UNIVERSITY OF MANNHEIM

Evaluating Group Fairness Metrics for Rankings

Marlene Lutz, Tobias Schumacher, Sandipan Sikdar and Markus Strohmaier

8th International Conference on Computational Social Science

Ranking applications

Hiring

Admission

Lending

E-commerce

...

Algorithmic Ranking

- Ranking algorithms assist in decisions that impact peoples wellbeing and success
- Rankings should not only be accurate but also **fair**
- Growing body of research for designing and deploying **fairness metrics** for rankings

How can fairness metrics for rankings be compared and evaluated?

Marlene Lutz
Marlene.lutz@uni-mannheim.de

Contribution

1

Proposal of 13 properties

We propose properties that are informative for the construction and evaluation of group fairness metrics for rankings.

2

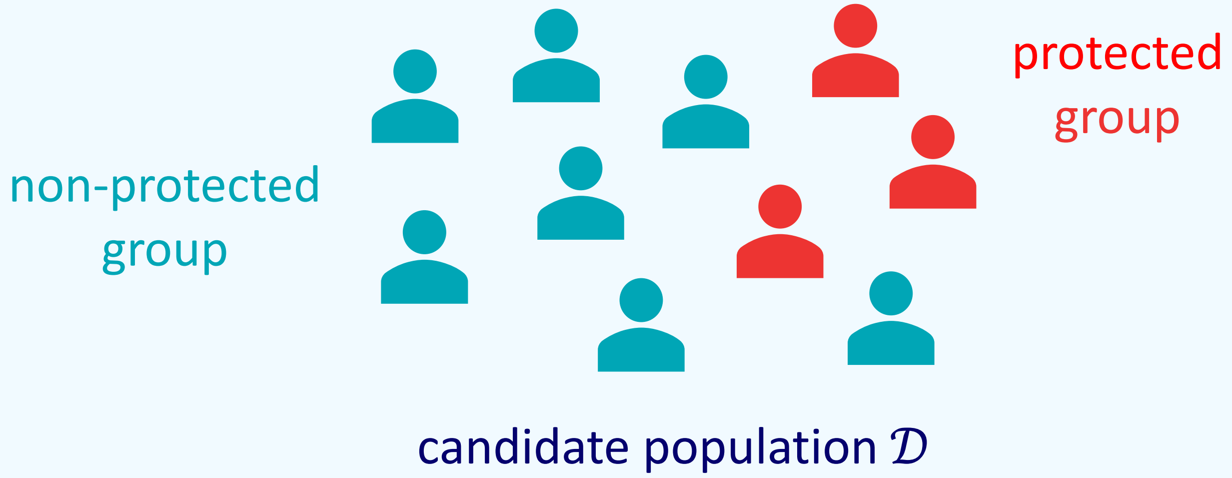
Application to fairness metrics

We deploy the properties to 10 existing group fairness metrics for rankings and study the extent to which they satisfy them.

Marlene Lutz
Marlene.lutz@uni-mannheim.de

Fair Ranking Setup

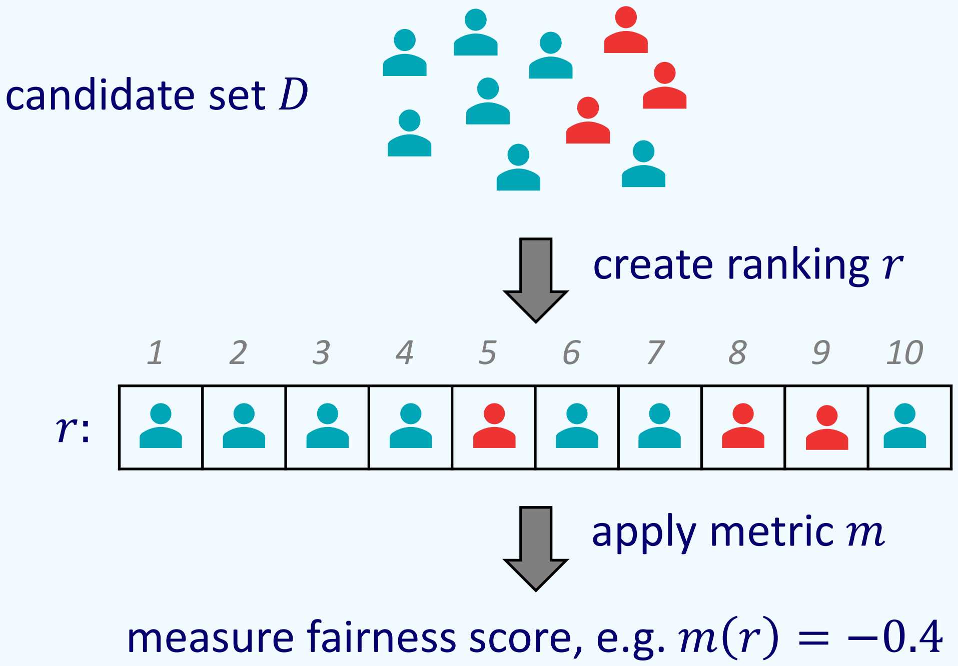
Goal: Ranking a set of candidates $D \subseteq \mathcal{D}$ s.t. the ranking r is fair with respect to a protected group (*group fairness*).



Marlene Lutz
Marlene.lutz@uni-mannheim.de

Fair Ranking Setup

Goal: Ranking a set of candidates $D \subseteq \mathcal{D}$ s.t. the ranking r is fair with respect to a protected group (*group fairness*).



Higher fairness score is better

Marlene Lutz
Marlene.lutz@uni-mannheim.de

Properties for Fair Ranking Metrics

P1: Group Distinctiveness

P2: Boundedness

P3: Monotonicity

P4: Deepness

P5: Intra-group Fairness

P6: Invariance to Linear Transformation of Relevance Scores

Universal Properties
(both ranking settings)

P7: Optimality of Random Rankings

P8: Invariance to Ranking Length

P9: Invariance to Group Proportions

P10: Symmetric Penalties for all Groups

Ranking the full popul.

P11: Deepness Threshold

P12: Closeness Threshold

P13: Confidence

Ranking a subset

Marlene Lutz
Marlene.lutz@uni-mannheim.de

Properties for Fair Ranking Metrics

P1: Group Distinctiveness

P2: Boundedness

P3: Monotonicity

P4: Deepness

P5: Intra-group Fairness

P6: Invariance to Linear Transformation of Relevance Scores

Universal Properties
(both ranking settings)

P7: Optimality of Random Rankings

P8: Invariance to Ranking Length

P9: Invariance to Group Proportions

P10: Symmetric Penalties for all Groups

Ranking the full popul.

P11: Deepness Threshold

P12: Closeness Threshold

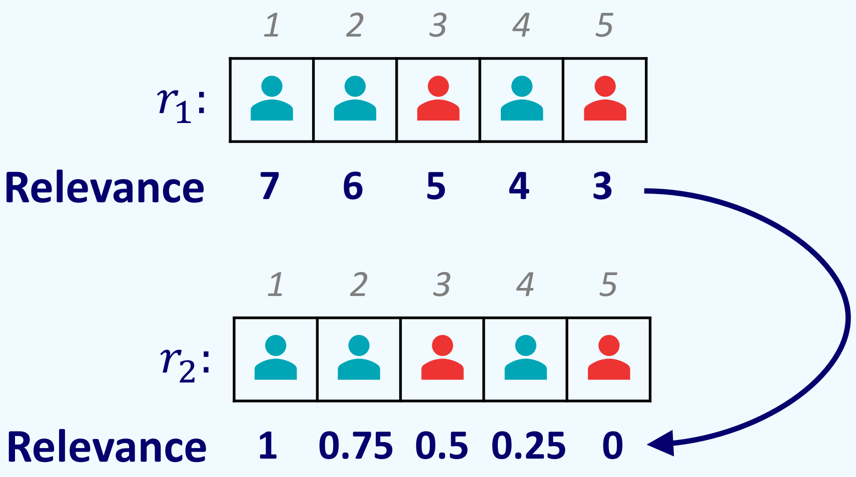
P13: Confidence

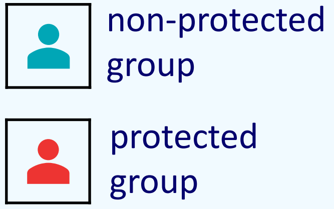
Ranking a subset

Marlene Lutz
Marlene.lutz@uni-mannheim.de

Invariance to Linear Transformation of Relevance Scores

A metric m is **invariant to linear transformation of relevance scores** if its values do not change after transforming the rel. scores of a candidate set.





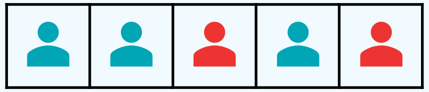
Linear transformation
e.g. min-max normalization

$$\boxed{m(r_1) \overset{!}{=} m(r_2)}$$

Marlene Lutz
Marlene.lutz@uni-mannheim.de

Invariance to Linear Transformation of Relevance Scores

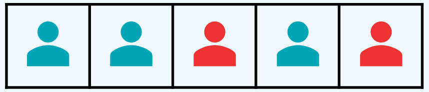
Example: Only transformed relevance scores are accessible (e.g. for privacy reasons).

r_1: 

	1	2	3	4	5
Relevance (true)	**7**	**6**	**5**	**4**	**3**

Fairness:

$DTD(r_1) = \boxed{-0.01}$ in favour of non-protected group

r_2: 

	1	2	3	4	5
Relevance (transformed)	**1**	**0.75**	**0.5**	**0.25**	**0**

Fairness:

$DTD(r_2) = \boxed{0.74}$ in favour of protected group

 non-protected group

 protected group

DTD Disparate treatment difference
0 - perfect fairness

Marlene Lutz
Marlene.lutz@uni-mannheim.de

Invariance to Ranking Length

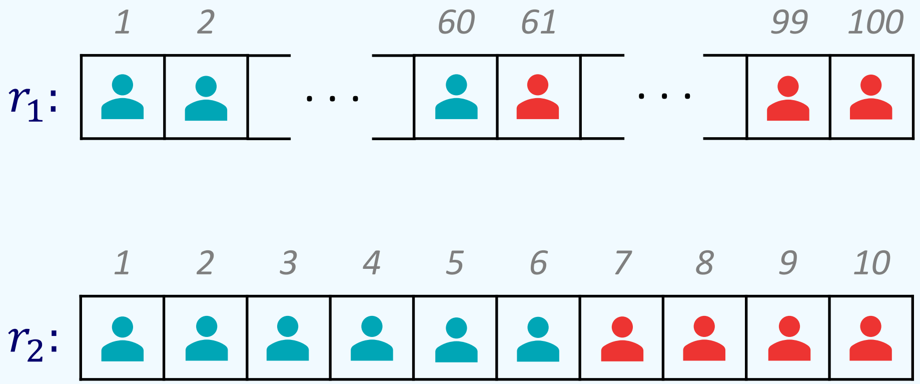
A metric m is **invariant to ranking length** if its *worst-case values* (total disadvantage of one group) do not change for different ranking lengths.

$$\boxed{m(r_1) \stackrel{!}{=} m(r_2)}$$

Marlene Lutz
Marlene.lutz@uni-mannheim.de

Invariance to Ranking Length

Example: Comparing the fairness of hiring processes (= rankings of job applicants) at different companies.



$ED(r_1) = -0.08$
e.g. a job at **PopularCompany**

$ED(r_2) = -0.24$
e.g. a job at **UnpopularCompany**

non-protected group (60%)

protected group (40%)

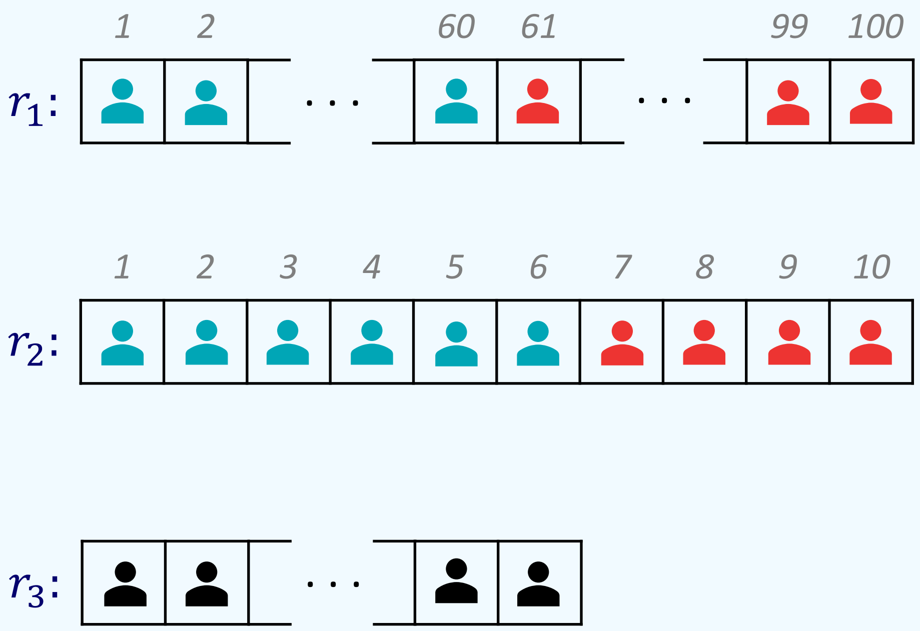
ED Exposure difference
0 - perfect fairness

Higher fairness score is better

Marlene Lutz
Marlene.lutz@uni-mannheim.de

Invariance to Ranking Length

Example: Comparing the fairness of hiring processes (= rankings of job applicants) at different companies.



$ED(r_1) = -0.08$
e.g. a job at **PopularCompany**

$ED(r_2) = -0.24$
e.g. a job at **UnpopularCompany**

$ED(r_3) = -0.07$
How can this value be interpreted?

non-protected group (60%)

protected group (40%)

ED Exposure difference, 0 - perfect fairness

Higher fairness score is better

Marlene Lutz
Marlene.lutz@uni-mannheim.de

Application to Fair Ranking Metrics

P6: Invariance to Linear Transformation of Relevance Scores

P8: Invariance to Ranking Length

Metric	P 1	P 2	P 3	P 4	P 5	P 6	P 7	P 8	P 9	P 10	P 11	P 12	P 13
Normalized discounted difference (rND)	×	✓	×	×	N/A	N/A	×	×	×	×	×	×	×
Normalized discounted KL-divergence (rKL)	×	✓	×	×	N/A	N/A	×	×	×	×	×	×	×
Normalized discounted ratio (rDR)	×	✓	×	×	N/A	N/A	×	×	×	×	×	×	×
Exposure difference (ED)	✓	✓	✓	✓	N/A	N/A	✓	×	×	×	✓	✓	✓
Exposure ratio (ER)	✓	×	✓	✓	N/A	N/A	×	×	×	×	✓	✓	✓
Disparate treatment difference (DTD)	✓	×	✓	✓	×	×	✓	×	×	×	✓	✓	✓
Disparate treatment ratio (DTR)	✓	×	✓	✓	×	×	×	×	×	×	✓	✓	✓
Disparate impact difference (DID)	✓	×	✓	✓	×	×	✓	×	×	×	✓	✓	✓
Disparate impact ratio (DIR)	✓	×	✓	✓	×	×	×	×	×	×	✓	✓	✓
Pairwise statistical parity (PSP)	✓	✓	✓	×	N/A	N/A	✓	✓	✓	✓	N/A	N/A	N/A

Universal Properties — Ranking full popul. — Ranking a subset

✓ property satisfied
× property not satisfied
N/A property not applicable

Marlene Lutz
Marlene.lutz@uni-mannheim.de

Conclusion

Not every application requires satisfaction of all properties!

- Highlight limitations of existing metrics
  - Lack of interpretability and comparability
  - Unexpected side effects
- Support informed evaluation and design of group fairness metrics for rankings
- Guide practitioners in choosing appropriate metrics

Marlene Lutz
Marlene.lutz@uni-mannheim.de

Thank You!

Contact:

Marlene Lutz – marlene.lutz@uni-mannheim.de

Tobias Schumacher – tobias.schumacher@uni-mannheim.de