sigma, \pi_2)$

σ	π_1	π_2
Melbourne	Melbourne	Vienna
Vienna	Vienna	Melbourne
Vancouver	Vancouver	Vancouver
Toronto	Toronto	Toronto
Calgary	Calgary	Calgary
Adelaide	Adelaide	Adelaide
Sydney	Sydney	Sydney
Helsinki	Perth	Helsinki
Perth	Helsinki	Perth
Auckland	Auckland	Auckland

Rank Aggregation: When Kendall Is Not Suitable

Click-through rate of a link: ratio of number of clicks to the number of displays

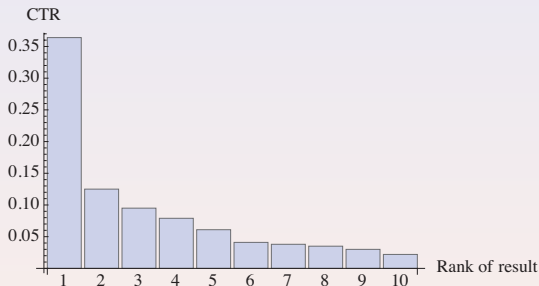


Figure : Click-through rates for 1st page of Google search results

In aggregating search results, **top of the ranking is more important**

Generalizing the Kendall Distance

How should the axioms be changed?

- Let us **remove the fourth axiom**

- 1 Distance function is a pseudo-metric
- 2 Relabeling of objects does not change distance.
- 3 $d(\sigma, \pi) = d(\pi, \omega) + d(\omega, \sigma)$ iff ω is between π and σ
- 4 ~~If two rankings agree except on a "segment," position of segment within ranking is not important:
 $d(\underline{abc}de, \underline{abd}ce) = d(\underline{cd}abc, \underline{dc}abc)$.~~

Generalizing the Kendall Distance

How should the axioms be changed?

- Let us **remove the fourth axiom**

- 1 Distance function is a pseudo-metric
- 2 Relabeling of objects does not change distance.
- 3 $d(\sigma, \pi) = d(\pi, \omega) + d(\omega, \sigma)$ iff ω is between π and σ
- 4 ~~If two rankings agree except on a "segment," position of segment within ranking is not important:
 $d(\underline{abcde}, \underline{abdce}) = d(\underline{cdabc}, \underline{dcabc})$.~~

The solution is again **Kendall τ** !

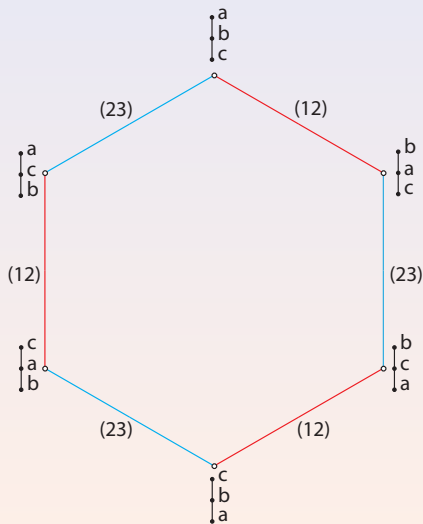
Removing the fourth axiom is **not sufficient**.

Why Modify the Third Axiom?

Lemma [F, Touri, Milenkovic]:
For complete rankings, **fourth axiom** follows from the third axiom.

Special case: $n = 3$

Consider the distinct paths between abc and cba .



Generalizing the Kendall Distance

Our relaxation of Kemeny's axioms:

- 1 Distance function is a pseudo-metric
- 2 Relabeling of objects does not change distance.
- 3 $d(\sigma, \pi) = d(\pi, \omega) + d(\omega, \sigma)$ iff ω is "between" π and σ for some ω between π and σ if π and σ disagree on more than one pair of elements.
- 4 ~~If two rankings agree except on a "segment," position of segment within ranking is not important:~~
 $d(\underline{abcde}, \underline{abdce}) = d(\underline{cdabe}, \underline{dcabe})$.

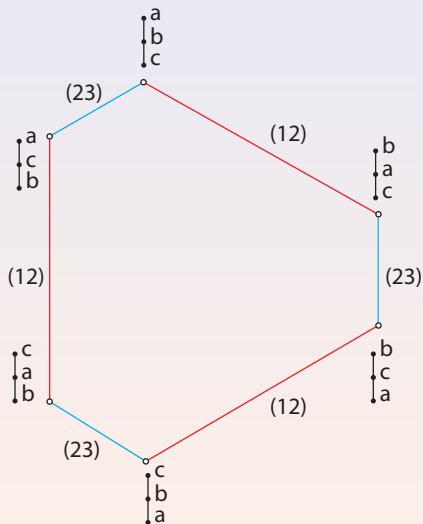
Unique solution: weighted Kendall τ [F, Touri, Milenkovic, 2012]

New Distance: Weighted Kendall Distance

Weighted Kendall distance:
minimum weight of transforming one permutation into the other using **adjacent transpositions** where each adjacent transposition has a given **weight**.

Weight of transposition (ij) is denoted by $\varphi(i, j)$.

$$d(abc, cba) = 2\varphi(2, 3) + \varphi(1, 2).$$



Decreasing Weight Functions

- Weighted Kendall distance between σ and $\pi_1 =$
 $d(\sigma, \pi_1) = \varphi(8, 9)$
- Weighted Kendall distance between σ and $\pi_2 =$
 $d(\sigma, \pi_2) = \varphi(1, 2)$
- If we choose $\varphi(i, i + 1)$ to be decreasing in i , then
 $d(\sigma, \pi_1) < d(\sigma, \pi_2) \Rightarrow$
decreasing weight function.

σ	π_1	π_2
Melbourne	Melbourne	Vienna
Vienna	Vienna	Melbourne
Vancouver	Vancouver	Vancouver
Toronto	Toronto	Toronto
Calgary	Calgary	Calgary
Adelaide	Adelaide	Adelaide
Sydney	Sydney	Sydney
Helsinki	Perth	Helsinki
Perth	Helsinki	Perth
Auckland	Auckland	Auckland

Weighted Kendall Distance: Example I

Consider the set of rankings

$$\Sigma = \left(\begin{array}{cccc} 1 & 4 & 2 & 3 \\ 1 & 4 & 3 & 2 \\ 2 & 3 & 1 & 4 \\ 4 & 2 & 3 & 1 \\ 3 & 2 & 4 & 1 \end{array} \right).$$

The Kemeny aggregate is $(4, 2, 3, 1)$.

The optimum aggregate ranking for the weight function φ with $\varphi(i, i+1) = (2/3)^{i-1}$, $i \in [4]$, equals $(1, 4, 2, 3)$.

A candidate with both strong showings and weak showings beats a candidate with a rather average performance.

Weighted Kendall Distance: Example II

Consider the set of rankings

$$\Sigma = \left(\begin{array}{c|c|c} 1 & 2 & 3 \\ \hline 1 & 2 & 3 \\ \hline 3 & 2 & 1 \\ \hline 2 & 1 & 3 \end{array} \right).$$

The Kemeny aggregates are $(1, 2, 3)$, $(2, 1, 3)$.

Weighted Kendall can be used to pick a unique solution: for any strictly decreasing weight function the solution is unique, namely, $(1, 2, 3)$.

Computing Weighted Kendall Distance

Computing Kendall τ is straightforward: count the number of disagreements.

How to compute the weighted Kendall distance for general weight functions is not known, but is known for a very important case:

Monotonic weight function: φ is monotonic if $\varphi(i, i + 1)$ is monotonic in i .

Theorem [F, Touri, Milenkovic]: Weighted Kendall distance with **monotonic weight** can be computed in time $O(n^4)$.

Theorem [F, Milenkovic]: 2-approximation for weighted Kendall distance with **general weights** can be computed in time $O(n^2)$.

Weighted Transposition Distance

Instead of allowing only adjacent transpositions, we can **allow all transpositions**

To each transposition (ij) assign weight $\varphi(i,j)$.

Weighted Transposition Distance: Minimum weight of a sequence of transpositions that transform one permutation to another.

Appropriate for **modeling similarity among candidates:**

$\varphi(\text{Godfather I, Godfather II})$

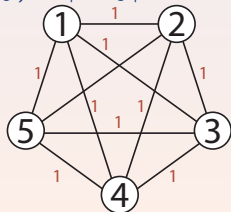
$< \varphi(\text{Godfather I, Goodfellas})$

$< \varphi(\text{Godfather I, Star Wars})$

Common Distance Functions

Several distance functions used for rank aggregation [Diaconis and Graham 88] are **special cases** of the weighted transposition distance:

- **Kendall's τ** : $K(\pi, \sigma) = \#$ of transpositions of adjacent ranks.
Equivalent to $\varphi_K(i, i+1) = 1$.
- **Spearman's Footrule**: $F(\pi, \sigma) = \sum_i |\pi^{-1}(i) - \sigma^{-1}(i)|$.
Equivalent to the path weight function $\varphi_F(i, j) = |i - j|$.
- **Cayley's distance**:
 $T(\pi, \sigma) = \#$ of transpositions
Equivalent to $\varphi_T(i, j) = 1$.



Weighted Transposition Distance: Example

Consider the votes listed in Σ ,

$$\Sigma = \left(\begin{array}{cccc} 1 & 2 & 3 & 4 \\ \hline 3 & 2 & 1 & 4 \\ \hline 4 & 1 & 3 & 2 \end{array} \right).$$

Even and odd numbers represent different types of candidates:

$$\varphi(i,j) = \begin{cases} 1, & \text{if } i,j \text{ are both odd or both even,} \\ 2, & \text{else.} \end{cases}$$

Votes are “diverse” : they alternate between odd and even numbers.

Kemeny aggregate is $(1, 3, 2, 4)$: odd numbers ahead even numbers.

Aggregation using φ gives $(1, 2, 3, 4)$.

Computing Weighted Transposition Distance

What is the distance of a single transposition from the identity?

Example: Distance of (red yellow) to identity

Find a path such that **two copies of the path minus its heaviest edge** has minimum weight

Computing Weighted Transposition Distance

- **4-approximation algorithm** for arbitrary weight functions in $O(n^4)$ operations
- **2-approximation algorithm** if weight function is a metric, in $O(n^4)$ operations
- **2-approximation algorithm** for path weight functions (e.g. weighted Kendall) in $O(n^4)$ operations
- **Exact algorithms** for metric-path weight functions (e.g. weighted Spearman's Footrule) in $O(n^2)$ operations.

See F.Farnoud and O.Milenkovic, "Sorting of permutations by cost-constrained transpositions," IT Transaction, 58(1):3–23, Jan. 2012.

Rank Aggregation

Recall: Given voter rankings $\sigma_1, \sigma_2, \dots, \sigma_m$, the rank aggregation problem can be stated as

$$\pi^* = \arg \min_{\pi} \sum_{i=1}^m d_{\varphi}(\pi, \sigma_i).$$

For many distance functions, problem is **NP-hard**.

Alternative ways to find reasonable solutions:

- **Approximation:** 2-approximation or 4-approximation (depending on the properties of φ) [Dwork et al. 2001] + **local search**
- Linear programming relaxation [Conitzer et al. 2006]
- **Heuristic Markov chain** methods developed in the spirit of **PageRank** [Dwork et al. 2001]

Rank Aggregation: Approximation

For **general weight function** φ , to find

$$\pi^* = \arg \min_{\pi} \sum_{i=1}^m d_{\varphi}(\pi, \sigma_i)$$

we approximate d_{φ} by $D = \sum_i f(\pi^{-1}(i), \sigma^{-1}(i))$ such that

$$(1/2)D(\pi, \sigma) \leq d_{\varphi}(\pi, \sigma) \leq 2D(\pi, \sigma).$$

Rank Aggregation: Approximation

For **general weight function** φ , to find

$$\pi^* = \arg \min_{\pi} \sum_{i=1}^m d_{\varphi}(\pi, \sigma_i)$$

we approximate d_{φ} by $D = \sum_i f(\pi^{-1}(i), \sigma^{-1}(i))$ such that

$$(1/2)D(\pi, \sigma) \leq d_{\varphi}(\pi, \sigma) \leq 2D(\pi, \sigma).$$

Using **perfect min weight bipartite matching algorithms**, can find

$$\pi' = \arg \min_{\pi} \sum_{i=1}^m D(\pi, \sigma_i)$$

exactly, and show that $\sum_{i=1}^m d_{\varphi}(\pi', \sigma_i) \leq 4 \sum_{i=1}^m d_{\varphi}(\pi^*, \sigma_i)$.

Rank Aggregation: Approximation

For **general weight function** φ , to find

$$\pi^* = \arg \min_{\pi} \sum_{i=1}^m d_{\varphi}(\pi, \sigma_i)$$

we approximate d_{φ} by $D = \sum_i f(\pi^{-1}(i), \sigma^{-1}(i))$ such that

$$(1/2)D(\pi, \sigma) \leq d_{\varphi}(\pi, \sigma) \leq 2D(\pi, \sigma).$$

Using **perfect min weight bipartite matching algorithms**, can find

$$\pi' = \arg \min_{\pi} \sum_{i=1}^m D(\pi, \sigma_i)$$

exactly, and show that $\sum_{i=1}^m d_{\varphi}(\pi', \sigma_i) \leq 4 \sum_{i=1}^m d_{\varphi}(\pi^*, \sigma_i)$.

Search for a **local optimum** starting from π' .

Rank Aggregation: Linear Programming Relaxation

Kendall τ distance = number of **disagreements**

c_{ij} is the **number of voters that prefer i to j**

π_{ij} equals 1 if the **aggregate π prefers i to j**

Aggregation problem as integer program [Conitzer et al, 2006]:

$$\text{minimize } \sum_{i,j} c_{ji} \pi_{ij}$$

$$\text{subject to } \pi_{ij} + \pi_{ji} = 1$$

$$\pi_{ij} + \pi_{jk} + \pi_{ki} \leq 2$$

$$\pi_{ij} \in \{0, 1\}$$

If we relax the condition $\pi_{ij} \in \{0, 1\}$ to $0 \leq \pi_{ij} \leq 1$, we have a **linear program**

Rank Aggregation: Linear Programming Relaxation

For weights that **decrease arithmetically**, we can do the same.
 π_{ijk} equals 1 if the **aggregate π** prefers **i to j** and **j to k** .
 α_{ijk} measures the **disagreement** of voters with ordering (ijk) .

Aggregation problem as integer program:

$$\begin{aligned} & \text{minimize } \sum_{i,j,k} \alpha_{ijk} \pi_{ijk} \\ & \text{subject to } \pi_{ijk} + \pi_{jik} + \dots + \pi_{kji} = 1 \\ & \quad \pi_{ijk} + \pi_{ikj} + \pi_{kij} = \pi_{ij} \\ & \quad \pi_{ijk} \in \{0, 1\} \end{aligned}$$

$$\alpha_{ijk} = \sum_{rst} \# \text{voters with preference } (rst) * d_{\varphi}(rst, ijk)$$

Again, removing integrality condition leads to a **linear program**.

Rank Aggregation: Markov Chain Methods

Based on ideas behind PageRank and work by Dwork et.al., 2001.

- Form a Markov chain with **nodes** indexed by candidates, and **transition probabilities** “determined” by voters.
- If a is preferred to b by large number of voters, the transition probability from a to b should be **high**.
- The **equilibrium distribution** reflects preference order of candidates.

How should a Markov chain approach be designed for **non-uniform weights**?

See F, Touri, Milenkovic, “Nonuniform Vote Aggregation Algorithms,” SPCOM 2012.

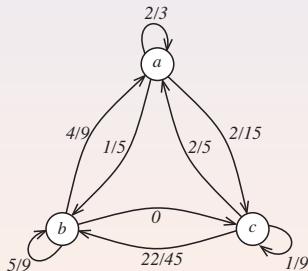
Rank Aggregation: Markov Chain Methods

The probability of going from i to j , where j is ranked higher, depends on sum of the weights of adjacent transpositions between the positions of a and b :

$$\beta_{ij}(\sigma) = \max_{l: j_{\sigma} \leq l < i_{\sigma}} \frac{\sum_{h=l}^{i_{\sigma}-1} \varphi(h, h+1)}{i_{\sigma} - l},$$

appropriately normalized.

For votes abc, abc, bca :



See F, Touri, Milenkovic, "Nonuniform Vote Aggregation Algorithms," SPCOM 2012.

Thank you!