The Definition of a New Correlation Variant for Rankings With Ties

Author: **Mikołaj Jakub Gazeel**
Supervised by: **Prof. Julián Urbano**

1 | Rank Similarity

- A *ranking* is an assignment of an order to some set of elements.

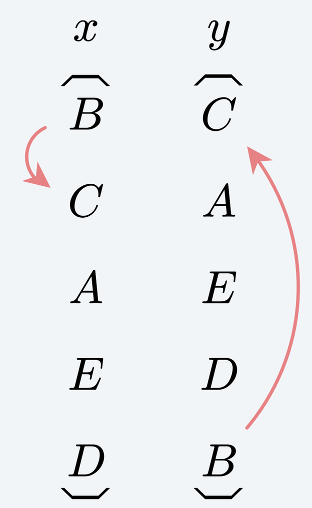
$$x = \langle B, C, A, E, D\rangle \qquad y = \langle C, A, E, D, B\rangle$$

- *Similarity measures* are used to compare rankings.
- Three of these measures which are widely used and are the main focus of this work are: τ due to Kendall [3], τ_{AP} due to Yilmaz et al. [6], and τ_h due to Vigna [4].
- Given two rankings these measures return a value between −1 and 1.
- 1 signifies perfect agreement between the rankings, and −1 perfect disagreement.
- As an example:

$$\tau(x, y) = 0.2 \qquad \tau_{AP}(x, y) \approx -0.042 \qquad \tau(x, x) = 1$$

High-level overview of τ, τ_{AP}, and τ_h

- τ is a measure which counts the number of pairs that agree and disagree in their order between the rankings.
- The amount of agreeing minus disagreeing pairs is the *numerator* of τ
- This value is then scaled by the *denominator* to fit in the range [−1, 1].
- τ_h along with the rankings requires a weighting function.
- This function $w(i, j)$ decides the contribution of each pair of elements (i, j) to the result.
- τ is just τ_h with the weight function $w(i, j) = 1$
- τ_{AP} is τ_h with the weight function $w(i, j) = \frac{1}{\max(x_i, x_j) - 1}$.
- x_i denotes the rank of element i in one of the rankings.



An example of a disagreeing pair (B, C)

- So τ_{AP} puts more importance on the elements near the top of the ranking.
- And τ_h can be thought of as a superset of τ and τ_{AP}

2 | Handling Ties

- What if ties appear in a ranking? Consider for example these two rankings:

$$k = \langle B, C, D, [E, A, F]\rangle \qquad l = \langle [E, A, F], D, C, B\rangle$$

- In the current interpretation tied items represent some kind of uncertainty about the actual order of the elements. During the comparison we assume that the tied elements have a certain objective order but the creator of the ranking was not able to set it.
- For example τ^a when comparing k and l, returns the average value of τ over all permutations of elements *E*, *A*, and *F*.
- Recently, Webber et al. [5] and Corsi & Urbano [1] introduce a previously unexamined meaning of ties. What if we instead assume that ties represent items which are **meant to be exactly at the same rank**?

Imagine an ice skating competition and two distinct juries each awarding a discrete amount of points to the participants based on their form, speed, etc. At the end of the competition these will produce two rankings of the participants, but what if two participants are tied? How do we compare the output of these juries?

- The current variants are not sufficient to answer this question.
- To tackle this problem we will try to define this new variant for the three measures introduced previously.

How can the w-variant be defined for conjoint measures, specifically τ, τ_{AP}, and τ_h?

3 | The Axiomatic Approach

- The approach to solving this problem is one where we stipulate **how** τ^w should behave, and **what it should represent** using a set of axioms.
- Let the set of all possible rankings of length n without ties be R_n. Whereas, the set of all rankings of length n possibly including ties will be denoted by $\hat{R}_n$. This also gives: $R_n \subset \hat{R}_n$

Observation: The numerator of τ is a valid distance metric for R_n.
Claim: To treat tied items as really occurring at the same place the numerator of τ^w must result in a valid distance metric for $\hat{R}_n$. (The other variants do not satisfy this)

- Kemeny [2] considered this problem and our axiomatic approach is based on his. Surprisingly we arrive at the same results, in terms of the distance metric.
- Armed with this claim: we will first provide axioms about $d\tau^w$, the distance metric over $\hat{R}_n$. This is the basis for the numerator of the new τ^w.
- Thereafter, we construct τ^w from $d\tau^w$ by determining the denominator.
- Finally, we define τ_h^w and τ_{AP}^w.

4 | Arriving at the definition

$d\tau^w$ should conform to the following:

- Axiom 1. $d\tau^w(l, r)$ *should be a valid distance function for the metric space over* $\hat{R}_n$.
  - Axiom 1.1. $d\tau^w(l, r) \geq 0$ *for all* l, r.
  - Axiom 1.2. $d\tau^w(l, r) = 0$ *if and only if* $l = r$.
  - Axiom 1.3. $d\tau^w(l, r) = d\tau^w(r, l)$.
  - Axiom 1.4. $d\tau^w(x, z) \leq d\tau^w(x, y) + d\tau^w(y, z)$ *for all* x, y, z.
- Axiom 3. $d\tau^w(l, r) = \sum_{i<j} d^w_{l,r}(i, j)$. *Where* $d^w_{l,r}(i, j)$ *is some real function dependent only on* $\mathrm{sgn}(l_i - l_j)$ *and* $\mathrm{sgn}(r_i - r_j)$.
- This axiom, not only follows the structure of τ is, but also ensures that the identity of the elements does not matter in our distance metric.
- Axiom 4. $d\tau^w(\langle A, B\rangle, \langle B, A\rangle) \geq d\tau^w(\langle A, B\rangle, \langle [A, B]\rangle)$

Just with these axioms, all of which are relevant to the w-variant, in detailed steps we show that: The **only** definitions of $d\tau^w$ which satisfy them are:

$$d\tau^w(x, y) = \sum_{i<j} d^w_{x,y}(i, j)$$

With a parameter $1 \leq \beta \leq 2$ and:

$$d^w_{x,y}(i, j) = \begin{cases} 0 & \text{if } \mathrm{sgn}(x_i - y_i) = \mathrm{sgn}(x_j - y_j) \\ 2 & \text{if } \mathrm{sgn}(x_i - y_i)\mathrm{sgn}(x_j - y_j) = -1 \\ \beta & \text{otherwise} \end{cases}$$

- More considerations laid out in the work, such as a choice for the denominator, lead to a definition of τ^w which is still dependent on β.
- The only degree of freedom left now is the value of β, to set this value we will consider the behaviour of τ^w for independent rankings.
- If we pick two rankings at random from the set of all possible rankings, we would expect the result of the metric to be 0. This leads to our final axiom:
- Axiom 8. *Define two independent variables* L *and* R *where for all* $x \in \hat{R}_n$ *we have* $\Pr[L = x] = \Pr[R = x] = \frac{1}{|\hat{R}_n|}$. *Now it should be the case that:* $\mathbb{E}[\tau^w(L, R)] = 0$.
- The value of β which satisfies *Axiom 8* is **different** for different lengths of rankings.
- *Axiom 8* cannot be satisfied for all lengths of rankings. Therefore we have to compromise. One value will be chosen for β and this will result in a slight amount of bias in τ^w. We investigate the values of β and this bias.

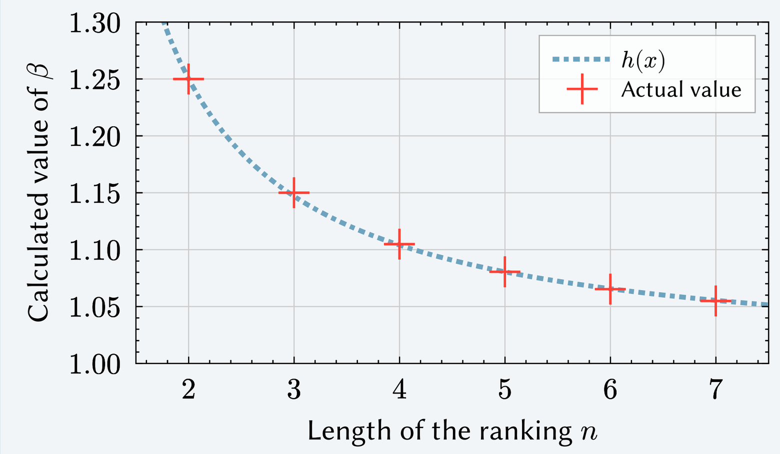


Figure 1: The value of β satisfying *Axiom 8* versus n.

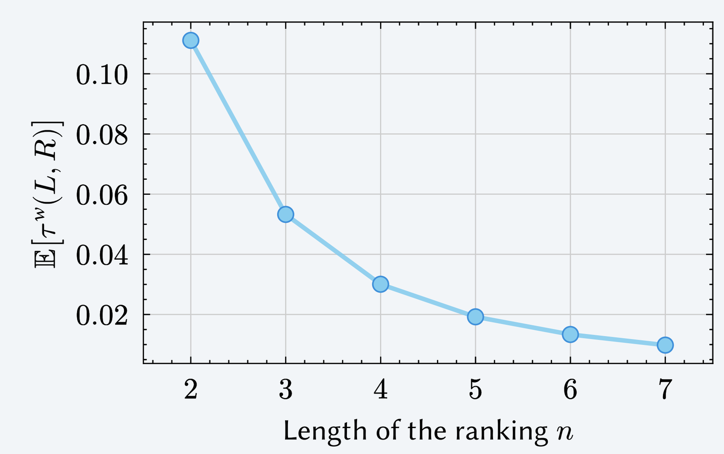


Figure 2: Bias of τ^w for small n.

- The value of $\beta = 1$ was chosen as the extrapolation of data in Figure 1 appears to converge to 1 as the length of the rankings increase.
- The bias resulting from this choice is still very small and most likely acceptable in real world applications, where comparisons are ran on very large rankings.

5 | Results and future considerations

- Setting $\beta = 1$ results in a final definition for τ^w:

$$\tau^w(x, y) = \frac{\sum_{i<j} g^w_{x,y}(i, j)}{\frac{n(n-1)}{2}}, \quad g^w_{x,y}(i, j) = \begin{cases} 1 & \text{if } \mathrm{sgn}(x_i - y_i) = \mathrm{sgn}(x_j - y_j) \\ -1 & \text{if } \mathrm{sgn}(x_i - y_i)\mathrm{sgn}(x_j - y_j) = -1 \\ 0 & \text{otherwise} \end{cases}$$

- The definitions for τ_h^w and τ_{AP}^w follow their original counterparts which were introduced earlier but utilize $g^w_{x,y}(i, j)$ for the contribution of pairs of elements.

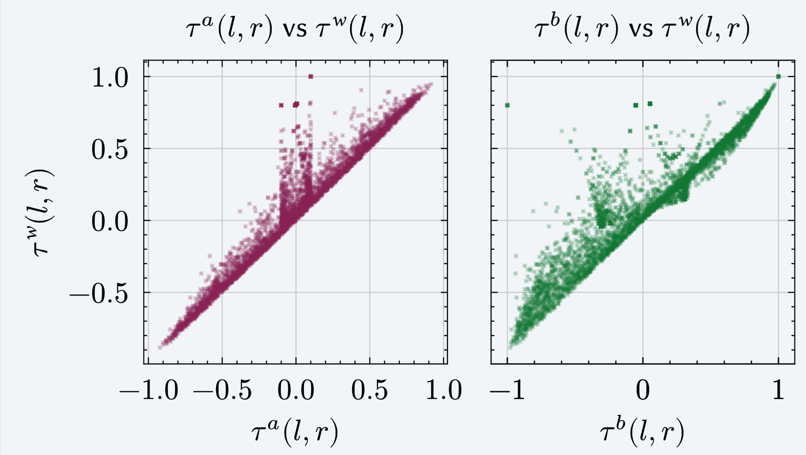


Figure 3: Comparison using synthetic data, ties high.

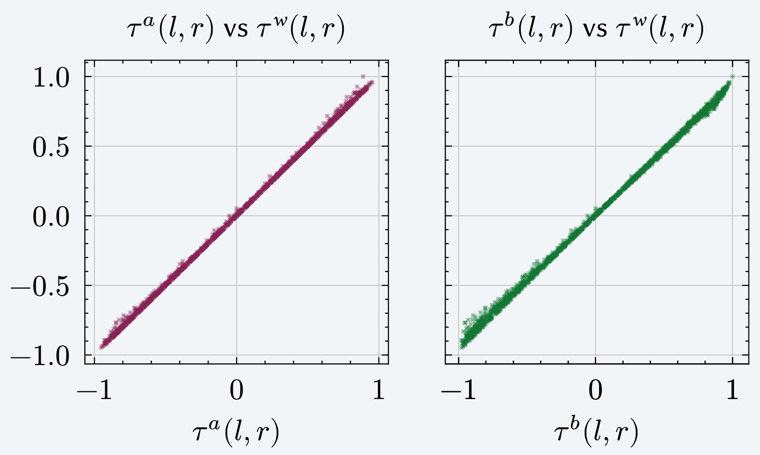
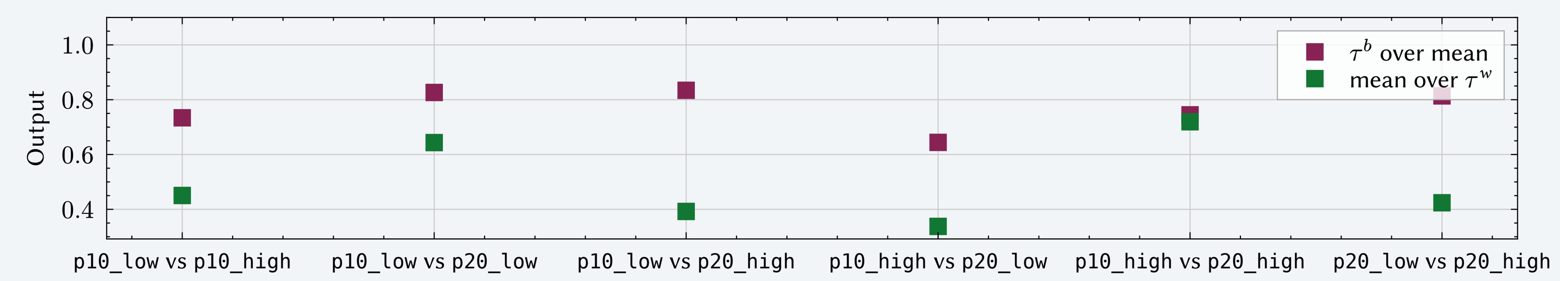


Figure 4: Comparison using synthetic data, ties low.

- The axioms show that τ^w forms a valid distance measure for rankings with ties.
- τ^w differs from other variants only when there is a high prevalence of ties in the compared rankings.
- Thanks to the axioms all w-variants show required behaviour: Returning values close to 1 if the rankings are strongly positively correlated and a value close to for 0 independent rankings.
- We also examine the real world impact of τ^w in the field of information retrieval. In many cases when comparing retrieval systems metrics output the same scores for two systems. For some metrics this means that the systems are completely tied and there is no intrinsic order to them. τ^w if useful here. It could possibly allow for calculation of more realistic similarity scores, which currently may be overestimated:



- The other end of the range poses a problem. If we compare the two rankings with ties k and l from section 2: $\tau^w(k, l) = -0.6$, whereas we expect −1. This is unwanted behaviour! Future work should consider ways of remedying this problem, various solutions for this are proposed in the work. The time constraints of the thesis did not allow for the extensive examination of these solutions.

[1] Matteo Corsi and Julián Urbano. 2024. The Treatment of Ties in Rank-Biased Overlap. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, July 2024. Washington DC, USA, 251–260. https://doi.org/10.1145/3626772.3657700
[2] John George Kemeny. 1959. Mathematics without Numbers. *Daedalus* 88, 4 (1959), 577–591. Retrieved May 23, 2025 from https://www.jstor.org/stable/20026529
[3] Maurice George Kendall. 1938. A New Measure of Rank Correlation. *Biometrika* 30, 1/2 (June 1938), 81–93. https://doi.org/10.2307/2332226
[4] Sebastiano Vigna. 2015. A Weighted Correlation Index for Rankings with Ties. In *Proceedings of the 24th International Conference on World Wide Web*, May 2015. Florence, Italy, 1166–1176. https://doi.org/10.1145/2736277.2741088
[5] William Webber, Alistair Moffat, and Justin Zobel. 2010. A similarity measure for indefinite rankings. *ACM Trans. Inf. Syst.* 28, 4 (2010). https://doi.org/10.1145/3626772.3657700
[6] Emine Yilmaz, Javed A Aslam, and Stephen Robertson. 2008. A new rank correlation coefficient for information retrieval. In *Proceedings of the 31st International ACM SIGIR Conference on Research and Development in Information Retrieval*, July 2008. Singapore, Singapore, 587–594. https://doi.org/10.1145/1390334.1390435