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

<table>
<thead>
<tr>
<th>Expert 1</th>
<th>Expert 2</th>
<th>Expert 3</th>
<th>Aggregate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caltech</td>
<td>UIUC</td>
<td>UCB</td>
<td>?</td>
</tr>
<tr>
<td>UIUC</td>
<td>UCB</td>
<td>UIUC</td>
<td>?</td>
</tr>
<tr>
<td>Stanford</td>
<td>Caltech</td>
<td>MIT</td>
<td>?</td>
</tr>
<tr>
<td>MIT</td>
<td>MIT</td>
<td>Stanford</td>
<td>?</td>
</tr>
<tr>
<td>UCB</td>
<td>Stanford</td>
<td>Caltech</td>
<td>?</td>
</tr>
</tbody>
</table>

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

Example: \((b, c, a)\) – a permutation of the set \(\{a, b, c\}\)
Given expert rankings $\sigma_1, \sigma_2, \cdots, \sigma_m$, 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 $\tau$.

- **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 $\tau$ and Spearman’s footrule [Kumar et al].
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 \(\omega\) is “between” \(\pi\) and \(\sigma\). Betweenness: for \(a, b \in [n]\), if \(\pi\) and \(\sigma\) both rank \(a\) before \(b\), then \(\omega\) also ranks \(a\) before \(b\).
4. If two rankings agree except on a “segment,” position of segment within ranking is not important: \(d(abcded, abdce) = d(cdaab, dcabe)\).

---

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

The unique distance that satisfies Kemeny’s axioms is Kendall $\tau$

Kendall $\tau = \text{minimum number of swaps of adjacent elements needed to transform one into the other} = \text{number of disagreements between two rankings}$.

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

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

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.
Kendall $\tau$ treats all positions in a ranking similarly.

For voters, top portion of rankings may be more important than the bottom.

A voter with vote $\sigma$ is likely to prefer $\pi_1$ to $\pi_2$.

But: $K(\sigma, \pi_1) = K(\sigma, \pi_2)$
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 \( \omega \) is between \( \pi \) and \( \sigma \)
4. If two rankings agree except on a “segment,” position of segment within ranking is not important:
   \( d(abcd\,e,\,abdc\,e) = d(cd\,abe,\,dc\,abe) \).
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 \(\omega\) is between \(\pi\) and \(\sigma\)
4. If two rankings agree except on a “segment,” position of segment within ranking is not important: 
   \[d(abcde, abdce) = d(cdabe, dcabe).\]

The solution is again Kendall \(\tau\)!
Removing the fourth axiom is not sufficient.
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 \( \omega \) is “between” \( \pi \) and \( \sigma \) for some \( \omega \) between \( \pi \) and \( \sigma \) if \( \pi \) and \( \sigma \) 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:
\[
d(abcd e, abdce) = d(cdabe, dcabe).\]

Unique solution: weighted Kendall \( \tau \) [F, Touri, Milenkovic, 2012]
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 \(\varphi(i, j)\).

\[
\begin{align*}
d(abc, cba) &= 2\varphi(2, 3) + \varphi(1, 2).
\end{align*}
\]
Decreasing Weight Functions

- Weighted Kendall distance between $\sigma$ and $\pi_1 = d(\sigma, \pi_1) = \varphi(8, 9)$
- Weighted Kendall distance between $\sigma$ 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.

<table>
<thead>
<tr>
<th></th>
<th>$\sigma$</th>
<th>$\pi_1$</th>
<th>$\pi_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Melbourne</td>
<td>Melbourne</td>
<td>Vienna</td>
<td></td>
</tr>
<tr>
<td>Vienna</td>
<td>Vienna</td>
<td>Melbourne</td>
<td></td>
</tr>
<tr>
<td>Vancouver</td>
<td>Vancouver</td>
<td>Vancouver</td>
<td></td>
</tr>
<tr>
<td>Toronto</td>
<td>Toronto</td>
<td>Toronto</td>
<td></td>
</tr>
<tr>
<td>Calgary</td>
<td>Calgary</td>
<td>Calgary</td>
<td></td>
</tr>
<tr>
<td>Adelaide</td>
<td>Adelaide</td>
<td>Adelaide</td>
<td></td>
</tr>
<tr>
<td>Sydney</td>
<td>Sydney</td>
<td>Sydney</td>
<td></td>
</tr>
<tr>
<td>Helsinki</td>
<td>Perth</td>
<td>Helsinki</td>
<td></td>
</tr>
<tr>
<td>Perth</td>
<td>Helsinki</td>
<td>Perth</td>
<td></td>
</tr>
<tr>
<td>Auckland</td>
<td>Auckland</td>
<td>Auckland</td>
<td></td>
</tr>
</tbody>
</table>
Consider the set of rankings

\[
\Sigma = \begin{pmatrix}
1 & 4 & 2 & 3 \\
1 & 4 & 3 & 2 \\
2 & 3 & 1 & 4 \\
4 & 2 & 3 & 1 \\
3 & 2 & 4 & 1
\end{pmatrix}.
\]

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

The optimum aggregate ranking for the weight function \(\varphi\) 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.
Consider the set of rankings

\[ \Sigma = \begin{pmatrix}
1 & 2 & 3 \\
1 & 2 & 3 \\
3 & 2 & 1 \\
2 & 1 & 3
\end{pmatrix}. \]

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 Kendall $\tau$ 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: $\varphi$ 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 $O(n^2)$. 
Instead of allowing only adjacent transpositions, we can **allow all transpositions**

To each transposition \((i, j)\) 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:**

\[
\varphi(\text{Godfather I}, \text{Godfather II}) < \varphi(\text{Godfather I}, \text{Goodfellas}) < \varphi(\text{Godfather I}, \text{Star Wars})
\]
Several distance functions used for rank aggregation [Diaconis and Graham 88] are special cases of the weighted transposition distance:

- **Kendall's $\tau$:** $K(\pi, \sigma) = \#$ 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 $\varphi_T(i, j) = 1$. 

![Diagram of a weighted graph]
Weighted Transposition Distance: Example

Consider the votes listed in $\Sigma$,

$$
\Sigma = \begin{pmatrix}
1 & 2 & 3 & 4 \\
3 & 2 & 1 & 4 \\
4 & 1 & 3 & 2
\end{pmatrix}
$$

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 $\varphi$ gives $(1, 2, 3, 4)$. 
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.

Recall: Given voter rankings $\sigma_1, \sigma_2, \ldots, \sigma_m$, 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 $\varphi$) [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]
For general weight function \( \varphi \), to find

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

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

\[
(1/2)D(\pi, \sigma) \leq d_\varphi(\pi, \sigma) \leq 2D(\pi, \sigma).
\]
For general weight function $\varphi$, to find

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

we approximate $d_\varphi$ 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 $\sum_{i=1}^{m} d_\varphi(\pi', \sigma_i) \leq 4 \sum_{i=1}^{m} d_\varphi(\pi^*, \sigma_i)$. 
For general weight function $\varphi$, to find

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

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

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

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 $\pi'$. 
Kendall $\tau$ distance $= \text{number of disagreements}$

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

$\pi_{ij}$ equals 1 if the aggregate $\pi$ prefers $i$ to $j$

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

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

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

Aggregation problem as integer program:

\[
\begin{align*}
\text{minimize} & \quad \sum_{i,j,k} \alpha_{ijk} \pi_{ijk} \\
\text{subject to} & \quad \pi_{ijk} + \pi_{jik} + \cdots + \pi_{kji} = 1 \\
& \quad \pi_{ijk} + \pi_{ikj} + \pi_{kij} = \pi_{ij} \\
& \quad \pi_{ijk} \in \{0, 1\}
\end{align*}
\]

\[\alpha_{ijk} = \sum_{rst} \# \text{voters with preference } (rst) \ast d_\phi(rst, ijk)\]

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?

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 $abc, abc, bca$:

Thank you!