Pursuing Privacy in Recommender Systems: the View of Users and Researchers from Regulations to Applications

Vito Walter Anelli
vitowalter.anelli@poliba.it
Polytechnic University of Bari
Bari, Italy

Luca Belli
lbelli@twitter.com
Twitter
Bari, USA

Yashar Deldjoo
yashar.dedljoo@poliba.it
Polytechnic University of Bari
Bari, Italy

Tommaso Di Noia
tommaso.dinoia@poliba.it
Polytechnic University of Bari
Bari, Italy

Antonio Ferrara*
antonio.ferrara@poliba.it
Polytechnic University of Bari
Bari, Italy

Fedelucio Narducci
fedelucio.narducci@poliba.it
Polytechnic University of Bari
Bari, Italy

Claudio Pomo
claudio.pomo@poliba.it
Polytechnic University of Bari
Bari, Italy

ABSTRACT

Recommender systems (RSs) have widely grown thanks to the outstanding capability of providing users with accurate and tailored recommendations. Recently, public awareness and new regulations forced RS researchers and practitioners to study solutions to user privacy endangerment. This tutorial will guide the attendees through the possible threats and the solutions towards private RSs.

CCS CONCEPTS

• **Information systems** → **Recommender systems**; • **Security and privacy**;

KEYWORDS

recommender systems, privacy

ACM Reference Format:
Vito Walter Anelli, Luca Belli, Yashar Deldjoo, Tommaso Di Noia, Antonio Ferrara, Fedelucio Narducci, and Claudio Pomo. 2021. Pursuing Privacy in Recommender Systems: the View of Users and Researchers from Regulations to Applications. In *Fifthteenth ACM Conference on Recommender Systems (RecSys '21), September 27-October 1, 2021, Amsterdam, Netherlands.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3460231.3473326

1 MOTIVATION AND TOPIC IMPORTANCE

In the last years, recommender systems (RSs) have become a dominant tool in online services, and their use has grown widely, thanks to the outstanding capability of mitigating the information overload problem in the user decision-making process. To provide users with accurate and tailored recommendations, RSs rely on the availability of personal user information, like past preferences about items, behavioral and demographic data, context, and social connections. However, data availability is usually the result of data collection practices, which may represent a threat to user privacy if we think about abuse of data and the extraction of information beyond what is intentionally revealed [28]. In the meanwhile, public awareness related to the privacy of data has been spurred and increased over the last few years [11, 18]. Furthermore, in recent years, an increasing number of countries have introduced regulations to protect user privacy and data security by prohibiting free data circulation and forcing personal data to remain isolated and fragmented (e.g., the GDPR in the European Union [17], the CCPA in California [12], and the Cybersecurity Law in China [30]). Despite the countless calls for better measures to protect sensitive and personally identifiable information, the topic of privacy has been quite under-researched in the community of RSs. After the last tutorial on privacy at RecSys in 2017 [19], only five research papers (we used the keywords "priva", "secur", "distributed", "decentralized", and "federated") were published on this topic [11, 15, 27, 32, 33]. In the meanwhile, new privacy-oriented learning paradigms (such as federated learning [23]) emerged, promoting new perspectives for studying privacy-preserving RSs.

To bridge the gap in perspectives and advances between the RecSys and other AI communities, we propose a **180-minutes tutorial** on privacy-awareness in RSs. We will provide the attendees with an overview of the privacy problems and the most adopted techniques for addressing such issues. The majority of the tutorial will focus on novel learning paradigms such as federated learning, providing a novel organization of the literature of privacy-oriented RSs. We will analyze theoretically and practically the possible threats and solutions to user privacy in recommendation models with dedicated hands-on sessions. We will cover both academic and industrial points of view about privacy, hopefully inspiring a discussion on a general rethinking of RSs for meeting the urge of privacy.

*Corresponding author

RecSys '21, September 27-October 1, 2021, Amsterdam, Netherlands

ACM ISBN 978-1-4503-8458-2/21/09…$15.00
https://doi.org/10.1145/3460231.3473326

2 OUTLINE OF THE TUTORIAL

The tutorial is split into eight parts presented by academic and industrial speakers.

2.1 The Data Paradox: Privacy and Utility in the Era of Regulations

Nowadays, one of the main challenges in RSs is protecting user privacy while still leveraging their data for the utility of business entities and the users themselves. Recent regulations have forced data to remain fragmented and isolated. We start our tutorial with an overview of such regulations in the context of the privacy-utility trade-off.

2.2 Privacy-Oriented Recommender Systems

In recent years, some paradigms and strategies have been proposed for privacy-oriented machine learning and RS privacy.

Learning Paradigms for Privacy-Oriented Recommender Systems. We provide the reader with a taxonomy about learning paradigms, which we deem fundamental to put in order the recent literature about privacy-oriented RecSys [18]. Indeed, centralized, distributed, and decentralized learning [13] are now flanked by federated learning [23, 35], which was explicitly conceived as a privacy-preserving learning paradigm. Each of these paradigms has to be thought and implemented according to the actors involved in training, the computational choices, and possible privacy threats. We propose a new classification of works in the literature based on architectural choices and learning strategies.

Threats to Privacy-Oriented Recommender Systems. Depending on the architecture, the parts of the architecture an actor can access, and how actors are trustworthy, there could be several privacy weaknesses and threats [20, 31]. This part of the tutorial explores and analyzes the privacy threats of RSs, focusing on the algorithmic solutions for the most relevant threats in the different learning paradigms.

2.3 Recent Techniques for Privacy-Preserving Recommender Systems

Private machine learning applications usually are built on two categories of solutions [20], whose fundamentals we outline in this part of the tutorial. On the one hand, cryptographic solutions (e.g., homomorphic encryption and multi-party cryptographic protocols) [29] provide rigorous security guarantees and can perform inference on encrypted data without degrading accuracy but are computationally expensive or introduce a significant communication overhead. On the other hand, solutions based on differential privacy [16] achieve mathematically strong privacy guarantees by adding a calibrated amount of noise to the model, the inputs, or the output results but cause degradation of model performance.

2.4 Trending Research and Open Challenges

In order to provide an overview of the current state of research, this part of the tutorial shows a diverse range of trending solutions for privacy in RSs, from differential privacy to cryptographic protocols, ranging from state-of-the-art techniques to more handcrafted approaches [13, 32, 34, 36]. We cover several contributions where different privacy weaknesses and learning paradigms are analyzed, and we show the main challenges remained open.

2.5 Preserving Privacy at Scale: The Case of Twitter

Following the principles, challenges, and possible solutions discussed before, we focus on the case of Twitter. We focus on how privacy concerns guide Twitter's machine learning development and what extra challenges they pose at scale.

2.6 Hands-on Session

This tutorial includes two practical sessions. The attendees will experiment with some of the algorithmic schemes for privacy-preserving machine learning and adapt RSs learning paradigms based on the possible threats.

Hands-on I: Implementing Privacy-Preserving Techniques in Recommender Systems. After the identification of the main works for privacy-preserving RSs, we provide the attendees with an environment where they can implement some basic privacy-preserving techniques in existing RSs. Moreover, we show how the users can interact with the brand new RSs reproducibility framework Elliot [3], where we added some privacy primitives for facilitating the experimentation of privacy-preserving techniques in a wide range of recommendation approaches.

Hands-on II: Endangering Privacy in Recommender Systems. The implementation of privacy-preserving techniques in RSs protects users' privacy, to an extent depending on the protection scheme and its configuration. In this hands-on session, we practically experiment with how these techniques can mitigate the privacy endangering in RSs.

3 ADDITIONAL INFORMATION

Intended Audience. We target an ***intermediate audience*** of researchers and practitioners willing to delve into the privacy aspects of RSs. We foresee a tutorial of ***180 minutes***. Particular ***prerequisite knowledge or skills are not required*** from the audience, except for a basic understanding of the main concepts in RSs and machine learning.

Supporting Material. The tutorial will be supported by: i) a GitHub repository containing an overview of the program with further details about the tutorial; ii) tutorial slides with references to all the relevant works; iii) two hands-on sessions for experimenting with privacy techniques and endangerment of RSs.

4 TUTORIAL PRESENTERS

Vito Walter Anelli is an Assistant Professor. His research interests fall in the areas of RSs, Knowledge representation, and User Modeling. He has presented two tutorials on Adversarial Learning, and has contributed to a chapter for the 3rd Edition of RS Handbook. He publishes in international venues: SIGIR, ECIR, RecSys, UMAP, SAC, ISWC (best research student paper), and ESWC [3, 5–7, 9]; journals: UMUAI, TKDE, and SWJ, and a book chapter on interpretable RSs. He has served as chair of RecSys challenge 2020/2021, three international Workshops, KaRS 2018/2019/2021 [2], and IIR 2021.

Luca Belli is a Staff Machine Learning researcher and research lead for the Machine learning Ethics, Transparency and Accountability that he co-founded. His research interests are around machine learning feedback loops and intrinsic value of RSs [24]. He was one of the drivers of the publishing of the Twitter RecSys 2020 [10] and

2021 datasets. He was the TA for many under- and graduate courses during his Ph.D.

Yashar Deldjoo is an Assistant Professor. His research focuses on designing multi-modal systems, model robustness, fairness, privacy, and interpretability. He has published two surveys at ACM CUSR. He regularly publishes at SIGIR, RecSys, ECIR, MMSys, ESWC [1, 4–7, 14], and journals, including UMUAI, TKDE, and TIST. He contributed two book chapters to the 3rd Edition of Recommender Systems Handbook. He presented tutorials at IR/RSs venues, including WSDM, RecSys, and ECIR [5]. He has been involved in organizing workshops including ACM RecSys challenge, MediaEval, and IIR '21.

Tommaso Di Noia is Professor of Computer Science. His research activities focus on AI and Data Management. They were initially devoted to knowledge representation and automated reasoning. Then, he studied how to apply knowledge representation techniques to automated negotiations. Following these ideas, he has devoted his interest to applying knowledge graphs and Linked Open Data to RSs with papers published in international journals, conferences, and book chapters [3, 5, 6, 9] During the last years, he moved his research into the Trustworthy AI topic with a particular interest in adversarial ML, explainability, fairness and privacy protection of RSs [5–7, 9].

Antonio Ferrara is a third-year Ph.D. student. He mainly focuses on Federated Learning and its challenges [4], with a particular interest on its relevance for designing privacy-oriented RSs and for modeling federated representations of the users' knowledge. During his studies, he published his works in national and international journals and conferences, including SAC [6], ECIR [5], and SIGIR [3]. He is a lecturer of AI and Computer Science academic courses.

Fedelucio Narducci is an Assistant Professor (tenure track position). He has authored papers in international conferences and journals such as RecSys, ECIR, UMUAI, DSS [5, 6, 21, 25, 26]. He was also co-organizer of the Worshop on Knowledge-aware and Conversational RSs 2018 [2]. His current research interests include RSs, conversational agents, natural language processing, e-health, user modeling and personalization. He has co-authored the book Semantics in Adaptive and Personalised Systems [22] and two chapters for the 2nd and 3rd Edition of RS Handbook.

Claudio Pomo is a second-year Ph.D. student. His research activities mainly focus on Explanation for RSs, with a specific interest in post-hoc methods. He published and presented his works in national and international journals and conferences, including RecSys [8], and SIGIR [3]. He is a lecturer of AI and Computer Science academic courses.

REFERENCES

[1] Jens Adamczak, Yashar Deldjoo, Farshad Bakhshandegan Moghaddam, Peter Knees, Gerard Paul Leyson, and Philipp Monreal. 2021. Session-based Hotel Recommendations Dataset: As part of the ACM Recommender System Challenge 2019. *ACM Trans. Intell. Syst. Technol.* 12, 1 (2021), 1:1–1:20.

[2] Vito Walter Anelli, Pierpaolo Basile, Derek G. Bridge, Tommaso Di Noia, Pasquale Lops, Cataldo Musto, Fedelucio Narducci, and Markus Zanker. 2018. Knowledge-aware and conversational recommender systems. In *12th ACM Conf. on Recommender Systems, RecSys, Canada.* ACM, 521–522. https://doi.org/10.1145/3240323.3240338

[3] Vito Walter Anelli, Alejandro Bellogín, Antonio Ferrara, Daniele Malitesta, Felice Antonio Merra, Claudio Pomo, Francesco M. Donini, and Tommaso Di Noia. 2021. Elliot: a Comprehensive and Rigorous Framework for Reproducible Recommender Systems Evaluation. In *44th Int. ACM Conf. on Research and Development in Information Retrieval, SIGIR, Canada.* https://doi.org/10.1145/3404835.3463245

[4] Vito Walter Anelli, Yashar Deldjoo, Tommaso Di Noia, and Antonio Ferrara. 2019. Towards Effective Device-Aware Federated Learning. In *AI*IA 2019 18th Int. Conf. of the Italian Association for Artificial Intelligence, Italy.* 477–491. https://doi.org/10.1007/978-3-030-35166-3_34

[5] Vito Walter Anelli, Yashar Deldjoo, Tommaso Di Noia, Antonio Ferrara, and Fedelucio Narducci. 2021. FedeRank: User Controlled Feedback with Federated Recommender Systems. In *ECIR 2021, Virtual Event, Part I (LNCS, Vol. 12656).* Springer, 32–47. https://doi.org/10.1007/978-3-030-72113-8_3

[6] Vito Walter Anelli, Yashar Deldjoo, Tommaso Di Noia, Antonio Ferrara, and Fedelucio Narducci. 2021. How to put users in control of their data in federated top-N recommendation with learning to rank. In *SAC '21: 36th Symposium on Applied Computing, Republic of Korea.* ACM, 1359–1362.

[7] Vito Walter Anelli, Yashar Deldjoo, Tommaso Di Noia, Eugenio Di Sciascio, and Felice Antonio Merra. 2020. SAShA: Semantic-Aware Shilling Attacks on Recommender Systems Exploiting Knowledge Graphs. In *17th Int. Conf. ESWC, Greece.* 307–323. https://doi.org/10.1007/978-3-030-49461-2_18

[8] Vito Walter Anelli, Tommaso Di Noia, Eugenio Di Sciascio, Claudio Pomo, and Azzurra Ragone. 2019. On the discriminative power of hyper-parameters in cross-validation and how to choose them. In *Proc.of the 13th ACM Conf. on Recommender Systems, RecSys 2019, Copenhagen, Denmark, September 16-20, 2019*, Toine Bogers, Alan Said, Peter Brusilovsky, and Domonkos Tikk (Eds.). ACM, 447–451. https://doi.org/10.1145/3298689.3347010

[9] Vito Walter Anelli, Tommaso Di Noia, Eugenio Di Sciascio, Azzurra Ragone, and Joseph Trotta. 2019. How to Make Latent Factors Interpretable by Feeding Factorization Machines with Knowledge Graphs. In *ISWC 18th Int. Semantic Web Conf., New Zealand, Part I (LNCS, Vol. 11778).* Springer, 38–56. https://doi.org/10.1007/978-3-030-30793-6_3

[10] Luca Belli, Sofia Ira Ktena, Alykhan Tejani, Alexandre Lung-Yut-Fon, Frank Portman, Xiao Zhu, Yuanpu Xie, Akshay Gupta, Michael Bronstein, Amra Delić, Gabriele Sottocornola, Walter Anelli, Nazareno Andrade, Jessie Smith, and Wenzhe Shi. 2020. Privacy-Aware Recommender Systems Challenge on Twitter's Home Timeline. *arXiv e-prints*, Article arXiv:2004.13715 (April 2020), arXiv:2004.13715 pages. arXiv:2004.13715 [cs.SI]

[11] Laura Burbach, Johannes Nakayama, Nils Plettenberg, Martina Ziefle, and André Calero Valdez. 2018. User preferences in recommendation algorithms: the influence of user diversity, trust, and product category on privacy perceptions in recommender algorithms. In *Proc.of the 12th ACM Conf. on Recommender Systems, RecSys 2018, Vancouver, BC, Canada, October 2-7, 2018.* 306–310. https://doi.org/10.1145/3240323.3240393

[12] California State Legislature. 2018. *The California Consumer Privacy Act of 2018.* Retrieved May 2020 from https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180AB375

[13] Chaochao Chen, Ziqi Liu, Peilin Zhao, Jun Zhou, and Xiaolong Li. 2018. Privacy Preserving Point-of-Interest Recommendation Using Decentralized Matrix Factorization. In *32 AAAI Conf. on Artificial Intelligence, USA.* 257–264.

[14] Yashar Deldjoo, Markus Schedl, Paolo Cremonesi, and Gabriella Pasi. 2020. Recommender systems leveraging multimedia content. *ACM Computing Surveys (CSUR)* 53, 5 (2020), 1–38.

[15] Erika Duriakova, Elias Z. Tragos, Barry Smyth, Neil Hurley, Francisco J. Peña, Panagiotis Symeonidis, James Geraci, and Aonghus Lawlor. 2019. PDMFRec: a decentralised matrix factorisation with tunable user-centric privacy. In *13th ACM Conf. on Recommender Systems, RecSys.* 457–461.

[16] Cynthia Dwork and Aaron Roth. 2014. The Algorithmic Foundations of Differential Privacy. *Found. Trends Theor. Comput. Sci.* 9, 3-4 (2014), 211–407. https://doi.org/10.1561/0400000042

[17] European Commission. 2018. *2018 reform of EU data protection rules.* Retrieved May 2020 from https://ec.europa.eu/info/priorities/justice-and-fundamental-rights/data-protection/2018-reform-eu-data-protection-rules/eu-data-protection-rules_en

[18] Arik Friedman, Bart P. Knijnenburg, Kris Vanhecke, Luc Martens, and Shlomo Berkovsky. 2015. Privacy Aspects of Recommender Systems. In *Recommender Systems Handbook.* 649–688. https://doi.org/10.1007/978-1-4899-7637-6_19

[19] Bart P. Knijnenburg and Shlomo Berkovsky. 2017. Privacy for Recommender Systems: Tutorial Abstract. In *Proc.of the Eleventh ACM Conf. on Recommender Systems, RecSys 2017, Como, Italy, August 27-31, 2017.* 394–395. https://doi.org/10.1145/3109859.3109935

[20] Bo Liu, Ming Ding, Sina Shaham, Wenny Rahayu, Farhad Farokhi, and Zihuai Lin. 2021. When Machine Learning Meets Privacy: A Survey and Outlook. *ACM Comput. Surv.* 54, 2 (2021), 31:1–31:36. https://doi.org/10.1145/3436755

[21] Pasquale Lops, Marco de Gemmis, Giovanni Semeraro, Fedelucio Narducci, and Cataldo Musto. 2011. Leveraging the linkedin social network data for extracting content-based user profiles. In *ACM Conf. on Recommender Systems, RecSys, USA.* ACM, 293–296. https://doi.org/10.1145/2043932.2043986

[22] Pasquale Lops, Cataldo Musto, Fedelucio Narducci, and Giovanni Semeraro. 2019. *Semantics in Adaptive and Personalised Systems - Methods, Tools and Applications.* Springer. https://doi.org/10.1007/978-3-030-05618-6

[23] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Agüera y Arcas. 2017. Communication-Efficient Learning of Deep Networks from Decentralized Data. In *20th Int. Conf. on Artificial Intelligence and Statistics, AISTATS, USA.* 1273–1282.

[24] Smitha Milli, Luca Belli, and Moritz Hardt. 2021. From Optimizing Engagement to Measuring Value. In *Proc.of the 2021 ACM Conf. on Fairness, Accountability, and Transparency* (Canada) *(FAccT)*. Association for Computing Machinery, New York, NY, USA, 714–722. https://doi.org/10.1145/3442188.3445933

[25] Fedelucio Narducci, Pierpaolo Basile, Marco de Gemmis, Pasquale Lops, and Giovanni Semeraro. 2020. An investigation on the user interaction modes of conversational recommender systems for the music domain. *UMUAI* 30, 2 (2020), 251–284. https://doi.org/10.1007/s11257-019-09250-7

[26] Fedelucio Narducci, Matteo Palmonari, and Giovanni Semeraro. 2014. CroSeR: Cross-language Semantic Retrieval of Open Government Data. In *36th European Conf. on IR Research, ECIR, The Netherlands (LNCS, Vol. 8416)*. Springer, 793–797. https://doi.org/10.1007/978-3-319-06028-6_98

[27] Amar Saini, Florin Rusu, and Andrew Johnston. 2019. PrivateJobMatch: a privacy-oriented deferred multi-match recommender system for stable employment. In *13th ACM Conf. on Recommender Systems, RecSys, Denmark*. 87–95. https://doi.org/10.1145/3298689.3346983

[28] Hyejin Shin, Sungwook Kim, Junbum Shin, and Xiaokui Xiao. 2018. Privacy Enhanced Matrix Factorization for Recommendation with Local Differential Privacy. *IEEE Trans. Knowl. Data Eng.* 30, 9 (2018), 1770–1782. https://doi.org/10.1109/TKDE.2018.2805356

[29] Erez Shmueli and Tamir Tassa. 2017. Secure Multi-Party Protocols for Item-Based Collaborative Filtering. In *Proc.of the Eleventh ACM Conf. on Recommender Systems, RecSys 2017, Como, Italy, August 27-31, 2017*. 89–97. https://doi.org/10.1145/3109859.3109881

[30] Standing Committee of the National People's Congress of Popular Republic of China. 2017. *China Internet Security Law.* Retrieved May 2020 from http://www.npc.gov.cn/npc/c1481/201507/82ce4cb5549c4f56be8a6744cf2b3273.shtml

[31] Soumya Wadhwa, Saurabh Agrawal, Harsh Chaudhari, Deepthi Sharma, and Kannan Achan. 2020. Data Poisoning Attacks against Differentially Private Recommender Systems. In *43rd ACM Conf. on research and development in IR SIGIR, China*. 1617–1620. https://doi.org/10.1145/3397271.3401301

[32] Aidmar Wainakh, Tim Grube, Jörg Daubert, and Max Mühlhäuser. 2019. Efficient privacy-preserving recommendations based on social graphs. In *Proc.of the 13th ACM Conf. on Recommender Systems, RecSys 2019, Copenhagen, Denmark, September 16-20, 2019*. 78–86. https://doi.org/10.1145/3298689.3347013

[33] Huazheng Wang, Qian Zhao, Qingyun Wu, Shubham Chopra, Abhinav Khaitan, and Hongning Wang. 2020. Global and Local Differential Privacy for Collaborative Bandits. In *RecSys 14th ACM Conf. on Recommender Systems, Brazil*. 150–159. https://doi.org/10.1145/3383313.3412254

[34] Jun Wang, Qiang Tang, Afonso Arriaga, and Peter Y. A. Ryan. 2019. Novel Collaborative Filtering Recommender Friendly to Privacy Protection. In *28th Int. Joint Conf. on Artificial Intelligence, IJCAI, China*. 4809–4815. https://doi.org/10.24963/ijcai.2019/668

[35] Qiang Yang, Yang Liu, Yong Cheng, Yan Kang, Tianjian Chen, and Han Yu. 2019. *Federated Learning*. Morgan & Claypool Publishers. https://doi.org/10.2200/S00960ED2V01Y201910AIM043

[36] Haotian Zhou, Xiao-Yang Liu, Cai Fu, Chen Shang, and Xinyi Chang. 2018. Differentially Private Matrix Completion via Distributed Matrix Factorization. In *17th IEEE Int. Conf. On Trust, Security And Privacy In Computing And Communications, USA*. 1628–1631.