Toward Recognizing Social Media Recommenders Under Absent Recommendations: A Graph Neural Network-Based Approach

Extended Abstract

Sabrina Guidotti
University of Milan-Bicocca
Milan, Italy
s.guidotti2@campus.unimib.it

Gregor Donabauer
University of Regensburg
Regensburg, Germany
Gregor.Donabauer@ur.de

Davide Taibi
CNR
Palermo, Italy
davide.taibi@itd.cnr.it

Giuseppe Vizzari
University of Milan-Bicocca
Milan, Italy
giuseppe.vizzari@unimib.it

Udo Kruschwitz
University of Regensburg
Regensburg, Germany
Udo.Kruschwitz@ur.de

Dimitri Ognibene
University of Milan-Bicocca
Milan, Italy
dimitri.ognibene@unimib.it

ABSTRACT

Assessing social media algorithms is hindered by platform opacity and data unavailability. To address this, we introduce *Social Media Recommenders Recognition under Absent Recommendations* (SM-ARR) and present SM-ARR-G, a Graph Neural Network framework designed to identify active algorithms without internal access. SM-ARR-G forecasts user actions by comparing past behavior against candidate "infospheres" (simulated exposure patterns), selecting the best predictor as the explanation for observed dynamics. Initial experiments using the DBLP dataset as a proxy indicate that our approach can detect hidden recommenders. If further developed, this framework could potentially offer a valuable tool for external auditing and enhancing algorithmic explainability.

KEYWORDS

Social Media; Recommender Systems; Graph Neural Networks; Algorithm Auditing; Transparency; User Behavior Modeling; SIM

ACM Reference Format:
Sabrina Guidotti, Gregor Donabauer, Davide Taibi, Giuseppe Vizzari, Udo Kruschwitz, and Dimitri Ognibene. 2026. Toward Recognizing Social Media Recommenders Under Absent Recommendations: A Graph Neural Network-Based Approach: Extended Abstract. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), Paphos, Cyprus, May 25 – 29, 2026*, IFAAMAS, 3 pages. https://doi.org/10.65109/YMXQ7472

1 INTRODUCTION

Engagement-focused algorithms raise significant ethical concerns regarding misinformation and polarization [1–3, 8, 15, 25]. While the personalization-polarization link is debated [9, 16], evidence that algorithms intensify societal tensions [7, 19, 23] highlights the urgent need for transparency [17, 18]. Existing auditing techniques reveal biases [4, 20, 22, 24, 26], while simulation studies illuminate long-term impacts [6, 13, 24]. Unlike research predicting impact through simulation, we propose a likelihood-driven framework to reverse-engineer the recommender from observed behaviors. Since many datasets lack the necessary structural richness [5, 14, 21], we use large-scale academic networks as a proxy. By analyzing the Graph Neural Network model loss under hypothetical recommenders, our approach infers the most likely recommender, enabling accountability even where direct auditing is infeasible.

2 METHODOLOGY

SM-ARR: Problem Statement. We formalize *Social Media Recommenders Recognition under Absent Recommendations (SM-ARR)* as the task of inferring, from observed user interactions alone and *without* access to platform recommendation logs, which candidate recommender R most plausibly generated the observed dynamics. Concretely, for each hypothesized R we condition a user–behavior predictor on the corresponding simulated *infosphere* and evaluate out-of-sample negative log-likelihood; the R yielding the lowest test loss is selected as the recognized recommender. We instantiate this with **SM-ARR-G**, a GNN-based framework operationalizing this criterion.

SM-ARR-G Implementation. Guidotti *et al.* [10] showed that injecting a simulated recommender into a GNN-based predictor changes the model's test loss, providing a quantitative measure of how well that recommender accounts for the observed social-network behavior. Building on this idea, we formally argue and empirically demonstrate that the candidate producing the lowest test loss constitutes the most plausible explanation.

Likelihood formulation. Let $D = \{(x_i, y_i)\}_{i=1}^{N}$ be the sequences of past and next actions for N users, and let R denote a hypothesized recommender. Training a GNN with parameters θ under hypothesis R yields the test loss $\mathcal{L}(\theta; D, R) = -\sum_{i=1}^{N} \log P_\theta(y_i|x_i, R)$. We identify the true generator $R^\star$ by minimizing $\mathcal{L}$ over candidate hypotheses: $R^\star = \arg\min_R \mathcal{L}(\hat{\theta}_R; D_{\text{test}}, R)$.

Data Generation and Selection Workflow. Testing requires datasets containing both real interactions and recommender logic, which are rare. We therefore generate datasets by learning a *recommender-neutral user model* (RNU) and using it simulate interactions under known recommenders.

Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), C. Amato, L. Dennis, V. Mascardi, J. Thangarajah (eds.), May 25 – 29, 2026, Paphos, Cyprus. © 2026 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). https://doi.org/10.65109/YMXQ7472

The selection process follows a three-step pipeline: (i) selecting a social media dataset, (ii) learning a new GNN user model for each available social media recommender, and (iii) selecting the recommender that results in the lowest learning loss.

Recommender-Neutral User Model (RNU). The RNU aims to extract the probability of interaction y between agents u, v given their states s_u, s_v and simulated infospheres r_u, r_v, minimizing dependence on the hidden platform recommender.

Marginalizing Over the Hidden Recommender. We employ a latent-variable marginal likelihood, maximizing the log-likelihood of observed actions: $\sum_i \log P_\theta(y_i \mid s_{u,i}, s_{v,i})$. We define $P_\theta(y \mid s_u, s_v)$ by marginalizing over latent recommendations (r_u, r_v) and hypothesized recommenders R:

$$P_\theta(y \mid s_u, s_v) = \int P(R) \left[\iint P_\theta(y \mid s_u, s_v, r_u, r_v) P(r_u, r_v \mid R, s_u, s_v)\, dr_u, dr_v \right] dR \quad (1)$$

Here, $P_\theta(y \mid \cdot)$ is the user model, $P(r_u, r_v \mid R, \cdot)$ is the recommendation probability under R, and $P(R)$ is a uniform prior over plausible recommenders. As full marginalization is computationally infeasible, we approximate this using a hindsight predictive model [10].

Hindsight Model of the Recommender. Following [10], the hindsight approach unfolds in three sequential steps: (i) creating a Hindsight Social Media Recommender; (ii) learning a Recommender Neutral User (RNU) model using GNN; and (iii) creating synthetic datasets by combining the RNU with known recommenders. This methodology uses actual future behavior to generate recommendations during training, which reduces dependency on unknown recommenders and lowers computational complexity while maintaining control over noisy recommendation paths.

3 EXPERIMENTS

Metrics and Datasets. We evaluate recognition using Negative Log-Likelihood (NLL) and edge classification accuracy on a held-out test split. We use the DBLP-Citation-Network V14 dataset as a proxy for social networks due to its relational richness, featuring over 5M paper nodes and 36M citation edges.

Synthetic Data Generation. We evaluate six infosphere types to assess recommendation predictability: **(1) No Infosphere**, a baseline using only real network data; **(2) Hindsight**, adopted from [10], is based on a seedgraph where paths trace the shortest connection from an author's history in year $y+1$ back to the graph in year y. This structure is enhanced with alternative branches as noise to improve realism; **(3) Top Paper**, featuring the n most popular papers; **(4) Top Paper × Topic**, combining global popularity with m topic-specific preferences; **(5) LightGCN**, a GNN-based collaborative filtering model [11]; and **(6) NAIS**, a neural attentive item similarity approach [12].

Generating interactions involves significant computational complexity due to the network's scale. To maintain realism and control complexity, we anchored the process to real-world values by using actual co-author counts as input for year $y+1$. For each infosphere, we simulated a sparse, symmetric connectivity matrix based on historical data, enforcing personalized node degree bounds and utilizing learned pairwise connection probabilities from author embeddings . For computational efficiency, ground-truth simulation was restricted to authors active in year $y+1$ (in this case 2020).

Applying the Recognition Method. For each candidate recommender R', we train a new Heterogeneous Graph Transformer (HGT) model from scratch. The architecture utilizes two HGT layers followed by a fully connected layer for link prediction. Training involves 5-fold cross-validation with binary cross-entropy loss, using the Adam optimizer and a OneCycleLR scheduler. Complete code is available at https://github.com/DimNeuroLab/academic_network_project.

4 RESULTS AND DISCUSSION

We evaluated three ground-truth generation methodologies: random sampling with hindsight embeddings, probability-based sorting with hindsight embeddings, and probability-based sorting with recommender-specific embeddings, but only the latter is reported here due to space constraints.

Consistent with the results obtained across all tested ground-truth recommenders (not reported), the 5-fold cross-validation results using LightGCN as the ground-truth recommender (Table 1) exemplify that the model matching the ground-truth generator achieves the highest likelihood with minimal sensitivity to fold variation. These findings suggest the framework successfully captures key characteristics of the underlying recommender system from user interactions alone.

5 CONCLUSION

By testing the method in [10] on synthetic datasets, we show that hidden recommender systems can be partially inferred from user interaction data alone. This validates the potential for generating recommender-neutral user models suitable for simulating algorithmic impacts in realistic settings. Our approach extends existing audits by detecting recommenders during actual operation, providing an initial step toward improving the societal impacts of social media dynamics.

Future efforts will target scalability and peculiarities of advance real world recommenders as specialization and dynamic campaigns.

ACKNOWLEDGMENTS

This research was supported by the Italian Ministry of University and Research under Grant No. 2023-NAZ-0206, PsyFuture – Dipartimento di Eccellenza 2023-2027.

Infosphere	Params	Accuracy ($\mu \pm \sigma$)	Loss ($\mu \pm \sigma$)	F1 ($\mu \pm \sigma$)
HINDSIGHT	5	0.9536 ± 0.0310	0.1553 ± 0.0947	0.9560 ± 0.0275
NO INFOSPHERE	N/A	0.8453 ± 0.0305	0.4282 ± 0.0534	0.8608 ± 0.0222
TOP PAPER	10	0.7558 ± 0.1778	0.4769 ± 0.2042	0.7017 ± 0.2403
TOP PAPER * TOPIC	[5,2]	0.8508 ± 0.0604	0.4123 ± 0.1102	0.8569 ± 0.0506
LightGCN	**N/A**	**0.9836 ± 0.0113**	**0.0900 ± 0.0579**	**0.9839 ± 0.0108**
NAIS	N/A	0.9638 ± 0.0226	0.1452 ± 0.0897	0.9643 ± 0.0228

Table 1: 5-fold cross validation results for *LightGCN* ground truth.

REFERENCES

[1] Hunt Allcott, Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow. 2020. The welfare effects of social media. *American economic review* 110, 3 (2020), 629–676.

[2] Mohamed Basel Almourad, John McAlaney, Tiffany Skinner, Megan Pleya, and Raian Ali. 2020. Defining digital addiction: Key features from the literature. *Psihologija* 53, 3 (2020), 237–253.

[3] Christopher A Bail, Lisa P Argyle, Taylor W Brown, John P Bumpus, Haohan Chen, MB Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky. 2018. Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences* 115, 37 (2018), 9216–9221.

[4] Reuben Binns, Max Van Kleek, Michael Veale, Ulrik Lyngs, Jun Zhao, and Nigel Shadbolt. 2018. 'It's Reducing a Human Being to a Percentage': Perceptions of Justice in Algorithmic Decisions. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3173574.3173951

[5] Christian G. Fink, Nathan Omodt, Sydney Zinnecker, and Gina Sprint. 2023. A Congressional Twitter network dataset quantifying pairwise probability of influence. *Data in Brief* 50 (2023), 109521. https://doi.org/10.1016/j.dib.2023.109521

[6] Daniel Fleder, Kartik Hosanagar, and Andreas Buja. 2010. Recommender systems and their effects on consumers: the fragmentation debate. In *Proceedings of the 11th ACM Conference on Electronic Commerce* (Cambridge, Massachusetts, USA) *(EC '10)*. Association for Computing Machinery, New York, NY, USA, 229–230. https://doi.org/10.1145/1807342.1807378

[7] Germain Gauthier, Roland Hodler, Philine Widmer, and Ekaterina Zhuravskaya. 2026. The political effects of X's feed algorithm. *Nature* (18 feb 2026). https://doi.org/10.1038/s41586-026-10098-2

[8] Nabeel Gillani, Ann Yuan, Martin Saveski, Soroush Vosoughi, and Deb Roy. 2018. Me, My Echo Chamber, and I: Introspection on Social Media Polarization. In *Proceedings of the 2018 World Wide Web Conference* (Lyon, France) *(WWW '18)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 823–831. https://doi.org/10.1145/3178876.3186130

[9] Andrew M. Guess, Neil Malhotra, Jennifer Pan, Pablo Barberá, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Drew Dimmery, Deen Freelon, Matthew Gentzkow, Sandra González-Bailón, Edward Kennedy, Young Mie Kim, David Lazer, Devra Moehler, Brendan Nyhan, Carlos Velasco Rivera, Jaime Settle, Daniel Robert Thomas, Emily Thorson, Rebekah Tromble, Arjun Wilkins, Magdalena Wojcieszak, Beixian Xiong, Chad Kiewiet de Jonge, Annie Franco, Winter Mason, Natalie Jomini Stroud, and Joshua A. Tucker. 2023. How do social media feed algorithms affect attitudes and behavior in an election campaign? *Science* 381, 6656 (2023), 398–404. https://doi.org/10.1126/science.abp9364 arXiv:https://www.science.org/doi/pdf/10.1126/science.abp9364

[10] Sabrina Guidotti, Gregor Donabauer, Simone Somazzi, Udo Kruschwitz, Davide Taibi, and Dimitri Ognibene. 2025. Modeling Social Media Recommendation Impacts Using Academic Networks: A Graph Neural Network Approach. In *Recommender Systems for Sustainability and Social Good*, Ludovico Boratto, Allegra De Filippo, Elisabeth Lex, and Francesco Ricci (Eds.). Springer Nature Switzerland, Cham, 63–72.

[11] Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, YongDong Zhang, and Meng Wang. 2020. LightGCN: Simplifying and Powering Graph Convolution Network for Recommendation. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval* (Virtual Event, China) *(SIGIR '20)*. Association for Computing Machinery, New York, NY, USA, 639–648. https://doi.org/10.1145/3397271.3401063

[12] Xiangnan He, Zhankui He, Jingkuan Song, Zhenguang Liu, Yu-Gang Jiang, and Tat-Seng Chua. 2018. NAIS: Neural Attentive Item Similarity Model for Recommendation. *IEEE Transactions on Knowledge and Data Engineering* 30, 12 (Dec. 2018), 2354–2366. https://doi.org/10.1109/tkde.2018.2831682

[13] Kartik Hosanagar, Daniel Fleder, Dokyun Lee, and Andreas Buja. 2014. Will the global village fracture into tribes? Recommender systems and their effects on consumer fragmentation. *Management Science* 60, 4 (2014), 805–823.

[14] Julian McAuley and Jure Leskovec. 2012. Learning to discover social circles in ego networks. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 1* (Lake Tahoe, Nevada) *(NIPS'12)*. Curran Associates Inc., Red Hook, NY, USA, 539–547.

[15] Dimitar Nikolov, Diego FM Oliveira, Alessandro Flammini, and Filippo Menczer. 2015. Measuring online social bubbles. *PeerJ computer science* 1 (2015), e38.

[16] Brendan Nyhan, Jaime Settle, Emily Thorson, Magdalena Wojcieszak, Pablo Barberá, Annie Chen, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Drew Dimmery, Deen Freelon, Matthew Gentzkow, Sandra González-Bailón, Andrew Guess, Edward Kennedy, Young Kim, David Lazer, Neil Malhotra, Devra Moehler, and Joshua Tucker. 2023. Like-minded sources on Facebook are prevalent but not polarizing. *Nature* 620 (07 2023), 1–8. https://doi.org/10.1038/s41586-023-06297-w

[17] Dimitri Ognibene, Gregor Donabauer, Emily Theophilou, Sathya Buršić, Francesco Lomonaco, Rodrigo Wilkens, Davinia Hernández-Leo, and Udo Kruschwitz. 2023. Moving beyond benchmarks and competitions: towards addressing social media challenges in an educational context. *Datenbank-Spektrum* 23, 1 (2023), 27–39.

[18] Dimitri Ognibene, Rodrigo Wilkens, Davide Taibi, Davinia Hernández-Leo, Udo Kruschwitz, Gregor Donabauer, Emily Theophilou, Francesco Lomonaco, Sathya Bursic, Rene Alejandro Lobo, et al. 2023. Challenging social media threats using collective well-being-aware recommendation algorithms and an educational virtual companion. *Frontiers in Artificial Intelligence* 5 (2023), 654930.

[19] Tiziano Piccardi, Martin Saveski, Chenyan Jia, Jeffrey Hancock, Jeanne L. Tsai, and Michael S. Bernstein. 2025. Reranking partisan animosity in algorithmic social media feeds alters affective polarization. *Science* 390, 6776 (2025), eadu5584. https://doi.org/10.1126/science.adu5584 arXiv:https://www.science.org/doi/pdf/10.1126/science.adu5584

[20] Manoel Horta Ribeiro, Raphael Ottoni, Robert West, Virgílio A. F. Almeida, and Wagner Meira. 2020. Auditing radicalization pathways on YouTube. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (Barcelona, Spain) *(FAT* '20)*. Association for Computing Machinery, New York, NY, USA, 131–141. https://doi.org/10.1145/3351095.3372879

[21] Benedek Rozemberczki, Carl Allen, and Rik Sarkar. 2019. Multi-scale Attributed Node Embedding. arXiv:1909.13021 [cs.LG]

[22] Piotr Sapiezynski, Avijit Ghosh, Levi Kaplan, Aaron Rieke, and Alan Mislove. 2022. Algorithms that "Don't See Color": Measuring Biases in Lookalike and Special Ad Audiences. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society* (Oxford, United Kingdom) *(AIES '22)*. Association for Computing Machinery, New York, NY, USA, 609–616. https://doi.org/10.1145/3514094.3534135

[23] H. Holden Thorp and Valda Vinson. 2024. Context matters in social media. *Science* 385, 6716 (2024), 1393–1393. https://doi.org/10.1126/science.adt2983 arXiv:https://www.science.org/doi/pdf/10.1126/science.adt2983

[24] Matus Tomlein, Branislav Pecher, Jakub Simko, Ivan Srba, Robert Moro, Elena Stefancova, Michal Kompan, Andrea Hrckova, Juraj Podrouzek, and Maria Bielikova. 2021. An Audit of Misinformation Filter Bubbles on YouTube: Bubble Bursting and Recent Behavior Changes. In *Proceedings of the 15th ACM Conference on Recommender Systems* (Amsterdam, Netherlands) *(RecSys '21)*. Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3460231.3474241

[25] Antonela Tommasel and Filippo Menczer. 2022. Do Recommender Systems Make Social Media More Susceptible to Misinformation Spreaders?. In *Proceedings of the 16th ACM Conference on Recommender Systems* (Seattle, WA, USA) *(RecSys '22)*. Association for Computing Machinery, New York, NY, USA, 550–555. https://doi.org/10.1145/3523227.3551473

[26] Zeynep Tufekci. 2018. YouTube, the great radicalizer. *The New York Times* 10, 3 (2018), 2018.