

Offline Evaluation Measures of Fairness in Recommender Systems

Theresia Veronika Rampisela

University of Copenhagen

Copenhagen, Denmark

thra@di.ku.dk

Abstract

The evaluation of recommender system fairness has become increasingly important, especially with recent legislation that emphasises the development of fair and responsible artificial intelligence. This increasing importance has led to the emergence of various fairness evaluation measures, which quantify fairness based on different fairness definitions. However, many of such measures are simply proposed and used without further analysis on their robustness. As a result, there is insufficient understanding and awareness of the measures' limitations. Among other issues, it is not known what kind of model outputs produce the (un)fairest score, how the measure scores are empirically distributed, and whether there are cases where the measures cannot be computed (e.g., due to division by zero). These issues cause difficulty in interpreting the measure scores and confusion on which measure(s) should be used for a specific case.

To address the above issues, this thesis presents a series of papers that assess and overcome various theoretical, empirical, and conceptual limitations of existing recommender system fairness evaluation measures. Altogether, the papers in this thesis investigate a wide range of offline evaluation measures for different fairness notions, which are divided based on the evaluation subjects (users and items) and for different evaluation granularities (groups of subjects and individual subjects). Firstly, we perform theoretical and empirical analysis on a set of fairness measures, exposing flaws that limit their interpretability, expressiveness, or applicability. Secondly, we contribute novel evaluation approaches and measures that overcome these limitations. Finally, considering the measures' limitations, we recommend guidelines for the appropriate measure usage, thereby allowing for more precise selection of fairness evaluation measures in practical scenarios.

Overall, the papers in this thesis collectively contribute to advancing the state-of-the-art offline evaluation of fairness in recommender systems.

Awarded by: University of Copenhagen, Copenhagen, Denmark **on** 15 September 2025.

Supervised by: Christina Lioma, Tuukka Ruotsalo, Maria Maistro.

Available at: <https://researchprofiles.ku.dk/en/publications/offline-evaluation-measures-of-fairness-in-recommender-systems/>.

Selected Publications

- Theresia Veronika Rampisela, Maria Maistro, Tuukka Ruotsalo, and Christina Lioma. Evaluation Measures of Individual Item Fairness for Recommender Systems: A Critical Study. *ACM Trans. Recomm. Syst.*, 3(2), November 2024a. doi: 10.1145/3631943. URL <https://doi.org/10.1145/3631943>.
- Theresia Veronika Rampisela, Tuukka Ruotsalo, Maria Maistro, and Christina Lioma. Can We Trust Recommender System Fairness Evaluation? The Role of Fairness and Relevance. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '24, page 271–281, New York, NY, USA, 2024b. Association for Computing Machinery. ISBN 9798400704314. doi: 10.1145/3626772.3657832. URL <https://doi.org/10.1145/3626772.3657832>.
- Theresia Veronika Rampisela, Maria Maistro, Tuukka Ruotsalo, Falk Scholer, and Christina Lioma. Stairway to Fairness: Connecting Group and Individual Fairness. In *Proceedings of the Nineteenth ACM Conference on Recommender Systems*, RecSys '25, page 677–683, New York, NY, USA, 2025a. Association for Computing Machinery. ISBN 9798400713644. doi: 10.1145/3705328.3748031. URL <https://doi.org/10.1145/3705328.3748031>.
- Theresia Veronika Rampisela, Tuukka Ruotsalo, Maria Maistro, and Christina Lioma. Joint Evaluation of Fairness and Relevance in Recommender Systems with Pareto Frontier. In *Proceedings of the ACM on Web Conference 2025*, WWW '25, page 1548–1566, New York, NY, USA, 2025b. Association for Computing Machinery. ISBN 9798400712746. doi: 10.1145/3696410.3714589. URL <https://doi.org/10.1145/3696410.3714589>.
- Theresia Veronika Rampisela, Maria Maistro, Tuukka Ruotsalo, Falk Scholer, and Christina Lioma. Relevance-aware Individual Item Fairness Measures for Recommender Systems: Limitations and Usage Guidelines. *ACM Trans. Recomm. Syst.*, 4(3), March 2026. doi: 10.1145/3765624. URL <https://doi.org/10.1145/3765624>.