<Effects of the assumption on ties in unseen parts of a ranking>

<What will happen if we relax the assumption that ties do not occur in unseen parts?>

< Lukas Roels[1] >

Supervisor(s): <Julian Urbano[1] >, <Matteo Corsi[1] >

[1]EEMCS, Delft University of Technology, The Netherlands

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering
June 23, 2024

Name of the student: <Lukas Roels>
Final project course: CSE3000 Research Project
Thesis committee: <Julian Urbano>, <Matteo Corsi>, <Mathijs Molenaar>

Effects of the assumption on ties in unseen parts of a ranking

LUKAS ROELS, Delft University of Technology, The Netherlands

Abstract

Rankings are more present in our daily lives than most people realize. Whether you are browsing Netflix and getting movies or shows based on your previous likes or dislikes, or you want to compare search engine results. To use rankings in the field of Computer Science a rank similarity is needed. Rank-Biased Overlap is one of those. It is top-weighted, can be used on uneven rankings, and when only a part of the ranking is known. A well-known problem in rank similarity measures is ties. There have been some ways of dealing with ties proposed since RBO was introduced. These ways have been shown to be promising but they only relate to the seen part. The unseen part of rankings is still a new concept with little research done about it. This paper aims to change that a bit. First, a full explanation is given of the three variations of dealing with ties. Then using these variants we show how the assumption that no ties exist in the unseen part affects these variants. Also, the current extrapolation method is researched as there is also a big influence of the above-mentioned assumption. We then use simulated data to give a clear data visualization to show how the theory relates to practice. We have tried to be clear and concise with our explanations and data visualizations so future researchers can use this paper to improve and progress RBO in the world of rank similarity measures.

CCS Concepts: • **Information systems** → **Retrieval effectiveness**.

Additional Key Words and Phrases: Rank correlation, Rank similarity measure, Rank-Biased Overlap, Ties,

ACM Reference Format:
Lukas Roels. 2024. **Effects of the assumption on ties in unseen parts of a ranking**. 1, 1 (June 2024), 7 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Rankings and similarity between rankings are everywhere in modern computer science. For example, the similarity between two people who rank their top 30 football teams which can be used in recommender systems, or similarity between search engines, or even fraud detection. Many algorithms have been proposed over the past decennia to find this similarity. Among these similarity measures some well-known formulae can be found like ρ by Spearman [5] or τ by Kendall [2]. Both are already over 80 years old but still in 2024 no perfect solution has been found to compute the rank similarity measure between rankings.

The origin of this paper starts with Rank-Biased Overlap, a rank similarity measure proposed by Webber et al.[3]. What was found was a new way of handling uneven, different lengths, and indefinite rankings. Rank-biased overlap sees rankings as infinite. When computing the similarity it stops at a certain depth. Everything before this depth is referred to as seen items and everything after this depth is referred to as unseen. Extrapolation is used to compensate for the loss of information in the unseen part. The way this extrapolation is done is explained in detail in section 3.5 as more background about Rank-Biased Overlap is needed.

A major property of rankings is ties. Ties, in this context, are defined as two or more unique items that share the same rank inside a ranking. Webber et al. [3], who proposed the original Rank-biased overlap, also gave a way of handling ties. A recent paper by Corsi and Urbano [1] proposed new ways of handling ties in Rank-Biased Overlap. In this the following assumption is made: "We assume no tied items in the unseen part" [1, Corsi and Urbano p.6]. This assumption means that when calculating the Rank-Biased Overlap, we assume no ties in the unseen part of a ranking.

This paper is focused on the following research question: **"What will happen if we relax the assumption that ties do not occur in unseen parts?"**. We will give insight into how much this assumption impacts the similarity between two rankings by simulating seen and unseen rankings. Another focus is providing an overview of the impact of this assumption on the different ways both Webber et al. [3] and Urbano and Corsi [1] handle ties.

We will start in chapter 2 where more detail is given about the background of our research and the problem behind our research. Then in chapter 3, a theoretical reasoning is given about the impact of the assumption. After this in chapter 4, we back our theory by showing data and explaining how these results correlate to chapter 3. In chapter 5, a reflection on the ethical aspects of our research is made. Lastly, in chapter 6 we talk about what we concluded factoring in all aspects of the paper and how we believe this research could be used to improve RBO.

Author's address: Lukas Roels, Delft University of Technology, Delft, The Netherlands.

XXXX-XXXX/2024/6-ART $15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

2 BACKGROUND

In this chapter, Rank-Biased Overlap is explained, in section 2.1, as well as our formal problem description, which can be found in section 2.2, which shows why this research is necessary.

2.1 Background

Rank-Biased Overlap, referred to now as RBO, is a rank similarity measure proposed by Webber et al. [3] more than 10 years ago. It has the potential for a wide range of applications. One example is shown in this paper by Sarica et al. [4]. It was the first measure that took indefinite rankings into account while being able to make it more or less top-weighted. As can be seen in equation 1, RBO uses the sum of agreements up to infinity. The agreement is defined as the proportion of overlap of two rankings at a certain depth.

$$RBO_{S,L,p} = \frac{1-p}{p} \sum_{d=1}^{\infty} A_{S,L,d} \cdot p^d \tag{1}$$

RBO uses a variable p which is called persistence. P gives a specific weight to a rank. It determines how top-weighted the calculation is. A larger p means less top-weighted and a lower p more top-weighted.

As mentioned in the introduction, Webber et al. [3] proposed a way of handling ties in rankings. They gave the top rank of a tied group to each element of that group like how sports rankings are seen. This way of handling ties is referred to as w-variant.

A new way of handling ties was proposed by Corsi and Urbano [1]. They propose two variants (referred to as a-variant and b-variant) where the a-variant computes the average contribution of an item in a tied group for every permutation of this group. The b-variant "accounts for the amount of information actually available to measure overlap" as explained in [1, Corsi and Urbano p.4]. Although both these variants have shown through data to be a promising solution it is not yet a perfect solution.

2.2 Problem Description

A key part of RBO is RBO_{min}, RBO_{max} and RBO_{ext}. RBO_{min} is the lower bound of RBO. It works by assuming every item in each list in the unseen part does not match an item in the other list. RBO_{max}, the upper bound of RBO, is the opposite. It assumes every item in the unseen part will be matched to an item in the other ranking. RBO_{ext} uses extrapolation to estimate the real RBO. As mentioned in chapter 1, it will be explained in detail in section 3.5. When using RBO_{min}, RBO_{max} and RBO_{ext} an assumption is made. The assumption that sits at the core of this paper. It assumes no ties occur in the unseen part of the ranking.

This is a logical assumption as by definition we do not know what is in the unseen part. However, this means that it is not yet a perfect solution for computing the similarity measure and its bounds. The problem behind the need for research shown in this paper is exactly that. RBO needs to be able to compute the similarity between rankings without assuming anything about the rankings. This paper will try and show what the effects are of one of the assumptions in RBO[1]. When talking about assumptions throughout this paper, unless otherwise specified, we are referring to the assumption defined in our research question.

x	b	(c	d)	e	f	g	h	i	j	k	l	m	n	(o	p)
z	y	x	w	(v	u	t)	s	e	(g	i	a)				

Table 1. Example of two rankings where the parentheses represent tied groups and the longer ranking will be referred to by L and the shorter by S

x	b	(c	d)	e	f	g	h	i	j	k	l	m	n	o	p	= L
z	y	x	w	(v	u	t)	s	e	g	i	a					= S

Table 2. Rankings shown in table 1 where ties are removed after depth 10 for S ranking and depth 11 for L ranking

x	b	(c	d)	e	f	g	h	i	j	k	= L
z	y	x	w	(v	u	t)	s	e	g		= S

Table 3. Rankings shown in table 1 but truncated to depth 10 for S ranking and depth 11 for L rankings

3 TIES IN UNSEEN PART

To understand what the effects are of the assumption, we will discuss the difference in RBO between a full ranking (example shown in table 1) and the same ranking with no ties after truncation depth (example shown in table 2). Then a discussion about the effect of the assumption on the extrapolation is given, using rankings shown in table 1 and 3. This chapter focuses more on discussing each part of RBO, its variants, the extrapolation, and what influence the assumption has on them. Firstly, in section 3.1, 3.2, and 3.3 the impact of the assumption on the different variants is given. This is done by assuming perfect extrapolation so only the effect of tie and no tie is given. Then in section 3.4 the influence of the length of a tied group is given. Lastly, in section 3.5, we reason how the current method of extrapolation is affected by the assumption.

3.1 W-variant

The w-variant as explained in chapter 2 is the original proposed way of dealing with ties in a ranking. The assumption affects this variant greatly as now only the first item in a tie group will have the rank that normally all items in the tie group have under the w-variant. The second item will have the rank below and so forth.

This can be shown using table 1 and table 2. In table 1, the items o and p belonging to the tied group at depths 15 and 16 will both have rank 15 when computing RBO under the w-variant. In table 2 however, the items o and p will have rank 15 and 16 respectively when computing RBO. Because of this item p will have less weight on the overall RBO.

Now using equation 2 and 3, made by Webber et al. [3, Webber et al. p. 21], we will calculate agreement, taking only items g, i,

[1]The assumption of no ties in unseen parts

and a into account, for depths 10, 11, and 12. Firstly for table 1, we get 4/22, 4/23, and 4/24 for depth 10, 11, and 12 respectively. For table 2, we get 2/20, 4/22, and 4/24 respectively. We are only taking the above-mentioned items into account as all items at lower ranks will not be influenced by the assumption at depths 10, 11, and 12. As you can see for depth 10 we get a higher agreement when calculating RBO under the assumption but for depth 11 we get a higher agreement when we would not make the assumption. At depth 12 both agreements are the same.

This is interesting to see as the assumption affects agreement when a tied group is not fully seen but when the whole group is seen or active, agreement at that depth becomes the same.

$$X_{S,L,d} = |S_{:d} \cap L_{:d}| \tag{2}$$

$$A^{w}_{S,L,d} = \frac{2 * X_{S,L,d}}{|S_{:d}| + |L_{:d}|} \tag{3}$$

3.2 A-variant

The variant where the assumption is expected to have the least influence is the a-variant. This variant computes the average agreement of all possible combinations of the order of items within a tied group. In general, only when a tied item is crossing an impact of the assumption can be found. Crossing is defined as follows: "An item within a crossing group will be able to contribute to overlap in as many permutations as it is placed at or above d" as explained in [1, Urbano and Corsi p.4].

When computing RBO and traversing the list calculating agreement at each depth the impact can be seen. For example at depth 10 in table 1, following the reasoning by Urbano and Corsi [1, Urbano and Corsi p.4 section 3.2], items g and i in the tied group will have a contribution of 1/3. Items g and i will have a contribution of 2/3 and 1 at depths 11 and 12 respectively. Item a will have the same contribution for each depth respectively but will not contribute to the agreement function as it is not matched in ranking L. When we are computing contribution at each depth in table 2, we can see that at depth 10 item g has a contribution of 1. At depths 11 and 12, items i and a will have a contribution of 1 at their respective depths. Item a however will not contribute to the agreement function as it is not matched in ranking L.

If we only take into account these three items when calculating agreement, using equation 4 [1, Urbano and Corsi p.4 section 3.2], we will see an agreement at depths 10, 11, and 12 of 2/30, 4/33 and 2/12 respectively for table 1. For table 2 we see that agreement at depth 10, 11, and 12 will be 1/10, 2/11, and 2/12 respectively. At depth 12, we can see that the agreement is the same for both tables. This is because at depth 12 every item in the tied group is now active and can contribute fully to the agreement.

$$A^{a}_{S,L,d} = \frac{1}{d} \sum_{e \epsilon \Omega} c_{e,S|d} * c_{e,L|d} \tag{4}$$

3.3 B-variant

The b-variant has the same contribution formula as the a-variant as can be seen in equation 5. The difference lies in the denominator. The a-variant uses the depth d as its denominator. The b-variant compensates for the information actually available. When a tied group is not active but crossing then this will show in the denominator.

Using the same example shown in section 3.2, we already have the contribution of items g, i, and a for both table 1 and 2. Again we will only take those three items into account when calculating the agreement, using equation 5, for depths 10, 11, and 12. For table 1, we get $\frac{\frac{2}{3}}{\sqrt{9+\frac{1}{3}}\sqrt{10}}$ for depth 10, $\frac{\frac{4}{3}}{\sqrt{9+\frac{4}{3}}\sqrt{11}}$ for depth 11 and $\frac{2}{\sqrt{12}\sqrt{12}}$ for depth 12. Which translates to 0.069, 0.125 and 0.1667 respectively. We already calculated agreement for table 2 in section 3.2 resulting in 1/10, 2/11, and 2/12 for depth 10, 11, and 12 respectively. These agreements are the same for the a-variant and b-variant as no item is crossing at depths 10, 11, and 12

We can see some similarities with the a-variant. At depth 12, agreements between variants a and b are the same as well as between table 1 and 2. For depths 10 and 11, we notice that the agreement of the a-variant is slightly smaller than that of the b-variant. This is a property of the agreement of the a-variant as mentioned by Urbano and Corsi [1, Urbano and Corsi p. 5].

$$A^{b}_{S,L,d} = \frac{\sum_{e \epsilon \Omega} c_{e,S|d} * c_{e,L|d}}{\sqrt{\sum_{e \epsilon \Omega} c^2_{e,S|d}} * \sqrt{\sum_{e \epsilon \Omega} c^2_{e,L|d}}} \tag{5}$$

3.4 Length of tied group

The one variable that influences all variants is the length of a tied group. In the above sections, we mention that as soon as a tied group is active the agreement between table 1 and 2 is the same. The impact the length has is that the longer the length of a tied group the longer it takes for the tied group to become active resulting in more deviations per depth. In our example, using only depths 10, 11, and 12, we see that there is only a difference in agreement at depths 10 and 11. If the items in the following 10 depths would also belong to the tied group then a difference will be found from depth 10 through 21 and then reaching depth 22 agreement would become the same.

3.5 Extrapolation

The extrapolation of RBO is an estimate of what the RBO would be if we saw the full ranking. The way extrapolation is done is by "assuming that the degree of agreement seen up to depth k is continued indefinitely" as mentioned in [3, Webber et al. p.19]. Under this way of calculating the extrapolation, the assumption will only have an impact when the last seen item in a ranking belongs to a tied group located at the start of the unseen part. This behavior can be seen in table 1 and 3. The property of this item changes from belonging to a tied group to an item with a unique rank.

To show the differences explained below we will once again only use item g, i, and a. In the case of the w-variant at the truncation depth 10 of ranking S in table 3, we will have an agreement of 1/10. From then on agreement is assumed to be constant. For table 1 however, we see that agreement is actually 2/12, at depth 10, as item i contributes to the overlap and item i and a to the denominator of the agreement function shown in equation 3. The difference in overlap and agreement is mainly caused by the loss of information but due to the assumption, the extrapolation is not altered to compensate for the above shown difference. Now 1/10 is used as a constant agreement but in reality, 2/12 should be used as that is the real agreement at depth 10.

The same can be shown for variants a and b. As we have calculated in section 3.2 and 3.3 when only taking items g, i, and a into account. For both variants the actual agreement at depth 10 and the agreement under the assumption, which results in the same agreement at depth 10 for table 3, is different. Then the same as with the w-variant happens. The wrong agreement is used as a constant.

4 METHODOLOGY - RESULTS

In this chapter, the focus lies more on showing data backing our reasoning and how we simulated the rankings to show these results. In section 4.1, we show how the variable p of RBO and the assumption affect RBO. We do this by first explaining how we simulated our data and then showing and discussing our results. Then in section 4.2, the influence of the assumption on the current extrapolation is shown. Once again first by explaining how we simulated and then showing our data.

4.1 P experimentation's

4.1.1 Simulation. To show how the difference in RBO changes based on p we are simulating rankings, using the code by Urbano and Corsi [2], as follows.

We start with specifying the max length a simulated ranking can have: **20**. We then sample a number, for each ranking, from 0.75 of the maximum length to the maximum length. In our case from 15 to 20. For the number of unique items in our domain, we take 1000 if nothing is specified. The reason we are taking such a high domain is to replicate infinity. We need to reproduce infinity to fully show the weight of extrapolation. We then generate two rankings with both random lengths between the above-mentioned 15 and 20.

All pairs of rankings are generated with random tau which is the percentage of conjointness. After this, we sample a truncated length. This is done between 0.25 of the maximum length and 0.5 of the maximum length to ensure we have enough data to be considered as unseen.

In the next step, we check if there are ties in the unseen part, and if not we simulate with a new length and a new truncated length until we have ties in the unseen part. This way it is ensured that for each pair of rankings, we can show a difference in RBO.

In the end, we compute the truncated part of the ranking and most importantly we reshape our original ranking so no ties exist after the truncated depth. We end up with six rankings which are three pairs of two rankings:

(1) Full ranking (example shown in table 1)
(2) Full ranking with no ties after truncation depth (example shown in table 2)
(3) Truncated ranking (example shown in table 3)

Often in plots, you will find RBO reality, RBO under assumption, and RBO truncated. These refer to the RBO of the rankings number 1, 2, and 3 from above respectively.

[2]https://github.com/julian-urbano/sigir2024-rbo

The choice for having the length of the truncated ranking at most half of the length of the full ranking is because we want to simulate an unseen part. So by the above choice, we make certain that there is at least the same amount of unseen items as seen items.

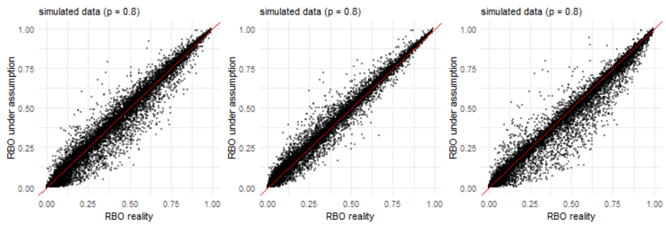
4.1.2 Results. In this subsection, we will discuss the influence of p on the difference in RBO between full rankings and full rankings under the assumption. This will also show how only the assumption affects the RBO without noise from the extrapolation. The parameter p is an important one as it is responsible for top-weightedness. P has a domain between 0 and 1. A larger p means less top-weightedness and a lower p means more. We will show results for the three different variants: w, b, and a. The scatter plot is computed based on 50000 pairs of rankings and the plot that shows the RBO based on p is generated with 10000 pairs of rankings and uses an interval of 0.002 for p.

Using figure 1 we can see that for variants w and b, there are a lot of outliers. For variant b we see that although there are outliers it is less dense than variant w. As for the a-variant, the cloud is much more dense, and fewer outliers can be found. The agreement function of the a-variant, shown in section 3.2, is the only agreement function that uses the actual depth. The w-variant uses, if applicable, the depth of the bottom rank of the tied group in the denominator. The b-variant uses the contributions in the denominator resulting in similar results to the a-variant for small tied groups but larger deviations for large tied groups. Because of this, the a-variant is less prone to outliers as only the contribution function is affected. Also, the a-variant tends to give a higher RBO under assumption. We can see somewhat the opposite from the b-variant where the cloud tends more to RBO reality. This means that on average the a-variant gets a higher RBO under the assumption than what the actual RBO is.

An interesting thing can be seen when comparing figure 1 and 2. We see that when we use a higher p, meaning less top-weightedness, the cloud is much more dense and outliers are limited. This, we believe, happens as sometimes RBO under assumption will give a higher RBO than it should be and sometimes it gives a lower, depending on where the tied group is placed, the length, and if the items inside the tied group contribute to agreement. So when you give more equal weight to each rank the differences would cancel each other out.

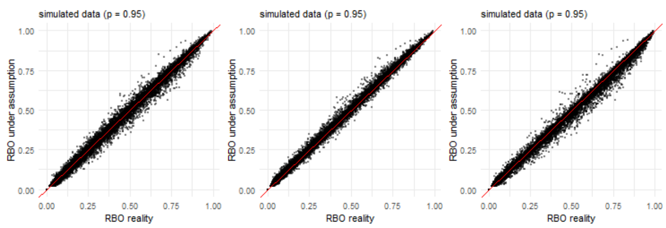
The difference in p-value we mentioned earlier can be seen in figure 3, 4, and 5. For each variant, we see that they follow the same trend. They get to a maximum at around p = 0.8 and then the difference in RBO drops to 0. We see that as mentioned above the a-variance performs better regardless of what p value is chosen. Between the w-variant and b-variant, we see that the b-variant starts slightly higher than the w-variant but goes to almost the same maximum at p around 0.8.

In table 4, we see the combination of figures 1 through 5. Here we can clearly see that the a-variant, on average, performs better at compensating for the assumption. The main thing to show with this table is the maximum differences. Although the average difference is also important it is really small compared to the maximum. Where the average is three numbers after the decimal point, the maximum is one number after the decimal point. Keep in mind, on a domain for RBO between 0 and 1 this is a big difference.



(a) w-variant (b) a-variant (c) b-variant

Fig. 1. RBO with p = 0.8 of full ranking and full ranking with no ties after truncation depth



(a) w-variant (b) a-variant (c) b-variant

Fig. 2. RBO with p = 0.95 of full ranking and full ranking with no ties after truncation depth

p	W-variant Avg.	W-variant Max.	B-variant Avg.	B-variant Max.	A-variant Avg.	A-variant Max.
0.5	0.007	0.520	0.008	0.564	0.004	0.611
0.8	0.009	0.370	0.009	0.350	0.006	0.302
0.9	0.007	0.267	0.008	0.248	0.005	0.205
0.95	0.005	0.162	0.005	0.156	0.003	0.142

Table 4. Shows the average difference in RBO between full ranking and ranking with no ties after a certain depth. It also shows the maximum difference measured for each variant.

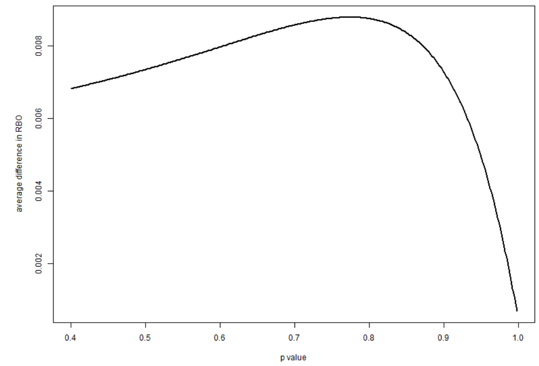


Fig. 3. difference in RBO using w-variant of full ranking and full ranking with no ties truncation depth based on P

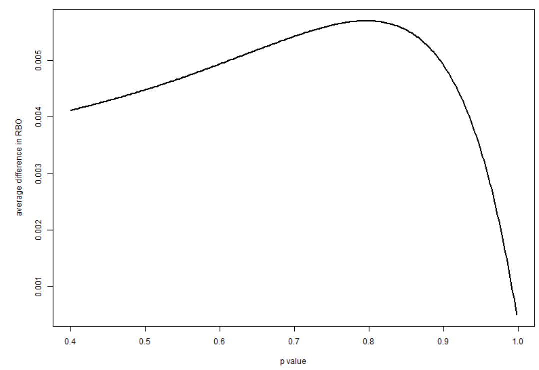


Fig. 4. difference in RBO using a-variant of full ranking and full ranking with no ties truncation depth based on P

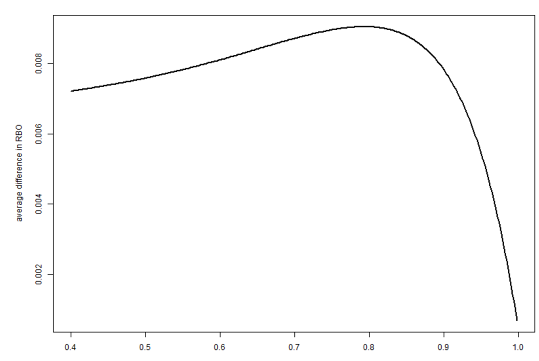


Fig. 5. difference in RBO using b-variant of full ranking and full ranking with no ties truncation depth based on P

4.2 Extrapolation experiments

4.2.1 simulation. Our general way of simulating rankings is explained above. For the specific purpose of showing how the number of unseen items, and therefore more items that could get untied, influence the assumption, the simulation is reshaped. We will use larger rankings now both with length 50 and ties enforced in the last 10% of each ranking.

Then for each percentage ranging from 10 to 75, we enforce our truncated depth is sampled such that when having a pair of rankings the percentage of unseen items is equal to the percentage we want to calculate our RBO with and then show in a plot.

In the end, we have 65 files where each full ranking is the same and the only difference lies in how many items are unseen.

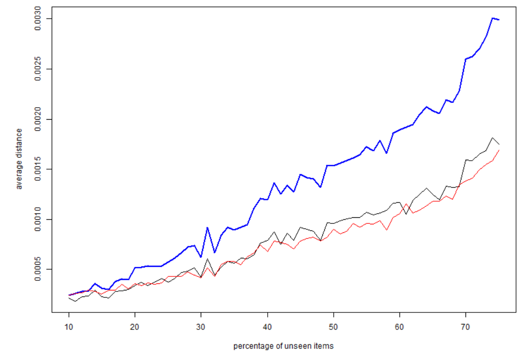
4.2.2 results. In section 3.5, we mentioned that the current method of extrapolating is only affected by the assumption under some circumstances. In figure 6 the blue line (difference in RBO between table 1 and 3) represents how the current extrapolation performs. The red line (difference in RBO between table 1 and 2) represents how with a perfect extrapolation, it would perfectly estimate if an item at a certain rank would match an item in the other ranking, only the assumption affects RBO. The black line (difference in RBO between table 2 and 3) represents how the loss of information, not knowing which items are in the unseen part, affects the RBO, not factoring in ties but only individual items.

We see that as suspected the blue line shows the biggest difference and the red line the lowest. This shows that with RBO, extrapolation is still its biggest restriction. This does not mean however that we should not regard the assumption made as not important. As can be seen by comparing the blue and black lines. The difference between these shows how the assumption affects the extrapolation as for the blue line ties may be present around the truncation depth and with the black line, we know with certainty, because of the way the rankings are generated, that no ties exist at the truncation depth. This neatly shows what we talked about in section 3.5.

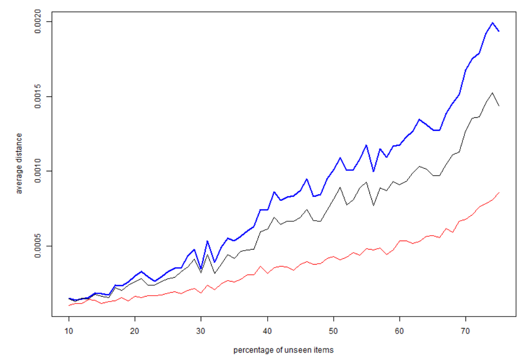
5 RESPONSIBLE RESEARCH

The subject of this paper is a rank similarity measure. The paper is not proposing a new technique or a new way of calculating the similarity. It is researching how an assumption made by others is influencing similarity. The data used in this paper is simulated so no privacy of any individuals is violated. All data is fully reproducible as an extensive explanation of how we came about the data is given.

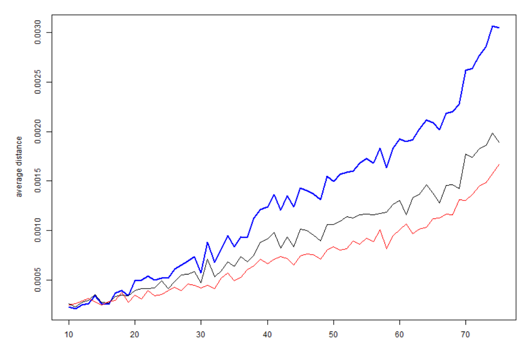
This paper is based on a very specific part of RBO namely ties. What is explained and shown throughout this paper might not correspond to when RBO is used in a different context. When using results from this paper one should fully understand the specific part of RBO that was used.



(a) w-variant



(b) a-variant



(c) b-variant

Fig. 6. Shows the difference in RBO based on the percentage of unseen items. The red line represents the average difference between full ranking and full ranking with no ties after truncation depth. The black line represents the average difference between truncated ranking and full ranking with no ties after truncation depth. The blue line shows the average difference between full ranking and truncated ranking

6 CONCLUSION AND FUTURE WORK

Throughout this paper, a lot of aspects of RBO were discussed. How the different variants perform. How well the extrapolation performs with ties. Most importantly how the assumption affects these variants and the extrapolation. When RBO was first introduced by Webber et al. [3] not much attention was given to ties as RBO in itself was an innovative way of measuring similarity. It was Urbano and Corsi [1] who researched ties to find two variants to handle them in a new way.

We believe as shown through this paper that the three variants do perform well in compensating for the assumption. Performing well does not mean perfect however as shown in chapter 4. We do think that with the information presented in this study, a solution could be found. Whether to reshape the variants so they compensate for the difference or to reshape the current extrapolation to factor in possible ties at the truncation depth.

In our opinion, the main focus in future work should lie on extrapolation. Not only to give a better estimation of the agreement after the truncation depth but to make up for possible ties at that depth, as talked about in section 3.5. This is also shown in section 4.2, where not only the loss of information creates a big difference.

`Lukas Roels`
`lroels@tudelft.nl`

REFERENCES

[1] Matteo Corsi and Julian Urbano. "The Treatment of Ties in Rank-Biased Overlap". In: *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '24), July 14–18, 2024, Washington, DC, USA. ACM, New York, NY, USA,* (2024). URL: https://doi.org/10.1145/3626772.3657700.

[2] Maurice G. Kendall. "A New Measure of Rank Correlation". In: *Biometrika* 30.1 (1938), pp. 81–93.

[3] William Webber Alistair Moffat and Justin Zobel. "A Similarity Measure for Indefinite Rankings". In: *ACM Transactions on Information Systems* 28.4 (2010), pp. 1–38.

[4] Alessia Sarica, Andrea Quattrone, and Aldo Quattrone. "Introducing the Rank-Biased Overlap as Similarity Measure for Feature Importance in Explainable Machine Learning: A Case Study on Parkinson's Disease". In: *Brain Informatics*. Ed. by Mufti Mahmud, Jing He, Stefano Vassanelli, André van Zundert, and Ning Zhong. Cham: Springer International Publishing, 2022, pp. 129–139. ISBN: 978-3-031-15037-1.

[5] Charles Spearman. "The Proof and Measurement of Association between Two Things". In: *The American Journal of Psychology* 15.1 (1904), pp. 72–101.