Quantifying Uncertainty due to Ties In Rank Correlation Coefficients

Andreas Tsatsanis
atsatsanis@tudelft.nl
Supervisor: Julián Urbano

1. Introduction

Rank similarity coefficients can measure *agreement* between *different rankings* of the *same items*. Kendall's τ uses *concordance*:

R = `a, b, c, [d, e]`
B = `b, a, [e, d, c]`

(a, b) is *discordant*, **(a,c),(a,d),(a,e),(b,c),(b,d)** and **(b,e)** are *concordant*.

$$\tau(R,B) = \frac{\sum_{\{i,j\}} c(i,j)}{n(n-1)/2} = \frac{5}{10}$$ *what fraction of item pairs are concordant?*

Extensions for τ: if some pairs matter more than others, a *weight function* can be used:

$$w(i,j) = \frac{1}{max(i,j)}$$ *item pairs later in the ranking contribute less*

$$\tau_w(R,B) = \frac{\sum c(i,j)w(i,j)}{\{weighted\ maximum\}} \approx 0.485$$

Extensions for τ: *tied items.* This can happen when there is *uncertainty* in the ranking. Variants of τ: τ^a, τ^b allow for computing τ with ties. But they do not reflect uncertainty:

$$\tau^a(R,B) = 0.5 \quad \tau(\langle a,b,c,d,e\rangle, \langle b,a,c,d,e\rangle) = 0.8$$
$$\tau^b(R,B) \approx 0.63 \quad \tau(\langle a,b,c,d,e\rangle, \langle b,a,e,d,c\rangle) = 0.2$$

Each way of breaking ties gives a different τ

It would be useful to *quantify* how much variability there is in the correlation.

RBO *bounds: RBO*^{high} *RBO*^{low},
- can show *impact* of ties on correlation value

2. Research Question

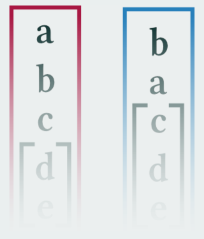
Can we efficiently compute τ^{min}/τ^{max}?
How do τ^{min}/τ^{max} compare to τ^a/τ^b?

3. Method

How can we compute τ^{min}/τ^{max}?
→ Trying all permutations: O(n!²)
→ Maximising $\sum_{\{i,j\}} c_{R,B}(i,j)w(i,j)$ is hard(Knapsack)
→ RBO: construct rankings to maximise *overlap*
! Concordance is not the same as overlap

Overlap depends on depth

Concordance depends only on relative order between item pairs



a > b b > c c > d d = e ...	a < b b > c c = d d = e ...

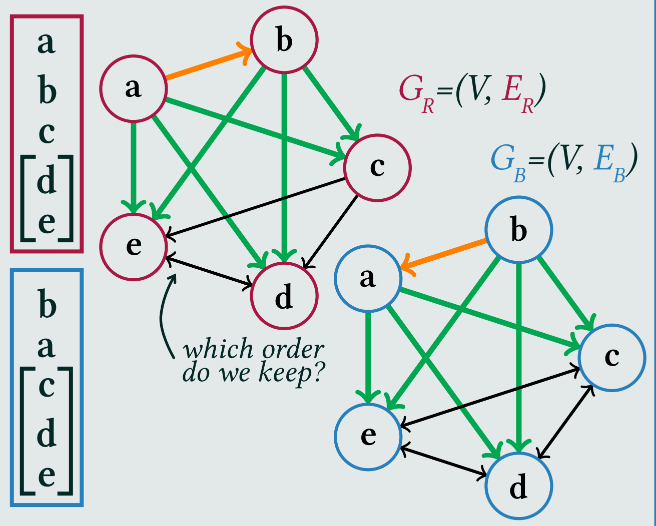
Concordance: sum of pairs with same relative order in both rankings

Idea:
Like RBO, construct extremal rankings.

Requirements:

- Pick an order for the items with maximum concordance.

- Valid order must be transitive



Same edge: concordant
Valid solution: acyclic
Concordance = $|E_R \cap E_B|$

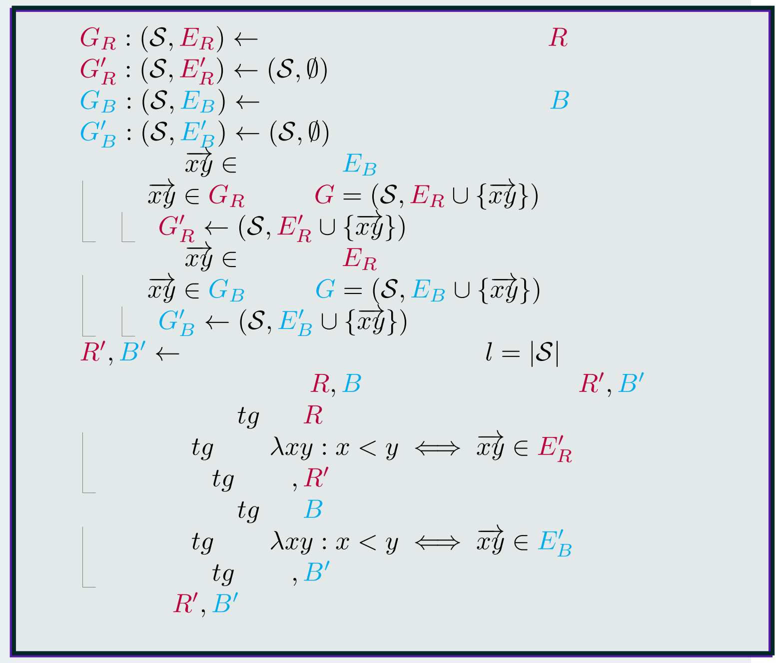
4. Proposed Algorithm

1. Take edges from $E_R \cap E_B$

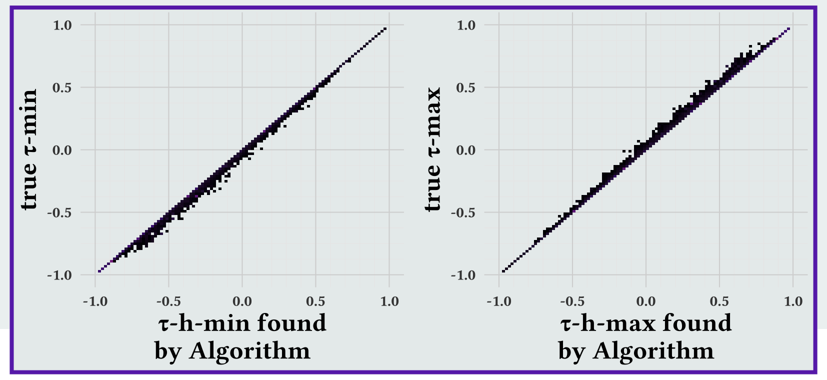
2. Check: can they be in the final ranking?

Condition: no cycles.

3. If allowed, greedily add to solution.



- Optimal for unweighted case (proof: exchange argument)
- Minimisation: invert sorting for E_R
- Weighted case: sort E by *descending weight*
  - τ_{AP}: optimal,
  - τ_h: approximation



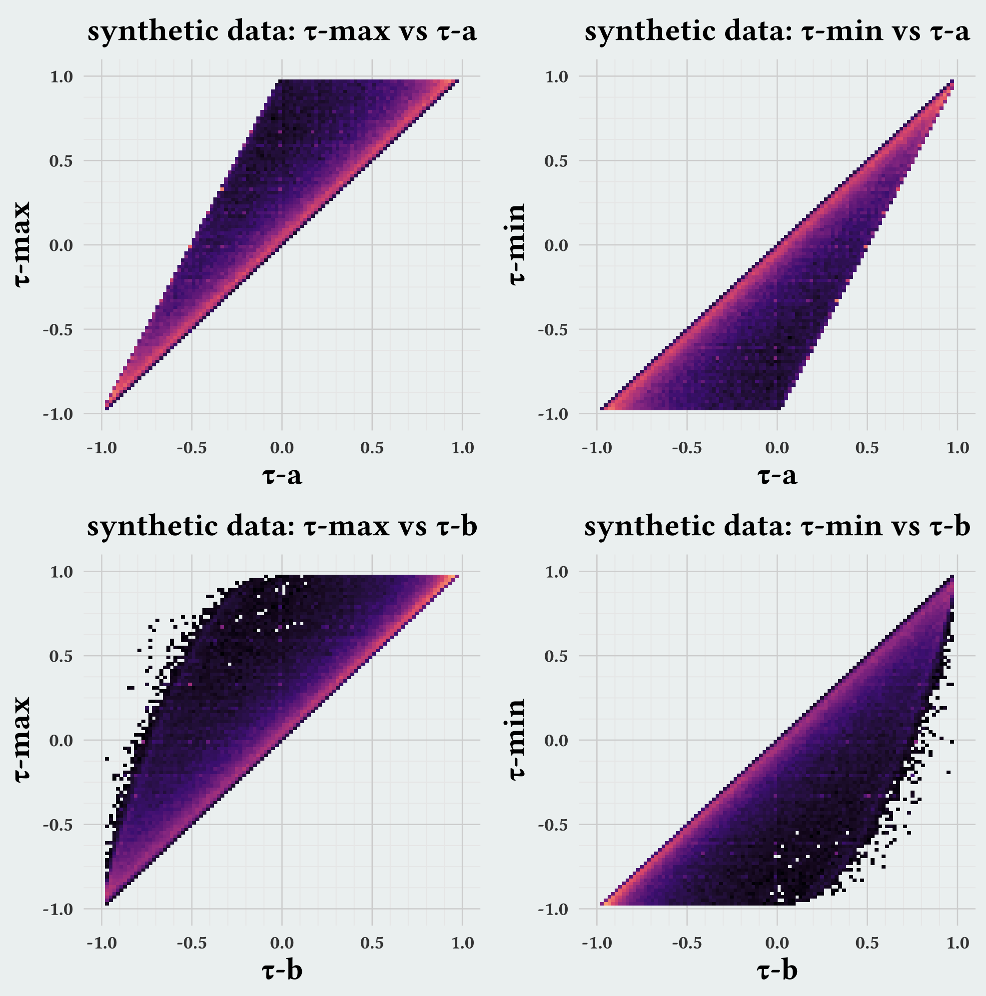
5. Results

→ Compare τ^a, τ^b with τ^{min}, τ^{max}
→ Expectation: if variants give accurate estimates in presence of ties, τ^a and τ^b should be tightly bounded by τ^{min}, τ^{max}.

In reality, the interval $[\tau^{min}, \tau^{max}]$ is wide around τ^a and τ^b.

This shows that a single value (τ^a or τ^b) cannot encapsulate the uncertainty in the rankings caused by the ties.

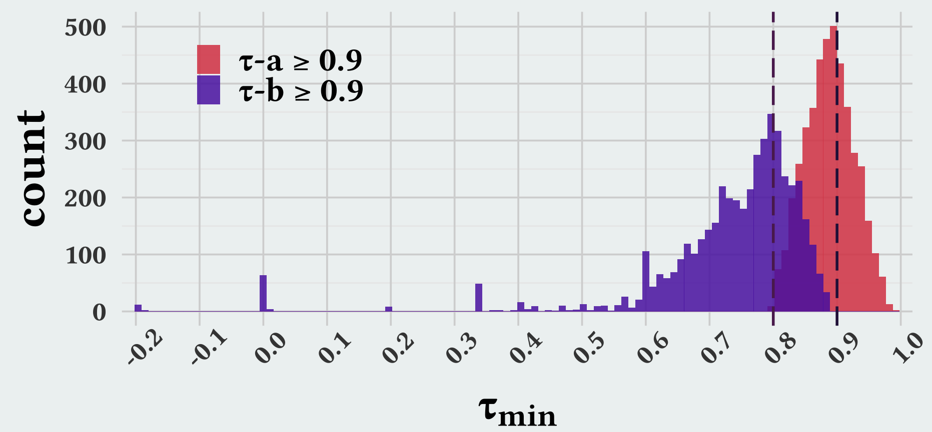
These findings were shown both for real (TREC) and synthetic datasets.



(Shown above) Synthetic data: uniform τ, uniform tie distribution, lengths 3-150, 250k samples

6. Implications

τ-min density for τ-a ≥ 0.9 and τ-b ≥ 0.9



		$\tau^{min} < 0.9$	$\tau^{min} < 0.85$	$\tau^{min} < 0.8$
$\tau^a \geq 0.9$	4413	2737	775	0
$\tau^b \geq 0.9$	9100	7424	5163	3238

When ties represent uncertainty:
→ τ^a can misrepresent correlation
→ τ^b can misrepresent correlation a lot
→ Possible consequence: false positive results
→ Problem: τ^b is the most widely used coefficient!

Bounds τ^{min}, τ^{max} can inform decisions by supporting or contradicting τ^a, τ^b estimates.

7. Conclusion

Uncertainty bounds are important for τ correlation with ties, just as in RBO.

We proposed an algorithm for computing τ^{min}, τ^{max}

- Does not generalise to all weight functions.

- Good approximations for reasonable weighting schemes.

Using uncertainty bounds can help inform decisions, *when ties in rankings are induced by uncertainty.*

Future Work

- Explore missing constraints on w

- Extend the algorithm to provide a distribution of values, akin to the work shown for RBO.

- Implementation in statistical software packages

- Encourage researchers to question how the tools and metrics they use reflect the properties of the rankings in their field.