Less Data, More Questions: Fairness and Accuracy Under Data Minimization in Recommender Systems*

Salvatore Bufi[1,*,†], Vincenzo Paparella[2,*,†], Vito Walter Anelli[1] and Tommaso Di Noia[1]

[1]*Politecnico di Bari, Bari, Italy*
[2]*ISTI-CNR, Pisa, Italy*

Abstract

Privacy laws now require data minimization, but its broader effects on recommender systems (RS) are still unclear. We systematically study how common minimization techniques reshape the three key RS goals—accuracy, user fairness, and provider fairness. Across multiple datasets and models we (i) measure performance shifts under data minimization strategies, (ii) pinpoint techniques that best balance the three objectives, and (iii) compare model robustness to data reduction. We find that while several strategies improve group-level consumer fairness, they often reduce accuracy and can even worsen provider fairness; the size of these trade-offs strongly depends on the chosen technique and model. Code and data are public at https://github.com/salvatore-bufi/DataMinimizationFairness.

Keywords

Data Minimization, Recommender Systems, Fairness, Multi-Objective Evaluation

1. Introduction

Recommender systems (RSs) form the backbone of many digital platforms, tailoring content in e-commerce, streaming, and social media [34, 25, 13, 23, 27, 8]. Their accuracy, however, is increasingly at odds with growing concerns over privacy, security risks, and the challenges of scalability, due to their reliance on fine-grained user data [17, 15, 5, 29]. Laws such as the GDPR, CCPA/CPRA, and China's PIPL mandate *data minimization*, i.e., retaining only what is strictly necessary [3, 1, 2, 4]. Initial work has examined how data minimization affects accuracy [7], privacy [16], bias [10, 26, 28], and per-user performance [35]. Yet, data minimization reshapes available user-item interactions by discarding or selectively retaining interactions, potentially altering the distribution of users and items. These alterations can amplify existing biases and undermine system fairness from consumer and producer perspectives. Despite the field's growing focus on fairness [38, 12, 31, 30, 6, 37], a rigorous account of data minimization impact on *both* accuracy and fairness is still missing. This study bridges the gap by systematically investigating the interplay between data minimization, accuracy, and fairness in recommendation, specifically, we: (i) **Evaluate** how minimization strategies impact accuracy and fairness on multiple real-world datasets; (ii) **Pinpoint** data minimization strategies that best balance the three objectives via a multi-objective evaluation; (iii) **Investigate** how different model architectures react to data minimization when balancing the competing objectives of accuracy and fairness. Code and data are available at https://github.com/salvatore-bufi/DataMinimizationFairness.

2. Experimental Setup

This research carries out a comprehensive investigation into the effects of data minimization on recommender systems, moving beyond a simple accuracy analysis to consider its broader implications

IIR2025: 15th Italian Information Retrieval Workshop, 3th - 5th September 2025, Cagliari, Italy
*Extended version [9] published at the 33rd ACM Conference on User Modeling, Adaptation and Personalization (UMAP 2025).
*Corresponding authors.
†These authors contributed equally.
s.bufi@phd.poliba.it (S. Bufi); vincenzo.paparella@isti.cnr.it (V. Paparella)

for both users and item providers. In this section, we detail the experimental setup employed to address the following research questions:

RQ1: How does data minimization impact the accuracy, provider fairness, and consumer fairness of recommendations? What are the observable trade-offs at the strategy level?

RQ2: Which data minimization strategies most effectively preserve recommender system performance compared to training on the full dataset?

RQ3: How robust are different recommendation models to data minimization, particularly concerning maintaining a balance between accuracy and fairness?

Data Preparation and Protocol. Experiments rely on two public benchmarks: *MovieLens 1M* (ML1M) [20, 14] and *Ambar*[19]. To make data-minimization effects observable while keeping models trainable, we follow Biega et al. [7]. Specifically, we retain users with at least 45 ratings (and for Ambar, items as well), then randomly sample 2500 users from each dataset.

Following this pre-processing, we implement our evaluation protocol which adapts the methodology from Biega et al. [7] to simulate a real-world user consent scenario, where the availability of data is determined by individual user choices. We randomly select 30% of users from the full dataset to form our experimental group, whose interaction data constitutes D_M. For each user in D_M, we partition their ratings into a candidate set (70%), a validation set (10%), and a test set (20%). The final training data is created by applying various minimization strategies to the candidate set, selecting only n ratings per user, with $n = \{1, 3, 7, 15, 100\}$.

Algorithms and Metrics. Our study evaluates the impact of data minimization on five representative algorithms from distinct methodological families: the graph-based LightGCN [21], neighborhood-based User-kNN [33], matrix factorization (BPRMF [32]), a linear model (EASER [36]), and a variational autoencoder (MultiVAE [24]). This diverse selection ensures our findings are comprehensive across different recommendation approaches. We assess performance across three objectives. We use nDCG [22] to measure accuracy. For fairness, we evaluate consumer fairness with Mean Absolute Deviation (MAD) [11], which quantifies quality disparities between user groups, and provider fairness with Ranking-based Statistical Parity (RSP) [39], which measures how uniformly item categories are exposed. Finally, we use the Hypervolume (HV) [18] to analyze the simultaneous performance on accuracy, consumer, and provider fairness.

Data Minimization Strategies. In our study, we evaluate several approaches to minimize user data. These strategies are designed to select subsets of user-item interactions that serve as input to the system. In this paper, we study the data minimization strategies explored by Biega et al. [7], aiming at broadening their assessment to fairness issues. These include a **Full data** baseline and methods that select n interactions per user based on: **Random selection**; recency (**Most Recent**)[1]; highest/lowest scores (**Most/Least Favorite**); global popularity (**Most Rated**); proximity to the average item profile (**Most Characteristic**); and highest rating variability (**Highest Variance**).

3. Results and Discussion

This section presents the empirical findings that address our three research questions by analyzing the impact of data minimization on accuracy and fairness[2].

Impact of Data Minimization on Key Objectives (RQ1). We find that data minimization significantly impacts recommendation quality, as shown in Table 1. Accuracy consistently drops as fewer interactions are retained (i.e., smaller n), with extreme minimization ($n = 1$) rendering recommendations impractical. This reduction, however, generally improves consumer fairness (lower MAD) by leveling the performance across user groups, albeit at the cost of relevance. The effect on provider fairness (RSP) is more complex. At low n values, the impact is unpredictable, but a clear trade-off with accuracy emerges: strategies that yield higher accuracy, e.g. “Most Rated”, tend to degrade provider fairness by narrowing

[1]The Most Recent strategy was not applied to the Ambar dataset due to a lack of timestamps.

[2]Full results on the ML1M and Ambar datasets are available in the original paper.

Table 1
Performance results of UserKNN and LightGCN on the Ambar dataset for $n = 1$, $n = 7$, and $n = 100$. The percentage difference between each value and the corresponding value obtained with the full strategy is indicated below each result. For each scenario, bold and underline stand for best and second-to-best values, respectively. All results are statistically significant based on paired t-tests ($p < 0.05$), except for the values denoted with †.

Strategy	$n = 1$				$n = 7$				$n = 100$			
	nDCG↑	RSP↓	MAD↓	HV↑	nDCG↑	RSP↓	MAD↓	HV↑	nDCG↑	RSP↓	MAD↓	HV↑
UserKNN												
Random	0.0036 (-97.26)	1.4422 (75.43)	0.0049 (-91.79)	0.0577 (-94.23)	0.0572 (-56.55)	0.7075 (-13.95)	0.0324 (-46.17)	0.8107 (-18.93)	$0.1373^{\dagger}$ (4.29)	0.8199 (-0.27)	0.062 (2.93)	0.7255 (-27.45)
Most Favorite	0.0101 (-92.35)	1.4621 (77.84)	0.0044 (-92.72)	0.0381 (-96.19)	**0.1092** (-17.03)	0.7307 (-11.12)	0.0361 (-40.13)	0.8225 (-17.75)	$0.1404^{\dagger}$ (6.64)	0.8755 (6.5)	0.06 (-0.42)	0.6694 (-33.06)
Least Favorite	0.0055 (-95.84)	1.4863 (80.79)	0.0354 (-41.26)	0.0133 (-98.67)	0.0345 (-73.81)	0.5747 (-30.1)	0.0153 (-74.6)	0.9426 (-5.74)	$0.1362^{\dagger}$ (3.48)	**0.8072** (-1.82)	0.0683 (13.31)	**0.7335** (-26.65)
Most Rated	**0.0164** (-87.55)	1.447 (76.0)	0.0101 (-83.21)	0.0534 (-94.66)	0.0938 (-28.74)	1.0878 (32.31)	0.0525 (-12.91)	0.4272 (-57.28)	$0.1311^{\dagger}$ (-0.4)	0.8177 (-0.53)	**0.0594** (-1.54)	0.7259 (-27.41)
Most Characteristic	0.0016 (-98.75)	1.5222 (85.16)	**0.0005** (-99.18)	0.0 (-100.0)	0.0285 (-78.37)	**0.491** (-40.28)	**0.0077** (-87.17)	**1.0297** (2.97)	**0.1437**† (9.13)	0.8548 (3.97)	0.0707 (17.29)	0.6858 (-31.42)
Highest Variance	0.0015 (-98.83)	**0.9225** (12.21)	0.0007 (-98.84)	**0.578** (-42.2)	0.062 (-52.9)	0.6658 (-19.01)	0.0396 (-34.26)	0.8508 (-14.92)	$0.1318^{\dagger}$ (0.09)	0.8138 (-1.01)	0.0598 (-0.72)	0.7301 (-26.99)
LightGCN												
Random	0.0083 (-93.95)	0.7742 (-9.77)	0.0302 (-46.03)	0.7097 (-29.03)	0.053 (-61.18)	0.7235 (-15.69)	0.0525 (-6.11)	0.7747 (-22.53)	0.1291 (-5.46)	**0.7887** (-8.09)	0.0532 (-4.85)	**0.7604** (-23.96)
Most Favorite	0.0108 (-92.1)	0.5157 (-39.9)	0.0058 (-89.55)	0.9891 (-1.09)	**0.1078** (-21.1)	0.6993 (-18.5)	**0.0231** (-58.74)	0.8665 (-13.35)	0.1287 (-5.73)	0.8545 (-0.42)	**0.0496** (-11.31)	0.6924 (-30.76)
Least Favorite	0.008 (-94.13)	0.6632 (-22.72)	0.0132 (-76.33)	0.8324 (-16.76)	0.0353 (-74.15)	0.6342 (-26.1)	0.1252 (123.74)	0.7842 (-21.58)	**0.134** (-1.9)	0.8338 (-2.84)	0.062 (10.85)	0.7086 (-29.14)
Most Rated	**0.0233** (-82.98)	1.1165 (30.11)	0.0082 (-85.38)	0.3892 (-61.08)	0.0853 (-37.51)	1.1706 (36.42)	0.044 (-21.29)	0.3417 (-65.83)	$0.1278^{\dagger}$ (-6.46)	0.8384 (-2.3)	0.0545 (-2.66)	0.7055 (-29.45)
Most Characteristic	0.001 (-99.26)	**0.3557** (-58.55)	**0.0004** (-99.31)	**1.145** (14.5)	0.0256 (-81.23)	**0.4283** (-50.09)	0.028 (-50.0)	**1.0684** (6.84)	0.1302 (-4.66)	0.8584 (0.04)	0.0584 (4.33)	0.6828 (-31.72)
Highest Variance	0.0047 (-96.59)	0.5715 (-33.4)	0.0026 (-95.35)	0.9304 (-6.96)	0.0587 (-57.02)	0.6896 (-19.64)	0.0317 (-43.4)	0.8308 (-16.92)	0.1244 (-8.88)	0.8497 (-0.98)	0.0544 (-2.84)	0.6915 (-30.85)

item exposure. Conversely, strategies like "Most Characteristic" improve provider fairness but severely compromise accuracy. "Most Favorite" strategy strikes a better balance, achieving competitive accuracy while mitigating the fairness degradation seen in other approaches. These intricate trade-offs are also reflected in the Hypervolume (HV) metric, underscoring that data minimization's effectiveness cannot be judged on accuracy alone. *We conclude that data minimization introduces a critical trade-off: while consumer fairness may improve, it comes at the expense of accuracy and can lead to unpredictable outcomes for provider fairness, requiring nuanced strategies to balance these objectives.*

Performance of Data Minimization Strategies (RQ2). To evaluate how well strategies preserve overall performance compared to using the full dataset, we adopt a multi-objective approach. We represent each outcome as a point in a 3D space (nDCG, RSP, MAD) and measure its Euclidean distance (δ) from the reference point (performance on full data). A smaller δ indicates better preservation. As shown in Figure 1, when n is large, most strategies perform similarly well, but a clear hierarchy emerges as n decreases. The "Most Rated" and "Most Characteristic" strategies consistently exhibit the highest δ values, indicating poor overall performance preservation due to their strong biases toward accuracy and fairness, respectively. In contrast, strategies that introduce variability into user profiles, such as "Most Favorite," and "Highest Variance", achieve lower δ values. *Thus, data minimization strategies that effectively shape user profiles by introducing variability are best at retaining the holistic performance of the recommender system.*

Robustness of Recommender Models (RQ3). To assess model robustness, we analyze the mean (μ) and standard deviation (σ) of the Euclidean distances (δ), across all minimization strategies for a given n, as reported in Table 2. A lower μ signifies better overall robustness, while a lower σ indicates

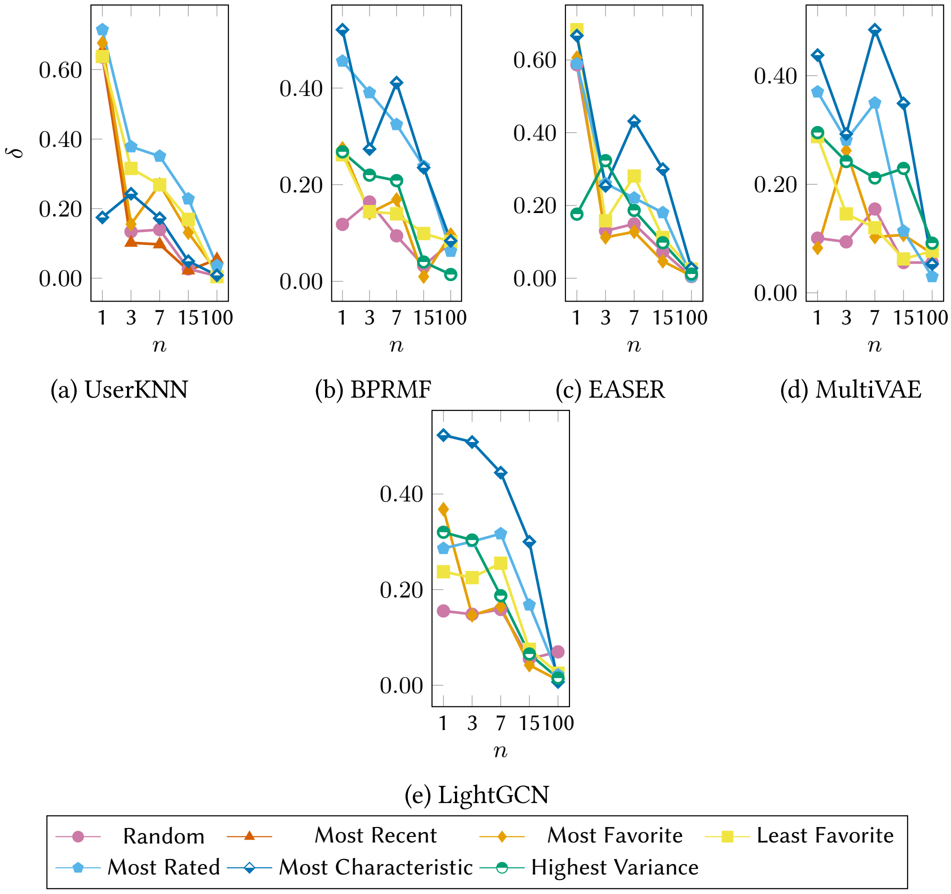

Figure 1: Comparison of the Euclidean distance between model performance under each minimization strategy and the full-data strategy, for the Ambar dataset. Smaller values indicate better preservation of overall performance.

Table 2
Mean (μ) and standard deviation (σ) of the distances among the points represented by each model's performance under different strategies and the point corresponding to the model performance under the "full" data strategy, for the Ambar dataset across different n. For each scenario, bold and underline stand for best and second-to-best values, respectively.

Model	$n=1$		$n=3$		$n=7$		$n=15$		$n=100$		Global	
	$\mu\downarrow$	$\sigma\downarrow$	$\mu\downarrow$	$\sigma\downarrow$	$\mu\downarrow$	$\sigma\downarrow$	$\mu\downarrow$	$\sigma\downarrow$	$\mu\downarrow$	$\sigma\downarrow$	$\mu\downarrow$	$\sigma\downarrow$
UserKNN	0.5821	0.2017	0.2212	0.1098	**0.2163**	**0.0956**	**0.1043**	**0.0850**	0.0212	0.0200	0.2290	0.2236
BPRMF	0.3166	0.1467	0.2223	0.0969	0.2245	0.1203	0.1090	0.1031	0.0698	0.0294	0.1884	0.1339
EASER	0.5516	0.1884	**0.2063**	0.0847	0.2328	0.1116	0.1350	0.0926	**0.0133**	**0.0105**	0.2278	0.2100
MultiVAE	**0.2625**	0.1433	0.2196	**0.0809**	0.2372	0.1506	0.1529	0.1145	0.0632	0.0216	**0.1871**	**0.1275**
LightGCN	0.3151	**0.1254**	0.2722	0.1351	0.2546	0.1113	0.1177	0.1000	0.0249	0.0231	0.1969	0.1483

more consistent performance. Our analysis reveals that graph-based and factorization-based models are the most resilient. LightGCN emerges as the most robust model overall, maintaining low μ and σ across all scenarios and proving particularly effective under extreme data sparsity ($n = 1$) due to its information propagation mechanism. BPRMF also demonstrates strong generalization. In contrast, traditional methods show limitations. UserKNN is robust for moderate minimization ($n \geq 3$) but falters with scarcer data. EASER, relying on co-occurrence statistics, is highly sensitive to sparsity. MultiVAE's performance is dataset-dependent, showing weaker robustness in ML1M but achieving the best performance in Ambar at the global level. *We conclude that graph-based modeling, which enhances classic factorization-based representations, is crucial for building recommenders that are robust to extreme data reduction when accuracy and fairness are considered jointly.*

4. Conclusion and Future Work

We audited the impact of data minimization on the accuracy and fairness trade-offs in RSs. Our findings reveal a critical trade-off, where consumer fairness often improves at the cost of accuracy and provider fairness. Strategies that introduce variability into user profiles, alongside robust graph-based models, proved most effective at balancing these objectives. Future work will focus on engineering recommender systems that are inherently fair and robust under severe data constraints.

Declaration on Generative AI

During the preparation of this work, the author did not use any AI tool.

References

[1] California consumer privacy act (ccpa). https://cloud.google.com/security/compliance/ccpa. accessed 2024-02-13.

[2] The california privacy rights act of 2020. https://thecpra.org/. accessed 2024-02-13.

[3] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679. accessed 2024-03-15.

[4] Translation: Personal information protection law of the people's republic of china – effective nov. 1, 2021. https://digichina.stanford.edu/work/translation-personal-information-protection-law-of-the-peoples-republic-of-china-effective-nov-1-2021/. accessed 2024-02-13.

[5] Vito Walter Anelli, Alejandro Bellogín, Tommaso Di Noia, Francesco Maria Donini, Vincenzo Paparella, and Claudio Pomo. Adherence and constancy in LIME-RS explanations for recommendation (long paper). In Vito Walter Anelli, Pierpaolo Basile, Tommaso Di Noia, Francesco Maria Donini, Cataldo Musto, Fedelucio Narducci, Markus Zanker, Himan Abdollahpouri, Toine Bogers, Bamshad Mobasher, Casper Petersen, and Maria Soledad Pera, editors, *Joint Workshop Proceedings of the 3rd Edition of Knowledge-aware and Conversational Recommender Systems (KaRS) and the 5th Edition of Recommendation in Complex Environments (ComplexRec) co-located with 15th ACM Conference on Recommender Systems (RecSys 2021), Virtual Event, Amsterdam, The Netherlands, September 25, 2021*, volume 2960 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2021. URL https://ceur-ws.org/Vol-2960/paper11.pdf.

[6] Vito Walter Anelli, Yashar Deldjoo, Tommaso Di Noia, Daniele Malitesta, Vincenzo Paparella, and Claudio Pomo. Auditing consumer- and producer-fairness in graph collaborative filtering. In Jaap Kamps, Lorraine Goeuriot, Fabio Crestani, Maria Maistro, Hideo Joho, Brian Davis, Cathal Gurrin, Udo Kruschwitz, and Annalina Caputo, editors, *Advances in Information Retrieval - 45th European Conference on Information Retrieval, ECIR 2023, Dublin, Ireland, April 2-6, 2023, Proceedings, Part I*, volume 13980 of *Lecture Notes in Computer Science*, pages 33–48. Springer, 2023. doi: 10.1007/978-3-031-28244-7_3. URL https://doi.org/10.1007/978-3-031-28244-7_3.

[7] Asia J. Biega, Peter Potash, Hal Daumé III, Fernando Diaz, and Michèle Finck. Operationalizing the legal principle of data minimization for personalization. In *SIGIR*, pages 399–408. ACM, 2020.

[8] Salvatore Bufi, Alberto Carlo Maria Mancino, Antonio Ferrara, Daniele Malitesta, Tommaso Di Noia, and Eugenio Di Sciascio. KGUF: simple knowledge-aware graph-based recommender with user-based semantic features filtering. In *IRonGraphs*, volume 2197 of *Communications in Computer and Information Science*, pages 41–59. Springer, 2024.

[9] Salvatore Bufi, Vincenzo Paparella, Vito Walter Anelli, and Tommaso Di Noia. Legal but unfair: Auditing the impact of data minimization on fairness and accuracy trade-off in recommender

systems. In *Proceedings of the 33rd ACM Conference on User Modeling, Adaptation and Personalization, UMAP 2025, New York City, NY, USA, June 16-19, 2025*, pages 114–123. ACM, 2025. doi: 10.1145/3699682.3728356. URL https://doi.org/10.1145/3699682.3728356.

[10] Gemma Galdon Clavell, Mariano Martín Zamorano, Carlos Castillo, Oliver Smith, and Aleksandar Matic. Auditing algorithms: On lessons learned and the risks of data minimization. In *AIES*, pages 265–271. ACM, 2020.

[11] Yashar Deldjoo, Vito Walter Anelli, Hamed Zamani, Alejandro Bellogín, and Tommaso Di Noia. A flexible framework for evaluating user and item fairness in recommender systems. *User Model. User Adapt. Interact.*, 31(3):457–511, 2021. doi: 10.1007/S11257-020-09285-1. URL https://doi.org/10.1007/s11257-020-09285-1.

[12] Yashar Deldjoo, Dietmar Jannach, Alejandro Bellogín, Alessandro Difonzo, and Dario Zanzonelli. Fairness in recommender systems: research landscape and future directions. *User Model. User Adapt. Interact.*, 34(1):59–108, 2024.

[13] Dario Di Palma, Giovanni Maria Biancofiore, Vito Walter Anelli, Fedelucio Narducci, and Tommaso Di Noia. Content-based or collaborative? insights from inter-list similarity analysis of chatgpt recommendations. In *UMAP (Adjunct Publication)*, pages 28–33. ACM, 2025.

[14] Dario Di Palma, Felice Antonio Merra, Maurizio Sfilio, Vito Walter Anelli, Fedelucio Narducci, and Tommaso Di Noia. Do llms memorize recommendation datasets? A preliminary study on movielens-1m. In *SIGIR*, pages 2582–2586. ACM, 2025.

[15] Tobias Eichinger and Axel Küpper. Distributed data minimization for decentralized collaborative filtering systems. In *ICDCN*, pages 140–149. ACM, 2023.

[16] Tobias Eichinger and Axel Küpper. On data minimization and anonymity in pervasive mobile-to-mobile recommender systems. *Pervasive Mob. Comput.*, 103:101951, 2024.

[17] Michèle Finck and Asia Biega. Reviving purpose limitation and data minimisation in personalisation, profiling and decision-making systems. *CoRR*, abs/2101.06203, 2021.

[18] M. Fleischer. The measure of pareto optima. In *EMO*, volume 2632 of *Lecture Notes in Computer Science*, pages 519–533. Springer, 2003.

[19] Elizabeth Gómez, David Contreras, Ludovico Boratto, and Maria Salamó. AMBAR: A dataset for assessing multiple beyond-accuracy recommenders. In *RecSys*, pages 137–147. ACM, 2024.

[20] F. Maxwell Harper and Joseph A. Konstan. The movielens datasets: History and context. *ACM Trans. Interact. Intell. Syst.*, 5(4):19:1–19:19, 2016.

[21] Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, Yong-Dong Zhang, and Meng Wang. Lightgcn: Simplifying and powering graph convolution network for recommendation. In *SIGIR*, pages 639–648. ACM, 2020.

[22] Walid Krichene and Steffen Rendle. On sampled metrics for item recommendation. In Rajesh Gupta, Yan Liu, Jiliang Tang, and B. Aditya Prakash, editors, *KDD '20: The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Virtual Event, CA, USA, August 23-27, 2020*, pages 1748–1757. ACM, 2020. doi: 10.1145/3394486.3403226. URL https://doi.org/10.1145/3394486.3403226.

[23] Nian Li, Chen Gao, Depeng Jin, and Qingmin Liao. Disentangled modeling of social homophily and influence for social recommendation. *IEEE Trans. Knowl. Data Eng.*, 35(6):5738–5751, 2023.

[24] Dawen Liang, Rahul G. Krishnan, Matthew D. Hoffman, and Tony Jebara. Variational autoencoders for collaborative filtering. In *WWW*, pages 689–698. ACM, 2018.

[25] Alberto Carlo Maria Mancino, Antonio Ferrara, Salvatore Bufi, Daniele Malitesta, Tommaso Di Noia, and Eugenio Di Sciascio. Kgtore: Tailored recommendations through knowledge-aware GNN models. In *RecSys*, pages 576–587. ACM, 2023.

[26] Vincenzo Paparella, Vito Walter Anelli, Ludovico Boratto, and Tommaso Di Noia. Reproducibility of multi-objective reinforcement learning recommendation: Interplay between effectiveness and beyond-accuracy perspectives. In Jie Zhang, Li Chen, Shlomo Berkovsky, Min Zhang, Tommaso Di Noia, Justin Basilico, Luiz Pizzato, and Yang Song, editors, *Proceedings of the 17th ACM Conference on Recommender Systems, RecSys 2023, Singapore, Singapore, September 18-22, 2023*, pages 467–478. ACM, 2023. doi: 10.1145/3604915.3609493. URL https://doi.org/10.1145/3604915.3609493.

[27] Vincenzo Paparella, Vito Walter Anelli, Franco Maria Nardini, Raffaele Perego, and Tommaso Di

Noia. Post-hoc selection of pareto-optimal solutions in search and recommendation. In *CIKM*, pages 2013–2023. ACM, 2023.

[28] Vincenzo Paparella, Dario Di Palma, Vito Walter Anelli, and Tommaso Di Noia. Broadening the scope: Evaluating the potential of recommender systems beyond prioritizing accuracy. In Jie Zhang, Li Chen, Shlomo Berkovsky, Min Zhang, Tommaso Di Noia, Justin Basilico, Luiz Pizzato, and Yang Song, editors, *Proceedings of the 17th ACM Conference on Recommender Systems, RecSys 2023, Singapore, Singapore, September 18-22, 2023*, pages 1139–1145. ACM, 2023. doi: 10.1145/3604915.3610649. URL https://doi.org/10.1145/3604915.3610649.

[29] Vincenzo Paparella, Alberto Carlo Maria Mancino, Antonio Ferrara, Claudio Pomo, Vito Walter Anelli, and Tommaso Di Noia. Knowledge graph datasets for recommendation. In Vito Walter Anelli, Pierpaolo Basile, Gerard de Melo, Francesco Maria Donini, Antonio Ferrara, Cataldo Musto, Fedelucio Narducci, Azzurra Ragone, and Markus Zanker, editors, *Proceedings of the Fifth Knowledge-aware and Conversational Recommender Systems Workshop co-located with 17th ACM Conference on Recommender Systems (RecSys 2023), Singapore, September 19th, 2023*, volume 3560 of *CEUR Workshop Proceedings*, pages 109–117. CEUR-WS.org, 2023. URL https://ceur-ws.org/Vol-3560/short2.pdf.

[30] Hossein A. Rahmani, Mohammadmehdi Naghiaei, and Yashar Deldjoo. A personalized framework for consumer and producer group fairness optimization in recommender systems. *Trans. Recomm. Syst.*, 2(3):19:1–19:24, 2024.

[31] Theresia Veronika Rampisela, Maria Maistro, Tuukka Ruotsalo, and Christina Lioma. Evaluation measures of individual item fairness for recommender systems: A critical study. *CoRR*, abs/2311.01013, 2023.

[32] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. BPR: bayesian personalized ranking from implicit feedback. *CoRR*, abs/1205.2618, 2012.

[33] Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl. Grouplens: An open architecture for collaborative filtering of netnews. In *CSCW*, pages 175–186. ACM, 1994.

[34] Abdulaziz Samra, Evgeny Frolov, Alexey Vasilev, Alexander Grigorevskiy, and Anton Vakhrushev. Cross-domain latent factors sharing via implicit matrix factorization. In *RecSys*, pages 309–317. ACM, 2024.

[35] Nasim Sonboli, Sipei Li, Mehdi Elahi, and Asia Biega. The trade-off between data minimization and fairness in collaborative filtering. *CoRR*, abs/2410.07182, 2024.

[36] Harald Steck. Embarrassingly shallow autoencoders for sparse data. In *WWW*, pages 3251–3257. ACM, 2019.

[37] Marco Valentini. Cooperative and competitive llm-based multi-agent systems for recommendation. In Claudia Hauff, Craig Macdonald, Dietmar Jannach, Gabriella Kazai, Franco Maria Nardini, Fabio Pinelli, Fabrizio Silvestri, and Nicola Tonellotto, editors, *Advances in Information Retrieval - 47th European Conference on Information Retrieval, ECIR 2025, Lucca, Italy, April 6-10, 2025, Proceedings, Part V*, volume 15576 of *Lecture Notes in Computer Science*, pages 204–211. Springer, 2025. doi: 10.1007/978-3-031-88720-8_33. URL https://doi.org/10.1007/978-3-031-88720-8_33.

[38] Yifan Wang, Weizhi Ma, Min Zhang, Yiqun Liu, and Shaoping Ma. A survey on the fairness of recommender systems. *ACM Trans. Inf. Syst.*, 41(3):52:1–52:43, 2023.

[39] Ziwei Zhu, Jianling Wang, and James Caverlee. Measuring and mitigating item under-recommendation bias in personalized ranking systems. In Jimmy X. Huang, Yi Chang, Xueqi Cheng, Jaap Kamps, Vanessa Murdock, Ji-Rong Wen, and Yiqun Liu, editors, *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval, SIGIR 2020, Virtual Event, China, July 25-30, 2020*, pages 449–458. ACM, 2020. doi: 10.1145/3397271.3401177. URL https://doi.org/10.1145/3397271.3401177.