The 45th European Conference on Information Retrieval

Dublin, Ireland, 03-04-2023

ECIR 2023 - Main Conference

Auditing Consumer- and Producer-Fairness in Graph Collaborative Filtering

Vito Walter Anelli, Yashar Deldjoo, Tommaso Di Noia, Daniele Malitesta, Vincenzo Paparella, Claudio Pomo

Politecnico di Bari

Bari, Italy

email: firstname.lastname@poliba.it

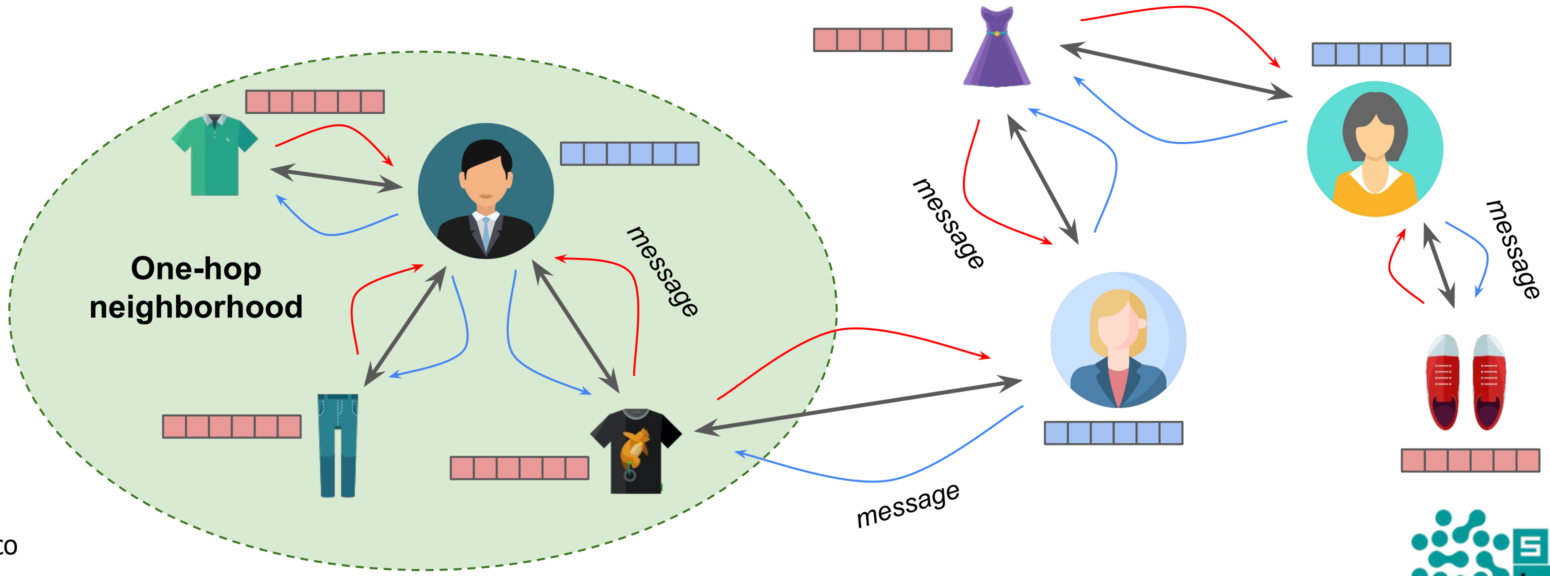
Politecnico di Bari

SISINF LAB

Introduction and Motivation

Graph collaborative filtering

In collaborative filtering (CF), graph convolutional networks (GCNs) have gained momentum thanks to their ability to **aggregate neighbor** nodes **information** into ego nodes at multiple hops (i.e., **message-passing**), thus effectively **distilling** the **collaborative signal**.



Politecnico di Bari

SISINF LAB

Consumer- & Producer- Fairness in RecSys

Recent works have **raised issues** regarding the **fairness** of recommender systems (e.g., **Burke R.**, **Ekstrand et al.**, **Wang et al.**).

Generally, recommendation fairness is categorized through two core aspects:

- **engaged parties** (i.e., consumers vs. producers)
- **type of benefit** (i.e., relevance vs. exposure)

→ **Consumer- and Producer- Fairness**

Their **joint evaluation**, with **accuracy**, is **rarely assessed** in the literature (**Naghiaei et al.**).

Burke, R.. **Multisided fairness for recommendation.** arXiv preprint arXiv:1707.00093 (2017).
Ekstrand, M. D, Das, A., Burke, R., Diaz, F. **Fairness and Discrimination in Information Access Systems.** arXiv preprint arXiv:2105.05779 (2021).
Wang, Y., Ma, W., Zhang, M., Liu, Y., Ma, S.: **A Survey on the Fairness of Recommender Systems.** CoRR abs/2206.03761 (2022).
Naghiaei, M., Rahmani, H.A., Deldjoo, Y.: **Cpfair: Personalized consumer and producer fairness re-ranking for recommender systems.** In: SIGIR, pp. 770–779, ACM (2022).

Consumer- & Producer- Fairness in Graph RecSys

While **state-of-the-art** **graph**-based recommendation models **have centered** on the enhancement of **accuracy** performance, **recent** approaches have designed **ad-hoc** solutions to address:

Consumer-Fairness

- Rahman, T.A., Surma, B., Backes, M., Zhang, Y.: **Fairwalk: Towards fair graph embedding.** IJCAI 2019.
- Fu, Z., et al.: **Fairness-aware explainable recommendation over knowledge graphs.** SIGIR 2020.
- Voit, M.M., Paulheim, H.: **Bias in knowledge graphs - an empirical study with movie recommendation and different language editions of dbpedia.** LDK, OASIcs 2021.
- Wu, L., Chen, L., Shao, P., Hong, R., Wang, X., Wang, M.: **Learning fair representations for recommendation: A graph-based perspective.** WWW 2021.
- Li, C., Hsu, C., Zhang, Y.: **Fairsr: Fairness-aware sequential recommendation through multi-task learning with preference graph embeddings.** TIST 2022.
- Wang, N., Lin, L., Li, J., Wang, H.: **Unbiased graph embedding with biased graph observations.** WWW 2022.

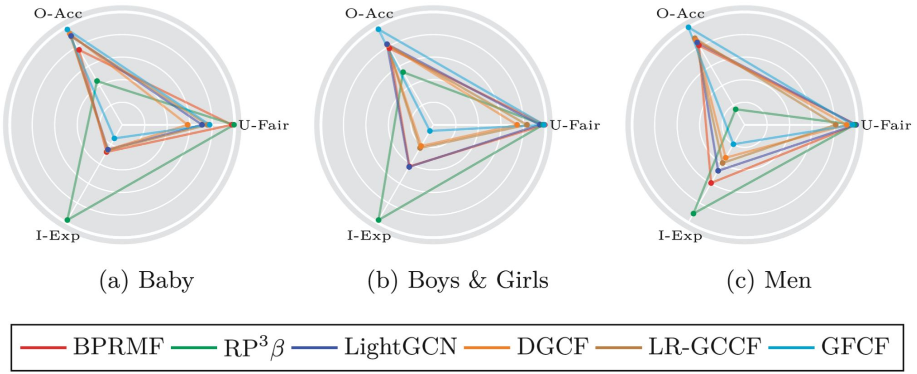
Producer-Fairness

- Mansoury, M., Abdollahpouri, H., Pechenizkiy, M., Mobasher, B., Burke, R.: **Fairmatch: A graph-based approach for improving aggregate diversity in recommender systems.** UMAP 2020.
- Sun, J., et al.:**A framework for recommending accurate and diverse items using bayesian graph convolutional neural networks.** KDD 2020.
- Zheng, Y., Gao, C., Chen, L., Jin, D., Li, Y.: **DGCN: diversified recommendation with graph convolutional networks.** WWW 2021.
- Boltsis, G., Pitoura, E.: **Bias disparity in graph-based collaborative filtering recommenders.** SAC 2022.
- Zhao, M., et al.: **Investigating accuracy-novelty performance for graph-based collaborative filtering.** SIGIR 2022.
- Mansoury, M., Abdollahpouri, H., Pechenizkiy, M., Mobasher, B., Burke, R.: **A graph-based approach for mitigating multi-sided exposure bias in recommender systems.** TIS 2022.

Motivating Example

We seek to investigate how **state-of-the-art** graph-based models for recommendation perform on **three objectives**:

- Accuracy → Overall Accuracy (O-Acc)
- Consumer-Fairness → User Fairness (U-Fair)
- Producer-Fairness → Item Exposure (I-Exp)



(a) Baby (b) Boys & Girls (c) Men

Tested **state-of-the-art** graph-based baselines + classical collaborative filtering methods:

- ➢ LightGCN
- ➢ DGCF
- ➢ LR-GCCF
- ➢ GF-CF
- ➢ BPRMF
- ➢ RP^3Beta

Preliminary results suggest that:

- ➢ **Graph**-based models perform better on **O-Acc**.
- ➢ **Traditional** collaborative filtering models perform better on **I-Exp**.
- ➢ **No clear trend** on **U-Fair**.
- ➢ There exists a **trade-off** among the **three objectives**.

Research Questions and Contributions

RQ1) *“Can we **explain the variations** observed when testing **several graph models** on overall accuracy, item exposure, and user fairness **separately**?”*

- ➢ We propose a **formal taxonomy** of graph collaborative filtering models with an **extended set of models** (i.e., eight) which encompasses two main strategies (i.e., **nodes representation**, **neighborhood exploration**).
- ➢ We extend the previous experimental settings to provide a more nuanced view of the outlined findings.

RQ2) *“How and why **nodes representation** and **neighborhood exploration** algorithms can strike a **trade-off** between **overall accuracy**, **item exposure**, and **user fairness**?”*

- ➢ We use the concept of **Pareto optimality** to assess the **different influence** of each **taxonomy dimension** in a **two-objective** scenario (O-Acc vs. I-Exp, O-Acc vs. U-Fair, I-Exp vs. U-Fair).

A Formal Taxonomy of Graph Collaborative Filtering

Taxonomy Dimensions

Models	Nodes representation				Neighborhood exploration			
	Latent representation		Weighting		Explored nodes		Message passing	
	low	*high*	*weighted*	*unweighted*	*same*	*different*	*implicit*	*explicit*
GCN-CF		✓		✓	✓			✓
GAT-CF		✓	✓		✓			✓
NGCF	✓			✓		✓		✓
LightGCN	✓			✓		✓		✓
DGCF	✓		✓			✓		✓
LR-GCCF	✓			✓	✓	✓		✓
UltraGCN	✓				✓	✓	✓	
GFCF						✓	✓	

Kipf, T.N., Welling, M.: **Semi-supervised classification with graph convolutional networks.** In: ICLR (Poster), OpenReview.net (2017).

Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Li`o, P., Bengio, Y.: **Graph attention networks.** In: ICLR (Poster), OpenReview.net (2018).

Wang, X., He, X., Wang, M., Feng, F., Chua, T.: **Neural graph collaborative filtering.** In: SIGIR, pp. 165–174, ACM (2019).

He, X., Deng, K.,Wang, X., Li, Y., Zhang, Y.,Wang, M.: **Lightgcn: Simplifying and powering graph convolution network for recommendation.** In: SIGIR, pp. 639–648, ACM (2020).

Wang, X., Jin, H., Zhang, A., He, X., Xu, T., Chua, T.: **Disentangled graph collaborative filtering.** In: SIGIR, pp. 1001–1010, ACM (2020).

Chen, L., Wu, L., Hong, R., Zhang, K., Wang, M.: **Revisiting graph based collaborative filtering: A linear residual graph convolutional network approach.** In: AAAI, pp. 27–34, AAAI Press (2020).

Mao, K., Zhu, J., Xiao, X., Lu, B., Wang, Z., He, X.: **Ultragcn: Ultra simplification of graph convolutional networks for recommendation.** In: CIKM, pp. 1253–1262, ACM (2021).

Shen, Y., et al.: **How powerful is graph convolution for recommendation?** In: CIKM, pp. 1619–1629, ACM (2021).

(Extended) Experimental Settings

Graph-based Baselines & Datasets

We **extend** the set of graph-based models to the **eight** approaches used in our **taxonomy**:

- GCN-CF*
- GAT-CF*
- NGCF
- LightGCN
- DGCF
- LR-GCCF
- UltraGCN
- GFCF

Datasets**	Users	Items	Interactions
*Baby****	5,842	7,925	35,475
*Boys & Girls****	3,042	12,912	35,762
*Men****	3,909	27,656	51,519

Trained through **grid-search** on **48** unique **hyper-parameter settings**.

Splitting: **70/10/20** for train/validation/test.

Batch size: 256, number of **epochs**: 400 (these **parameters** are **shared** among all models).

* The postfix -CF indicates we re-adopt the original implementation to the task of personalized recommendation.
** Taken from the Amazon dataset (available at this link: https://cseweb.ucsd.edu/~jmcauley/datasets.html#amazon_reviews).
*** We randomly sample 60k interactions and drop users/items with less than 5 interactions.

Evaluation Metrics

	Metrics	Formulation	Reference
Overall Accuracy	*Recall* ↑	$\frac{\lvert Rel_u \cap Rec_u \rvert}{Rel_u}$	—
	nDCG ↑	$\frac{\sum_{u \in u^{te}} NDCG_u@K}{\lvert u^{te} \rvert}$	—
Item Exposure	*EFD* ↑	$C \sum_{i_k \in R} disc(k) p(rel \mid i_k, u)(-\log_2 p(i \mid seen, \theta))$	**Vargas and Castells**
	Gini Index ↑	$\frac{1}{n-1} \sum_{j=1}^{n} (2j - n - 1) p\left(i_j\right)$	**Shani et al.**
	APLT ↑	$\frac{1}{\lvert U_t \rvert} \sum_{u \in U_t} \frac{\lvert \{i, i \in (L(u) \cap \sim \Phi)\} \rvert}{\lvert L(u) \rvert}$	**Abdollahpouri et al.**
User Fairness	*UMADrank* ↓	$avg_{i,j}(MAD(R^{(i)}, R^{(j)}))$	**Deldjoo et al.**
	UMADrat ↓	$avg_{i,j}(MAD(R^{(i)}, R^{(j)}))$	**Deldjoo et al.**

Vargas, S., Castells, P.: **Rank and relevance in novelty and diversity metrics for recommender systems.** In: RecSys, pp. 109–116, ACM (2011).

Shani, G., Gunawardana, A.: **Evaluating Recommendation Systems.** In: Ricci, F., Rokach, L., Shapira, B., Kantor, P.B. (eds.) Recommender Systems Handbook, pp. 257–297. Springer, Boston, MA (2011).

Abdollahpouri, H., Burke, R., Mobasher, B.: **Controlling popularity bias in learning-to-rank recommendation.** In: RecSys, pp. 42–46, ACM (2017).

Deldjoo, Y., Anelli, V.W., Zamani, H., Bellogin, A., Di Noia, T.: **A flexible framework for evaluating user and item fairness in recommender systems.** User Model. User Adapt. Interact. 31(3), 457–511 (2021).

Results and Discussion

RQ1. Taxonomy-Aware Evaluation

- Of all taxonomy **sub-dimensions** (i.e., message-passing, explored nodes, weighting, and latent representations), we specifically **focus** on **message-passing** and **weighting** as representative of **neighborhood exploration** and **node representation**.
- The results refer to the **Amazon Men** dataset, for **top-20** recommendation **lists**.
- Best values for each metric is in **bold**.
- Message-Passing:
  - **implicit** (i.e., GCN-CF, GAT-CF, NGCF, LightGCN, DGCF, LR-GCCF).
  - **explicit** (i.e., UltraGCN, GFCF).
- Weighting:
  - **weighted** (i.e., GAT-CF, DGCF).
  - **unweighted** (i.e., GCN-CF, NGCF, LightGCN, LR-GCCF).

ECIR 2023

RQ1.1 Message-Passing

Dimensions	Values	Overall accuracy		Item exposure			User fairness	
		Recall↑	*nDCG*↑	*EFD*↑	*Gini*↑	*APLT*↑	*rank*↓	*rat*↓
Message passing	*implicit*	0.1222 (GFCF)	**0.0911 (GFCF)**	**0.2615 (GFCF)**	0.2871 (UltraGCN)	0.1808 (UltraGCN)	0.0123 (UltraGCN)	**0.0022 (UltraGCN)**
	explicit	**0.1223 (LR-GCCF)**	0.0884 (LR-GCCF)	0.2536 (LR-GCCF)	**0.5090 (LR-GCCF)**	**0.3823 (GAT-CF)**	**0.0002 (DGCF)**	0.0169 (LightGCN)

- **Both implicit** and **explicit** have the **same** number of **top-performing** models.
- **Explicit** methods **outperform explicit** ones on **item exposure** (e.g., the **absence** of **message-passing** prevents users from **exploring** vast **item** segments).
- **No** obvious **reason** to favour **implicit** or **explicit** models on **accuracy** and user **fairness**.

Politecnico di Bari

SISINF LAB

RQ1.2 Weighting

Dimensions	Values	Overall accuracy		Item exposure			User fairness	
		Recall↑	*nDCG*↑	*EFD*↑	*Gini*↑	*APLT*↑	*rank*↓	*rat*↓
Weighting	*weighted*	0.1210 (DGCF)	0.0857 (DGCF)	0.2428 (DGCF)	0.3240 (DGCF)	**0.3823 (GAT-CF)**	**0.0002 (DGCF)**	0.0301 (DGCF)
	unweighted	**0.1223 (LR-GCCF)**	**0.0884 (LR-GCCF)**	**0.2536 (LR-GCCF)**	**0.5090 (LR-GCCF)**	0.3438 (LR-GCCF)	0.0101 (GCN-CF)	**0.0169 (LightGCN)**

- Almost all **unweighted** provide better **performance**.
- Only deviation is **GAT-CF** on **APLT** (i.e., recommends more items from the long-tail).
- This happens since **weighted** techniques explore **further neighborhoods** of the ego node (see again GAT-CF on same-2), thus matching niche products.
- **EFD** and **Gini** do **not** follow the **same behaviour**, but they represent **other viewpoints** of item exposure.

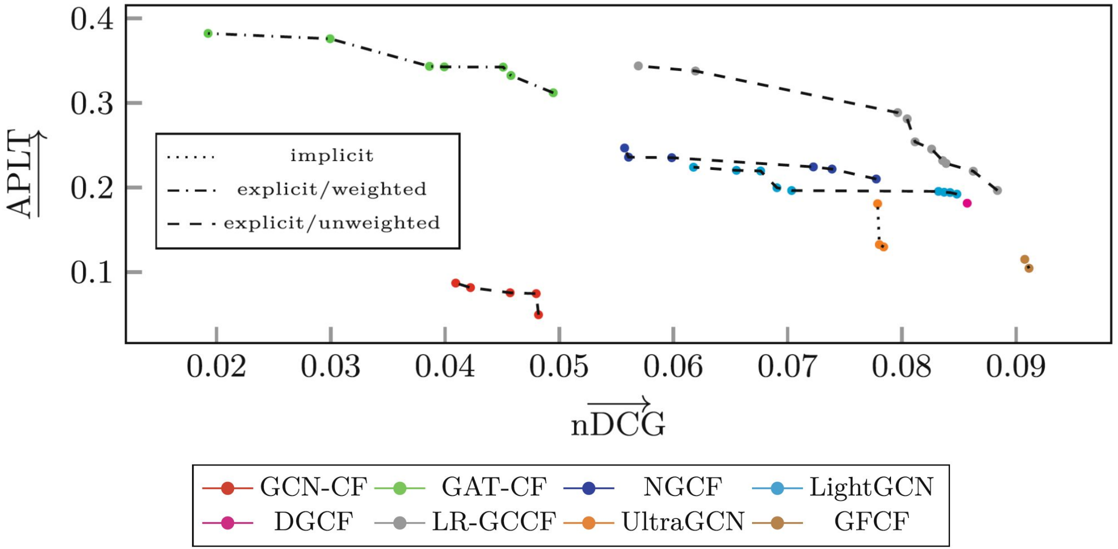
RQ2. Trade-Off Analysis

- Through **Pearson** correlation, we observe that **nDCG/APLT**, and **nDCG/UMADrank** are **negatively correlated**.
- A **trade-off** for these metrics pairs might be **necessary** and **desirable**.
- We also seek to investigate the **APLT/UMADrank** trade-off.
- To conduct the study, we select the **message-passing** and **weighting** architectural choices, having the following categories:
  - implicit message-passing,
  - explicit + weighted,
  - explicit + unweighted.
- We select the **Pareto** optimal solutions (among the 48 explored configurations) laying on the Pareto **frontier** (**Paparella V.**).

Paparella, V.: **Pursuing optimal trade-off solutions in multi-objective recommender systems.** In: RecSys, pp. 727–729, ACM (2022)

ECIR 2023

RQ2.1 Accuracy/Item Exposure

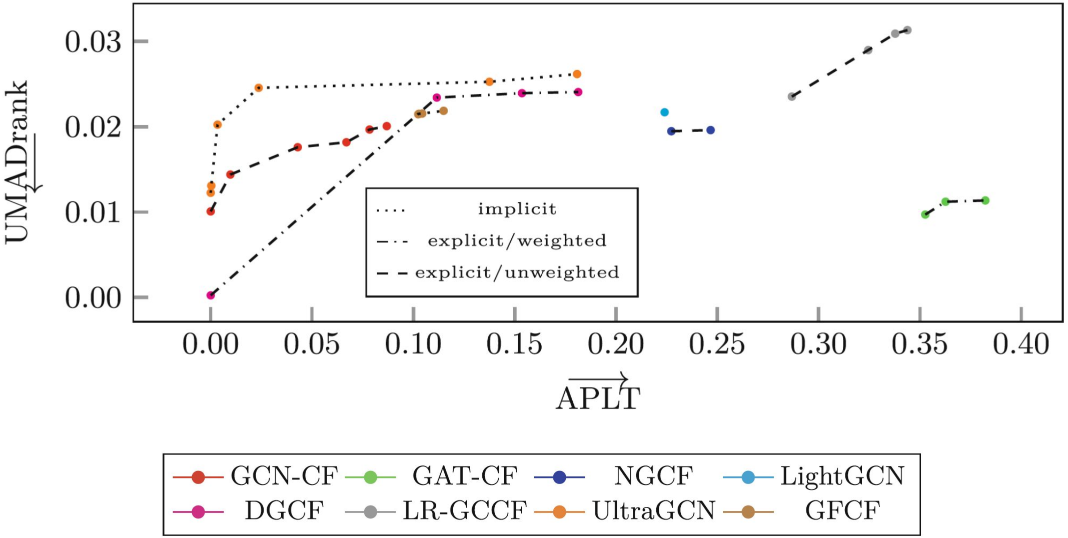


- **Explicit/weighted** maximize **either nDCG** or **APLT**.
- **Explicit/unweighted** show a **balanced** trade-off.
- **LR-GCCF**'s frontier **dominates** over the other **explicit/unweighted** ones.
- **Implicit** models increase **accuracy** at the **expense** of item **exposure**.

Politecnico di Bari

SISINF LAB

RQ2.2 Item Exposure/User Fairness



- **Two** groups: **poor** item exposure vs. **acceptable** item **exposure**.
- A subset of models belonging to **explicit/unweighted** recommend **niche** products and provide **acceptable accuracy**.
- **GAT-CF** is near the **utopia** point, but it **varies** quite a lot on the **accuracy**.

Conclusion and Future Work

Conclusion

- We **assess** the performance of **graph-based** recommendation models on **Consumer-** and **Producer-**Fairness measures.
- We recognize a **taxonomy** of graph-based approaches based upon **node representation** and **neighborhood exploration** strategies.
- We study **accuracy** vs. **CP-Fairness separately** and **simultaneously**.
- **Concerns** about the adoption of **recent** (i.e., **implicit**) approaches.

Future Work

- Analyze other **datasets** and **algorithms**.
- Propose **novel** graph-based **approaches** balancing **accuracy** and **CP-fairness**.

Thank you! Any questions?

Our official GitHub repository:

GitHub

https://github.com/sisinflab/ECIR2023-Graph-CF

Contacts:

vitowalter.anelli@poliba.it

yashar.deldjoo@poliba.it

tommaso.dinoia@poliba.it

daniele.malitesta@poliba.it (PRESENTER)

vincenzo.paparella@poliba.it

claudio.pomo@poliba.it