



Measuring Fairness in Ranked Results: An Analytical and Empirical Comparison

Amifa Raj and Michael D. Ekstrand
People and Information Research Team, Boise State University



People and
Information
Research
Team



grant IIS 17-51278

Search bar with 'math books' and search button. Filter options: all, title, author, genre.

Books Groups Quotes Stories People Listopia Trivia

Page 1 of about 4329 results (0.07 seconds)

- Short-Cut Math (Dover Books on Mathematics) by Gerard W. Kelly
Math Coloring Books for Kids: Coloring Book for Minecrafters: The Best Relaxing Color by Number Pixel Art Game...
My Math Books: Think Math! by Harcourt School Publishers



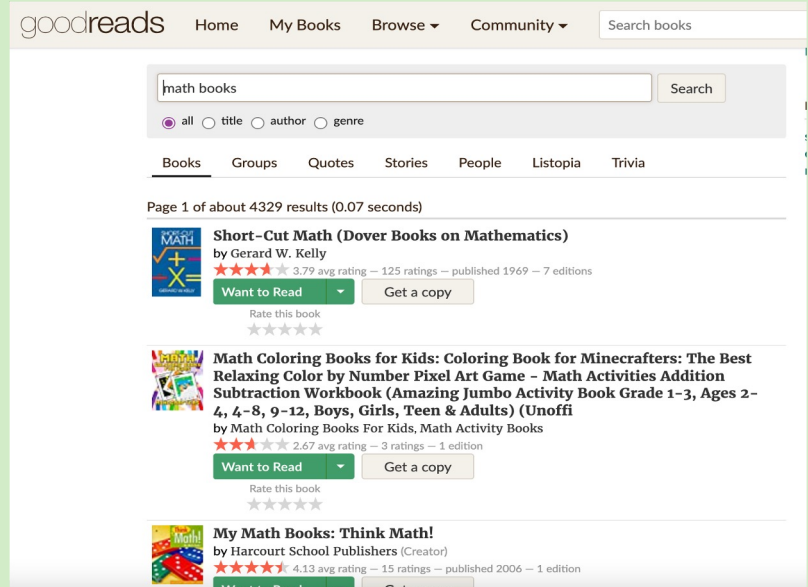
auto-rickshaw-wall-painting



NETFLIX Home TV Shows Movies Latest My List
Trending Now
the office
STAR TREK THE NEXT GENERATION
ANCIENT ALIENS NEW EPISODES



Bias in Ranking



Disparate Exposure

Fair Ranking Metrics Resources

Measuring Fairness in Ranked Output

Yang et. al.
SSDBM'17
PreFA

FA*IR: A Fair Top-k Ranking Algorithm

Zhelike et. al.
CIKM'17
FAIR

Fairness of Exposure in Rankings

Singh and Joachims
KDD'18
D* (DP, EUR, RUR)

Equity of Attention: Amortizing Individual Fairness in Rankings

Biega et. al.
SIGIR'18
IAA

Quantifying the Impact of User Attention Fair Group Representation in Ranked List

Sapiezynski et. al.
WWW'19
AWRF

Evaluating Stochastic Rankings with Expected Exposure

Diaz et. al.
CIKM'20
EE* (EEL, EER, EED)

Why the Problem is a Problem!



**Several Fair
Ranking Metrics**



**Finding Suitable
Metrics**



**Differences among
the Metrics**

Contributions

Describe and compare rank-fairness metrics in unified framework

Identify gaps between their original presentation and the practicalities of applying them to IR systems

Direct comparison of their outcomes with the same data and experimental setting

Sensitivity analysis to assess the impact of design choices and external factors on these metrics

Fairness Positioning



Consumer

Group

**Provider
Fairness**



Individual

Provider

Group Fairness

Factors We Considered

Fairness Goal

**System-
Target
Comparison**

Ratio-based
Distance function

Relevance

**Weighting
Strategy**

**Group
Membership**

Soft Association
Multinomial
Distribution

Fairness Definition

PreF Δ , FAIR, AWRF, DP, EED

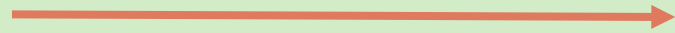
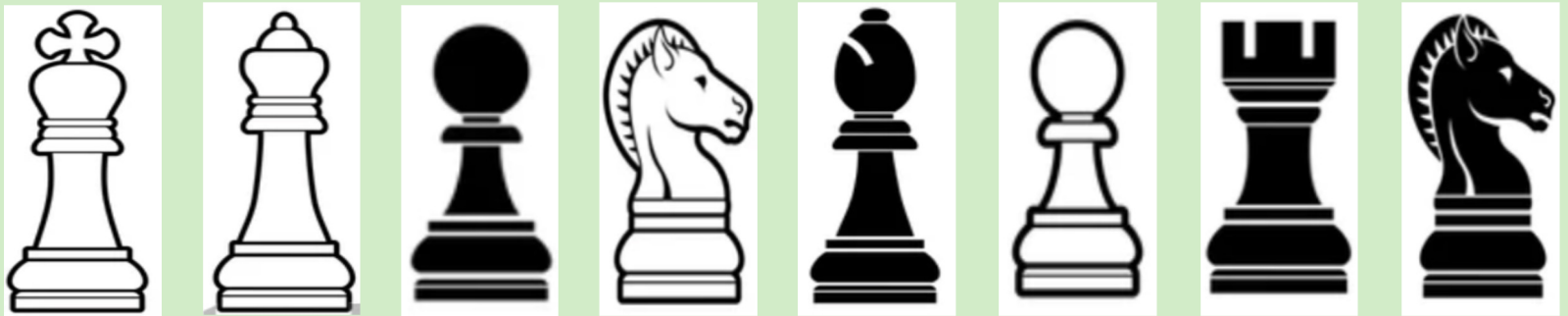
Item position should not be affected by membership

Statistical
Parity

IAA, EUR, RUR, EEL, EER

Exposure/attention should be proportional to **relevance**

Equal
Opportunity

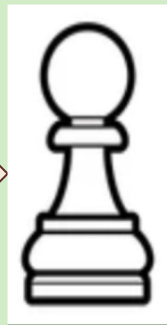


Low order

Ranked List

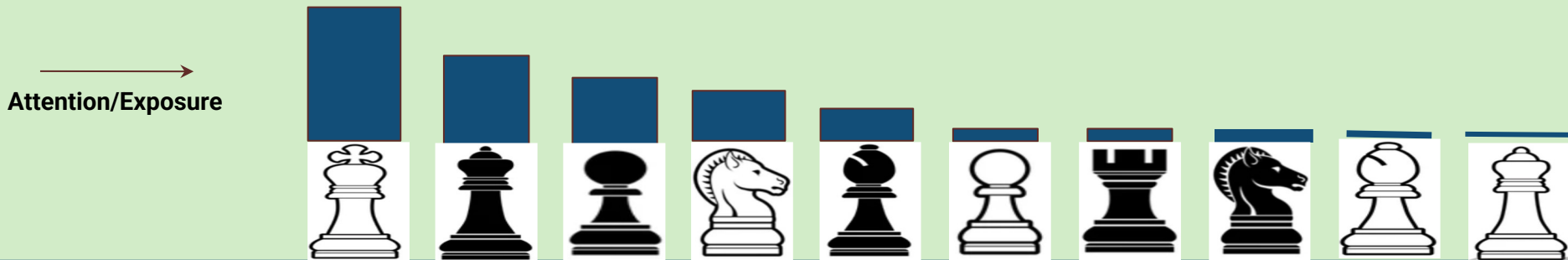


Protected Group




Non-Protected Group

Statistical Parity (Single Ranking)



AWRF (Sapienzyński et. al)

Expected cumulative exposure( x position weight) \wedge $\geq p$

Target distribution is the group distribution in entire ranked list (true demographics)

- no relevance information
- geometric attention decay
- non-binary group membership
- uses a target distribution to compare

PreFA (Yang et. al) and FAIR (Zehlike et. al) do not use position weight

Sequence of Ranking



•
•
•

Statistical Parity (Multiple Ranking)

Demographic Parity (Singh et.al)

Exp



=Exp



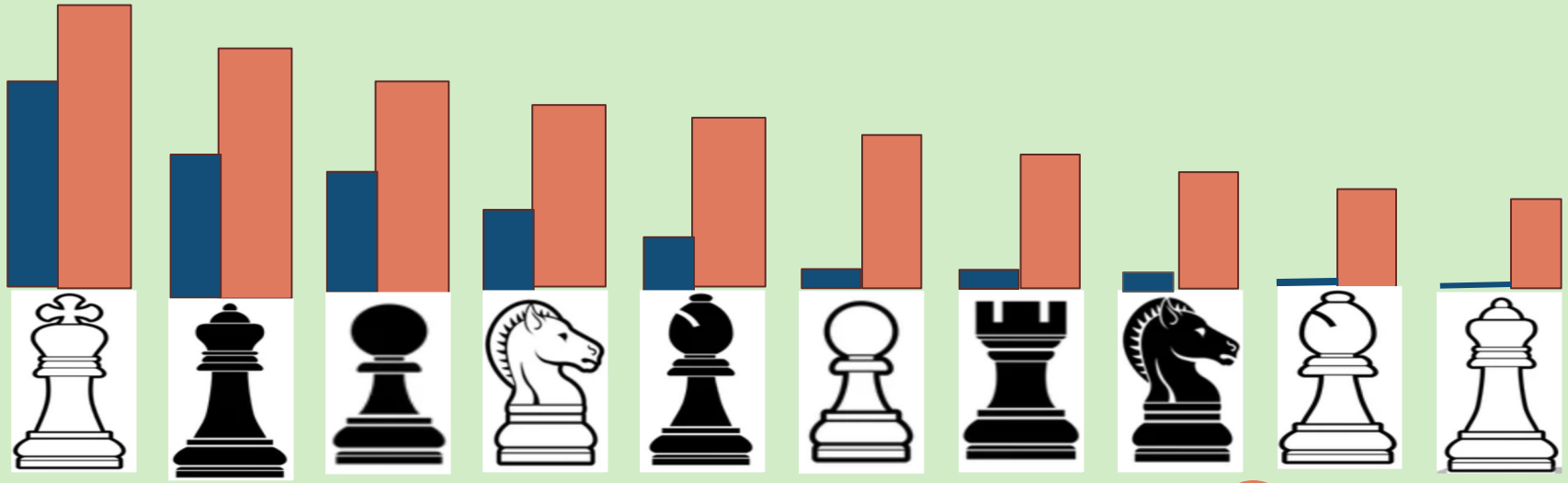
- logarithmic attention decay
- binary group membership

Expected-Exposure Disparity (Diaz et.al)

EED: Demographic Parity

- rbp & cascade attention decay
- non-binary group membership

Equal Opportunity



— relevance
— exposure

-
-
-

Exposure should be proportional to **relevance**

Equal Opportunity

EE* (Diaz et. al)

- stochastic ranking
- rbp & cascade attention decay
- non-binary group membership

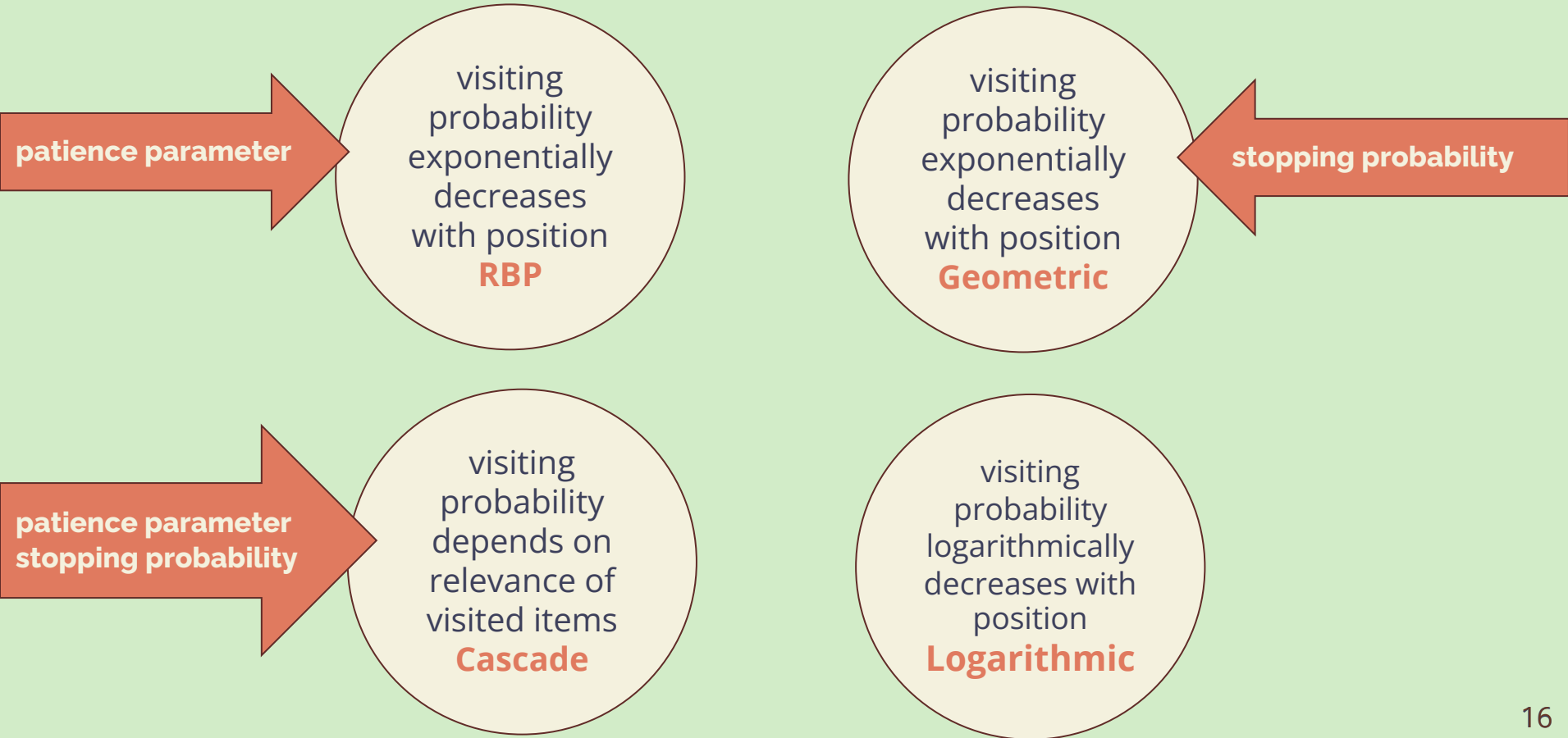
EEL(Expected Exposure Loss):
 $||\text{target-system}||_2$

EER (Expected Exposure
Relevance): Exposure-relevance
distribution

IAA (Biega et. al) differs in weighting strategy,
group membership, and relevance

EUR, RUR (Singh et. al) differs in weighting
strategy and group membership

Browsing Model (Weighting Strategy)



Summary of Fair Ranking Metrics

Metric(s)	Goal	Weighting	Relevance	Binomial?
PreFd	Each prefix representative of whole ranking	✗	✗	Dep on d
FAIR	Each prefix matches target distribution	✗	✗	✓
AWRF	Weighted representation matches population	Geometric	✗	✗
DP	Exposure equal across groups	Logarithmic	✗	✓
DTR	Exposure proportional to relevance	Logarithmic	✓	✓
DIR	Discounted gain proportional to relevance	Logarithmic	✓	✓
IAA	Exposure proportional to predicted relevance	Geometric	Predicted	✗
EEL, EER	Exposure matches ideal (from relevance)	Cascade, Geom	✓	✗
EED	Exposure well-distributed	Cascade, Geom	✗	✗

Implementing the Metrics

	Recommendations	Search (retrieval and re-ranking)
Dataset	GoodReads bookdata	FairTREC 2020
Sensitive Attributes	Gender of author	Economic development of the author's country of scholarly articles
Algorithms	CF (implicit feedback)	Participants provided

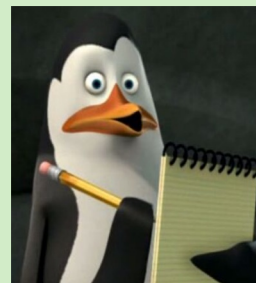
Challenges in Implementation



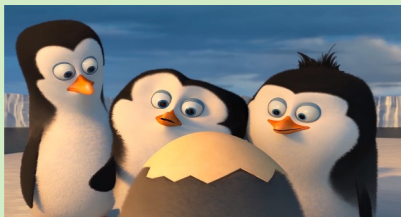
Missing Relevance Information

IAA, EE*, DP, EUR, RUR

AWRF, IAA, DP, EUR, RUR, EE*



Parameter Setting



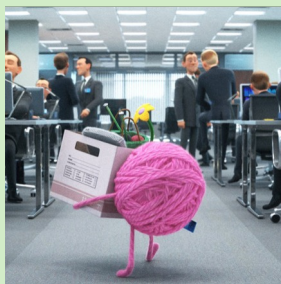
Missing Group Label

All the metrics

PreFA, FAIR, IAA, DP, EUR, RUR



Soft Group Association
Non-binary groups

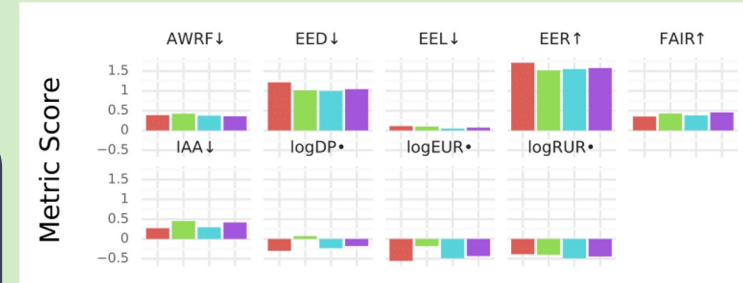


Extreme Imbalance

- **PreFA** and **RUR**: suffer from missing data (sparsity) problem
- Reformulated ratio-based metric to smoothed log ratio

Direct Comparison

- Metrics frequently disagree on system orderings.
- No clear agreement.
- The most consistently-agreeing pair is FAIR and AWRFA



Sensitivity Analysis

Ranked-list size

- No effect on metrics for FairTREC
- Ratio-based metrics and FAIR showed sensitivity

Weighting Strategy

- Default parameters
- EEL and logRUR showed high sensitivity

Stopping Probability

- Almost all metrics showed sensitivity
- logRUR is extremely sensitive

Patience Parameter

- logRUR showed high sensitivity

Key Findings

Defining metrics in unified framework

- Metrics are surprisingly similar

Implement the metrics in same experimental setup

- Missing data, missing relevance information, ranked list size are crucial/delicate factors in implementing metrics.

Sensitivity Analysis

- Metrics differ in their sensitivity towards external factors.
- High sensitivity towards design choices add complexity in the usability of metrics

Recommendations

	Allow multinomial protected attributes	Allow soft group association	Sensitivity towards design choices
Single-list metrics FAIR, AWRF	AWRF	AWRF	AWRF
Demographic Parity in Sequence DP, EED	EED	EED	EED
Equal Opportunity in Sequence EUR, RUR, IAA, EER, EEL	EER, EEL	EER, EEL	EER, EEL, IAA

Conclusion

Future Work

- Missing label
- Missing or sparse relevance
- Ambiguous or multiple group association
- Robust, explainable, and efficient metric design
- Simulation study to understand the impact of crucial factors in metric implementation.



THANK You!

Credits: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**