Reproducibility Report for ACM SIGMOD 2022 Paper: “Rank Aggregation with Proportionate Fairness”

FATEMEH NARGESIAN, University of Rochester, USA

The paper studies the classical rank aggregation problem subject to proportionate fairness or p-fairness for the case of multiple individual rank orders over a set of items, where the candidates belong to multiple (non-binary) protected groups and Kemeny distance is considered. The paper proposes three solutions: GrBinaryIPF, ApproxMultiValuedIPF, AlgRAPF, RandAlgRAPF. The experiments use real-world synthetic datasets and compare the proposed solutions with the state-of-the-art techniques: FairILP [2] and DetConstSort [1].

1 INTRODUCTION

This report summarizes the reproducibility evaluation for the paper entitled “Rank Aggregation with Proportionate Fairness” by Dong Wei, Md Mouinul Islam, Baruch Schieber, and Senjuti Basu Roy, from New Jersey Institute of Technology [3]. The paper appeared in the Proceedings of the 2022 ACM SIGMOD International Conference on Management of Data (SIGMOD’22). The experiments are reproducible and support the key findings of the paper. All experiments provide performance numbers similar or non-contradictory to the paper except for one experiment involving random data input.

2 SUBMISSION

The repository containing scripts and reproducibility instructions for this paper is available at: https://github.com/MdMouinulIslam/RankAggregationProportionate. The *graphs* directory contains a Jupyter notebook for each plot in the paper. The *data* directory contains the datasets used in the experiments, in csv format. The implementation of algorithms and baselines can be found in the *codes* directory. The readme file under the main *RankAggregationProportionate* contains a short description on how to prepare data and how to run algorithms.

3 HARDWARE AND SOFTWARE ENVIRONMENT

Table 1. Hardware & Software environment

	Paper	Repro Review
CPU	Intel E3-1245	Intel Xeon Gold 5218
GHz	3.70	2.30
RAM	32GB	512GB
Python	3.8	3.8

4 REPRODUCIBILITY EVALUATION

4.1 Process

We followed the step-by-step instructions provided in the reproducibility repository. We installed the required libraries, listed in requirements.txt, using pip install. Installing Gurobi requires obtaining a license file, if on pc, from https://www.gurobi.com/. For reproducing the plots, We ran the Jupyter notebook dedicated to each plot. There were some minor errors in the naming of dataset files referred in the code, which were resolved after some rounds of interactions with the authors. Apart from that, the whole reproducibility process went smoothly and without any issues.

4.2 Results

The conclusions of the reproducibility evaluation are as follows.

- **Figure 1: Percentage of positions satisfying p-fairness (IPF)**
  - **(a): Fantasy football: GrBinaryIPF vs. DetConstSort:** This experiment is partially reproducible. Although the numbers for *DetconstSort* do not exactly match the plot in the paper, the results do not violate the main point.
  - **(b): German Credit: ApproxMultiValuedIPF vs DetConstSort:** The results are fully reproducible.
  - (c): MovieLens: ApproxMultiValuedIPF vs DetConstSort: The results are fully reproducible.
- **Figure 2: Percentage of groups satisfying p-fairness (IPF)**
  - **(a) Fantasy football: GrBinaryIPF vs. DetConstSort:** The results are fully reproducible.
  - **(b) German Credit: ApproxMultiValuedIPF vs DetConstSort:** The results are fully reproducible.
  - **(c) MovieLens: ApproxMultiValuedIPF vs DetConstSort:** The results are fully reproducible.
- **Figure 3: Kendall-Tau distance IPF**
  - **(a) Fantasy football: OptIPF, GrBinaryIPF, DetConstSort:** The results for GrBinaryIPF and OptIPF are not consistent with the paper. Det-ConstSort can sometimes compute a ranking with a smaller distance than GrBinaryIPF and OptIPF and the paper shows such instance. This did not happened during multiple reproducibility runs.
  - **(b) German Credit: ApproxMultiValuedIPF vs. DetConstSort:** The results are fully reproducible.
  - **(c) MovieLens: ApproxMultiValuedIPF vs. DetConstSort:** The results are fully reproducible.
- **Figure 4: Varying δ analysis IPF**
  - **Fantasy Football: GrBinaryIPF** The results are fully reproducible.
  - **German Credit: ApproxMultiValuedIPF:** The results are fully reproducible.
  - **Movie Lens: ApproxMultiValuedIPF:** The results are fully reproducible.
- **Figure 5: Running time analysis of IPF**
  - **GrBinaryIPF:** The results are fully reproducible.
  - **ApproxMultiValuedIPF:** The results are fully reproducible.
- **Figure 6: % of positions satisfying p-fairness (RAPF)**
  - **(a) Fantasy football: p-fairness: AlgRAPF vs. FairILP:** This experiment is partially reproducible. Although the numbers for *FairILP* do not exactly match the plot in the paper, the results do not violate the main point.
  - **(b) MovieLens: p-fairness: AlgRAPF vs. FairILP:** This experiment is partially reproducible. Although the numbers for *FairILP* do not exactly match the plot in the paper, the results do not violate the main point.
- **Figure 7: Kemeny Distance RAPF**
  - **(a) Fantasy football: AlgRAPF vs. RandAlgRAPF vs. OptRA vs. FairILP:** The results are fully reproducible.
  - **(b) MovieLens: AlgRAPF vs. RandAlgRAPF vs. OptRA vs. FairILP:** The results are fully reproducible.
- **Figure 8: Running time analysis**
  - **(a) Vary $n, m = 100$: RandAlgRAPF:** The results are fully reproducible.

- **(b) Vary** $m, n = 1000$**: RandAlgRAPF:** The results are fully reproducible.
- **(c) Vary** $n, m = 100$**: AlgRAPF:** The results are fully reproducible.
- **(d) Vary** $m, n = 1000$**: AlgRAPF:** The results are fully reproducible.

- **Figure 9: Varying δ analysis RAPF**
  - **Fantasy Football : AlgRAPF vs. RandAlgRAPF varying δ:** The results are fully reproducible.
  - **Movie Lens : AlgRAPF vs. RandAlgRAPF varying δ:** The results are fully reproducible.

5 SUMMARY

Overall most paper claims have been reproduced with minimal effort. This deemed results reproducible.

REFERENCES

[1] Sahin Cem Geyik, Stuart Ambler, and Krishnaram Kenthapadi. 2019. Fairness-Aware Ranking in Search & Recommendation Systems with Application to LinkedIn Talent Search. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2019, Anchorage, AK, USA, August 4-8, 2019*, Ankur Teredesai, Vipin Kumar, Ying Li, Rómer Rosales, Evimaria Terzi, and George Karypis (Eds.). ACM, 2221–2231. https://doi.org/10.1145/3292500.3330691

[2] Caitlin Kuhlman and Elke A. Rundensteiner. 2020. Rank Aggregation Algorithms for Fair Consensus. *Proc. VLDB Endow.* 13, 11 (2020), 2706–2719. http://www.vldb.org/pvldb/vol13/p2706-kuhlman.pdf

[3] Dong Wei, Md Mouinul Islam, Baruch Schieber, and Senjuti Basu Roy. 2022. Rank Aggregation with Proportionate Fairness. In *SIGMOD ’22: International Conference on Management of Data, Philadelphia, PA, USA, June 12 - 17, 2022*, Zachary G. Ives, Angela Bonifati, and Amr El Abbadi (Eds.). ACM, 262–275. https://doi.org/10.1145/3514221.3517865