Leveraging Graph Neural Networks for User Profiling: Recent Advances and Open Challenges

Erasmo Purificato
Otto von Guericke University
Magdeburg, Germany
Leibniz Institute for Educational
Media | Georg Eckert Institute
Brunswick, Germany
erasmo.purificato@ovgu.de

Ludovico Boratto
Department of Mathematics and
Computer Science,
University of Cagliari
Cagliari, Italy
ludovico.boratto@acm.org

Ernesto William De Luca
Otto von Guericke University
Magdeburg, Germany
Leibniz Institute for Educational
Media | Georg Eckert Institute
Brunswick, Germany
ernesto.deluca@ovgu.de

ABSTRACT

The proposed tutorial aims to familiarise the CIKM community with modern user profiling techniques that utilise Graph Neural Networks (GNNs). Initially, we will delve into the foundational principles of user profiling and GNNs, accompanied by an overview of relevant literature. We will subsequently systematically examine cutting-edge GNN architectures specifically developed for user profiling, highlighting the typical data utilised in this context. Furthermore, ethical considerations and beyond-accuracy perspectives, e.g. fairness and explainability, will be discussed regarding the potential applications of GNNs in user profiling. During the hands-on session, participants will gain practical insights into constructing and training recent GNN models for user profiling using open-source tools and publicly available datasets. The audience will actively explore the impact of these models through case studies focused on bias analysis and explanations of user profiles. To conclude the tutorial, we will analyse existing and emerging challenges in the field and discuss future research directions.

CCS CONCEPTS

• Human-centered computing → User models; **• Social and professional topics → User characteristics**; **• Applied computing → Law, social and behavioral sciences**.

KEYWORDS

User Profiling, Graph Neural Networks, Fairness, Explainability

ACM Reference Format:
Erasmo Purificato, Ludovico Boratto, and Ernesto William De Luca. 2023. Leveraging Graph Neural Networks for User Profiling: Recent Advances and Open Challenges. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (CIKM '23), October 21–25, 2023, Birmingham, United Kingdom.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3583780.3615292

1 MOTIVATION

The primary objective of **user profiling** is to deduce an individual's interests, personality traits, or behaviours based on collected data in order to create an effective user representation known as a *user model*. Constructing such a user model holds significant importance, particularly in the context of adaptive and personalised systems [9], as it is commonly used to develop recommender systems [18]. The significance of this field of research has grown in recent years due to the vast amount of data shared by millions of online users on a daily basis. As a result, *user profiling* has become a crucial topic in various domains, including social networks [17], e-commerce [31], and recommender systems [12].

In the past, early user profiling techniques primarily focused on examining static characteristics, which involved gathering information from online forms and surveys (*explicit user profiling*) [19]. However, these methods faced limitations as users were often unwilling to share their personal information directly. Consequently, modern systems have shifted towards *implicit user profiling*, also known as *behavioural user profiling* [15], which involves inferring user profiles based on their actions and interactions. Graph structures have emerged as a popular approach for representing these behaviours, where nodes represent users and edges signify interactions between them. By leveraging these graph structures, user profiling can capture the dynamic nature of user behaviour and provide a more comprehensive understanding of their preferences and interests. This shift from explicit to implicit user profiling has allowed for more effective and privacy-preserving methods of profiling users. **Graph Neural Networks** (GNNs) [13, 16, 29, 37, 38] have proven to be successful in modelling graph data across various fields, such as recommender systems [14, 35], natural language processing [33] and text mining [27], and several contributions have been already emerged for user profiling [5, 6, 23, 32].

The evaluation of existing user profiling techniques commonly involves assessing the models' ability to classify different types of profiles [6]. In recent years, there has been a heightened emphasis on ethical principles guiding the development and utilization of artificial intelligence (AI) systems across various research fields [10]. This increased focus on ethics has prompted attention towards transparency [30], privacy [22], and social equity [11] issues. Researchers now recognise the importance of ensuring transparency in AI systems, allowing users to understand how their data is being utilised and how profiling decisions are made. Additionally, privacy concerns have become a significant consideration, emphasising the need for robust mechanisms to protect users' sensitive information during the profiling process. Moreover, there is a growing recognition of the social equity implications of user profiling, aiming to

CIKM '23, October 21–25, 2023, Birmingham, United Kingdom

ACM ISBN 979-8-4007-0124-5/23/10.
https://doi.org/10.1145/3583780.3615292

address biases and promote fairness in developing and applying these techniques.

The proposed tutorial will explore and analyse two crucial *beyond-accuracy perspectives* in GNN-based models for user profiling: *fairness* and *explainability*. A comprehensive understanding of these perspectives will be offered, emphasising their significance and impact in the context of user profiling.

When considering *fairness*, one particular challenge when using GNNs to classify user profiles is their susceptibility to learning biases from historical data, leading to biased outputs. The graph structure and message-passing process utilised by GNNs can contribute to this issue by amplifying discrimination. This occurs because nodes within the same group are more likely to be connected to each other compared to nodes from different groups, as highlighted in prior work [24]. **Algorithmic fairness** has gained prominence within the realm of decision-making systems. Numerous studies have been conducted to detect and mitigate biases in machine learning models [3, 4, 21]. However, only a limited number of these studies specifically address fairness concerns in GNNs used for user profiling models. Noteworthy examples include FairGNN [8] and our analysis of state-of-the-art models for behavioural user profiling [1, 20]. As the focus on fairness in GNN-based user profiling models is still relatively nascent, the tutorial will shed light on these important research developments and their implications.

Explainability is another emerging research topic in automated systems, aiming to provide human interpretable insights into AI models [25]. This research area has gained significance due to specific regulations such as the EU GDPR, which explicitly requires users to understand how and why a particular system generates a specific outcome. Although some sporadic work has been conducted in the field of explainable user models [2, 7], it remains an open research area that demands further exploration. General explainability approaches designed for "general" GNNs can be applied to address this challenge, such as the approaches presented in [34, 36]. The tutorial will delve into this direction, exploring the application of these explainability approaches to user profiling architectures. The research results in this area will be shared during the tutorial sessions, offering valuable insights into the current advancements and future prospects of explainable user models in the context of GNN-based user profiling.

By the end of this tutorial, participants will acquire a comprehensive understanding of essential concepts about GNN models for user profiling. They will gain the ability to assess the fairness of these models, evaluate the explainability of state-of-the-art architectures, and critically analyse their impact on relevant stakeholders. Furthermore, participants will be equipped to identify the challenges and opportunities that exist in this field. The tutorial aims to empower attendees with the knowledge and skills necessary to navigate and contribute to the ongoing advancements in GNN-based user profiling, enabling them to make informed decisions and drive future research in this domain.

2 OUTLINE

During the initial session of our tutorial, we will present and discuss the existing literature on user profiling approaches in machine learning, with a specific focus on GNN-based models tailored for this purpose. Our discussion will encompass the latest advancements in this field, shedding light on the cutting-edge techniques and approaches used to address two critical beyond-accuracy perspectives: fairness and explainability. These topics will highlight how state-of-the-art methods tackle these timely concerns. To ensure an engaging and interactive experience, we will prepare practical exercises for each session. These exercises will provide participants with hands-on opportunities to apply the concepts and techniques discussed in the tutorial. We intend to publish all the resources used throughout the tutorial on our webpage to facilitate further exploration and learning. This includes slides, notebooks, reference lists, pre-processed datasets, and any relevant source code. By making these resources readily available, participants can access and review the materials at their convenience, enabling them to reinforce their understanding and delve deeper into the presented topics.

Below, you will find a brief overview of the expected outline, including a description of the practical session[1].

- **User profiling session** - *Lecture slides and Jupyter Notebooks in Python* - ~90 mins
  - **Opening and instructors' presentation** (5 mins)
  - **Introduction to user profiling** (15 mins): A historical overview of the research on user profiling is given to set the basis for understanding the recent advances. In particular, starting from the definition of the key terms in the field (e.g. the difference between *explicit* and *implicit user profiling* [15, 19]), we shortly illustrate several user profiling contributions in different domains [12, 17, 18, 31].
  - **Introduction to GNNs** (15 mins): As for the user profiling part, in this part of the tutorial, we cover the most important notions about *graph neural networks*, such as basic terminology and most popular architectures (e.g. GCN [16] and GAT [29]), to create a common background to make the audience able to follow the core tutorial sections.
  - **GNN-based models for user profiling** (25 mins): We present in detail the current state-of-the-art GNN-based model for user profiling, such as CatGCN [5], RHGN [32] and others, describing their architectures, their training procedures, and discussing their strengths and weaknesses.
  - **Hands on state-of-the-art GNN-based models for user profiling** (25 mins): To show how the described GNNs are designed and implemented, we execute and explain some of the models, e.g. CatGCN [5] and RHGN [32], in their original configuration, illustrating the used datasets for every contribution (i.e. Alibaba[2] and JD[3]).
  - **Q&A** (5 mins)
- **Beyond-accuracy perspectives session** - *Lecture slides and Jupyter Notebooks in Python* - ~90 mins
  - **Overview on beyond-accuracy perspectives for user profiling with GNNs: Algorithmic Fairness** (15 mins): The fairness aspects encountered in user profiling research are discussed. Several points of view are taken into account: (1) analysis of the beyond-accuracy perspectives in

[1] Due to the page limit for this proposal, the literature referenced in the outline is not meant to be complete. It provides leads to the most relevant papers. The tutorial covers over 50 references, with most of the challenges and solutions presented in detail.
[2] https://tianchi.aliyun.com/dataset/dataDetail?dataId=56
[3] https://github.com/guyulongcs/IJCAI2019_HGAT

existing state-of-the-art GNN-based models [20]; (2) description of models designed for the precise purpose, such as FairGNN [8]; (3) application of general GNN approaches for fairness to a specific user profiling case study.

- **Use cases on related beyond-accuracy perspectives (Fairness)** (35 mins): First, we illustrate the implementation of four standard fairness metrics to the illustrated models. Then, we describe FairGNN [8], a state-of-the-art GNN-based framework for debiasing, and run it on the datasets used in the original publication (i.e. NBA[4] and Pokec [28]). We also apply the three approaches described in the last section of the theoretical session by making use of our standardised framework developed for analysing fairness in GNN-based models for user profiling [1]. In particular, the framework can get a graph structure *NetworkX*[5] or *Neo4J*[6] format and generate the specific input for each of the GNNs.
- **Overview on beyond-accuracy perspectives for user profiling with GNNs: Explainability** (10 mins): The explainability aspects encountered in user profiling research are discussed by presenting some of the most recent works which try to address this challenge (e.g. [26]).
- **Open challenges and concluding remarks** (15 mins): We discuss the existing open challenges in user profiling research and wrap up the tutorial content.
- **Q&A** (10 mins)
- **Closing** (5 mins)

3 TARGETED AUDIENCE AND PREREQUISITES

The tutorial is designed to benefit the CIKM community, including researchers, industry technologists and practitioners with a *beginner* or *intermediate* level of expertise in the field. It caters to individuals without prior knowledge of user profiling and GNNs, ensuring accessibility to a wide range of participants.

While a basic understanding of Python programming is preferred, along with familiarity with common data science libraries like Pandas and NumPy, the tutorial provides all the necessary background information to accommodate individuals with varying levels of programming experience. It is important to note that the tutorial goes beyond technical aspects, delving into key concepts that have applications in interdisciplinary fields. This broader scope makes the tutorial valuable and engaging for an interdisciplinary audience as well, expanding the potential reach and impact of the outlined topics.

4 INSTRUCTORS' BIOGRAPHY

Erasmo Purificato (webpage: https://erasmopurif.com/; email: erasmo.purificato@ovgu.de) has been a Research Assistant in the "Human-Centred AI" group at the Otto von Guericke University Magdeburg and in the "Human-Centred Technologies for Educational Media" department at the Leibniz Institute for Educational Media | Georg Eckert Institute (GEI), since February 2020. He is doing his Ph.D. in Computer Science at OVGU with a project entitled "Human-Centred Fairness Analysis on Graph Neural Network-Based Models for Behavioural User Profiling" focusing on Human-Centred AI techniques. From July 2021 he has been appointed by the Guglielmo Marconi University of Rome, Italy, as an Adjunct Professor in "Software Engineering". Between September 2017 and December 2019, he worked at Blue Reply company in Turin as a Machine Learning Engineer and Technical Leader in the Cognitive&Data business unit, developing projects in computer vision and content-based image retrieval fields. He co-organised the APEx-UI 2022 workshop co-located with IUI 2022 and the IEEE Autumn School ISACT 2022 co-located with ICHMS 2022, where he gave a talk about "Beyond-Accuracy Perspectives: Explainability and Fairness". He has been part of the program committee of several conferences and workshops, such as SIGIR, RecSys, UMAP, HT and ExUM, and served as a reviewer for journals, such as TORS.

Ludovico Boratto (webpage: https://www.ludovicoboratto.com/; email: ludovico.boratto@acm.org) is an Assistant Professor at the Department of Mathematics and Computer Science of the University of Cagliari (Italy). His research interests focus on recommender systems and their impact on stakeholders. He has authored more than 60 papers and published his research in top-tier conferences and journals. His research activity brought him to give talks and tutorials at top-tier conferences and research centres. He is the editor of the book "Group Recommender Systems: An Introduction" by Springer. He is an editorial board member of the "Information Processing & Management" journal (Elsevier) and "Journal of Intelligent Information Systems" (Springer), and guest editor of several journals' special issues. He is part of the program committees of the main Web conferences, where he received four outstanding contribution awards. He has given several tutorials in top-tier conferences, including RecSys 2022, ICDE 2021, ECIR 2021, WSDM 2021, ICDM 2020, UMAP 2020, DSAA 2018, ICDM 2017, RecSys 2016, ECAI 2016, and ECIR 2016. In 2012 he got his Ph.D. at the University of Cagliari. From May 2016 to April 2021, he joined Eurecat as Senior Research Scientist in the Data Science and Big Data Analytics research group. In 2010 and 2014, he spent ten months at Yahoo! Research in Barcelona as a visiting researcher. He is a member of ACM and IEEE.

Ernesto William De Luca (webpage: https://ernestodeluca.eu/; email: deluca@gei.de) is head of the "Human-Centred Technologies for Educational Media" department at the Leibniz Institute for Educational Media | Georg Eckert Institute (GEI) and from October 2019 has been appointed as a Full Professor in "Research Infrastructures for Digital Humanities" at the Otto von Guericke University Magdeburg, Germany. In addition, in May 2015, the Guglielmo Marconi University of Rome appointed him an associate professor in "Computational Engineering". He studied computational linguistics and then gained his Ph.D. in computer science. His research includes machine learning, human-machine interaction, Natural Language Processing, user and data modelling, the Semantic Web and Information Retrieval. He has written over 100 papers for national and international conferences and journals and organized and chaired numerous workshops and conferences. He is a regular reviewer and programme committee member of different high-profile journals and conferences.

[4]https://www.kaggle.com/datasets/noahgift/social-power-nba
[5]https://networkx.org/
[6]https://neo4j.com/

REFERENCES

[1] Mohamed Abdelrazek, Erasmo Purificato, Ludovico Boratto, and Ernesto William De Luca. 2023. FairUP: A Framework for Fairness Analysis of Graph Neural Network-Based User Profiling Models. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Taipei, Taiwan) *(SIGIR '23)*. Association for Computing Machinery, New York, NY, USA, 3165–3169. https://doi.org/10.1145/3539618.3591814

[2] Krisztian Balog, Filip Radlinski, and Shushan Arakelyan. 2019. Transparent, scrutable and explainable user models for personalized recommendation. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 265–274.

[3] Solon Barocas, Moritz Hardt, and Arvind Narayanan. 2019. *Fairness and Machine Learning*. fairmlbook.org. http://www.fairmlbook.org.

[4] Simon Caton and Christian Haas. 2020. Fairness in machine learning: A survey. *arXiv preprint arXiv:2010.04053* (2020).

[5] Weijian Chen, Fuli Feng, Qifan Wang, Xiangnan He, Chonggang Song, Guohui Ling, and Yongdong Zhang. 2021. CatGCN: Graph Convolutional Networks with Categorical Node Features. *IEEE Transactions on Knowledge and Data Engineering* (2021).

[6] Weijian Chen, Yulong Gu, Zhaochun Ren, Xiangnan He, Hongtao Xie, Tong Guo, Dawei Yin, and Yongdong Zhang. 2019. Semi-supervised user profiling with heterogeneous graph attention networks. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*. 2116–2122.

[7] David Jaime Tena Cucala, Bernardo Cuenca Grau, Egor V Kostylev, and Boris Motik. 2021. Explainable GNN-Based Models over Knowledge Graphs. In *International Conference on Learning Representations*.

[8] Enyan Dai and Suhang Wang. 2021. Say no to the discrimination: Learning fair graph neural networks with limited sensitive attribute information. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*. 680–688.

[9] Christopher Ifeanyi Eke, Azah Anir Norman, Liyana Shuib, and Henry Friday Nweke. 2019. A survey of user profiling: State-of-the-art, challenges, and solutions. *IEEE Access* 7 (2019), 144907–144924.

[10] European-Commission. 2019. *Ethics guidelines for trustworthy AI*. Publications Office. https://doi.org/doi/10.2759/346720

[11] Elizabeth Gómez, Carlos Shui Zhang, Ludovico Boratto, Maria Salamó, and Mirko Marras. 2021. The winner takes it all: geographic imbalance and provider (un) fairness in educational recommender systems. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1808–1812.

[12] Yulong Gu, Zhuoye Ding, Shuaiqiang Wang, and Dawei Yin. 2020. Hierarchical user profiling for e-commerce recommender systems. In *Proceedings of the 13th International Conference on Web Search and Data Mining*. 223–231.

[13] Will Hamilton, Zhitao Ying, and Jure Leskovec. 2017. Inductive representation learning on large graphs. *Advances in neural information processing systems* 30 (2017).

[14] 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*. 639–648.

[15] Sumitkumar Kanoje, Sheetal Girase, and Debajyoti Mukhopadhyay. 2015. User profiling trends, techniques and applications. *arXiv preprint arXiv:1503.07474* (2015).

[16] Thomas N. Kipf and Max Welling. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *5th International Conference on Learning Representations, ICLR 2017, Conference Track Proceedings*.

[17] Lizi Liao, Xiangnan He, Hanwang Zhang, and Tat-Seng Chua. 2018. Attributed social network embedding. *IEEE Transactions on Knowledge and Data Engineering* 30, 12 (2018), 2257–2270.

[18] Mohammad Naiseh, Nan Jiang, Jianbing Ma, and Raian Ali. 2020. Personalising explainable recommendations: Literature and conceptualisation. In *World Conference on Information Systems and Technologies*. Springer, 518–533.

[19] Danny Poo, Brian Chng, and Jie-Mein Goh. 2003. A hybrid approach for user profiling. In *36th Annual Hawaii International Conference on System Sciences, 2003. Proceedings of the*. IEEE, 9–13.

[20] Erasmo Purificato, Ludovico Boratto, and Ernesto William De Luca. 2022. Do Graph Neural Networks Build Fair User Models? Assessing Disparate Impact and Mistreatment in Behavioural User Profiling. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*. 4399–4403.

[21] Erasmo Purificato, Flavio Lorenzo, Francesca Fallucchi, and Ernesto William De Luca. 2022. The Use of Responsible Artificial Intelligence Techniques in the Context of Loan Approval Processes. *International Journal of Human–Computer Interaction* (2022), 1–20.

[22] Erasmo Purificato, Sabine Wehnert, and Ernesto William De Luca. 2021. Dynamic Privacy-Preserving Recommendations on Academic Graph Data. *Computers* 10, 9 (2021), 107.

[23] Afshin Rahimi, Trevor Cohn, and Timothy Baldwin. 2018. Semi-supervised user geolocation via graph convolutional networks. *arXiv preprint arXiv:1804.08049* (2018).

[24] Tahleen Rahman, Bartlomiej Surma, Michael Backes, and Yang Zhang. 2019. Fairwalk: towards fair graph embedding. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*. 3289–3295.

[25] Wojciech Samek, Grégoire Montavon, Andrea Vedaldi, Lars Kai Hansen, and Klaus-Robert Müller. 2019. *Explainable AI: interpreting, explaining and visualizing deep learning*. Vol. 11700. Springer Nature.

[26] Emily Sullivan, Dimitrios Bountouridis, Jaron Harambam, Shabnam Najafian, Felicia Loecherbach, Mykola Makhortykh, Domokos Kelen, Daricia Wilkinson, David Graus, and Nava Tintarev. 2019. Reading news with a purpose: Explaining user profiles for self-actualization. In *Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization*. 241–245.

[27] Zhiqing Sun, Jian Tang, Pan Du, Zhi-Hong Deng, and Jian-Yun Nie. 2019. Divgraphpointer: A graph pointer network for extracting diverse keyphrases. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 755–764.

[28] Lubos Takac and Michal Zabovsky. 2012. Data analysis in public social networks. In *International scientific conference and international workshop present day trends of innovations*, Vol. 1. Present Day Trends of Innovations Lamza Poland.

[29] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2017. Graph attention networks. *arXiv preprint arXiv:1710.10903* (2017).

[30] Qianwen Wang, Yao Ming, Zhihua Jin, Qiaomu Shen, Dongyu Liu, Micah J Smith, Kalyan Veeramachaneni, and Huamin Qu. 2019. Atmseer: Increasing transparency and controllability in automated machine learning. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.

[31] Chuhan Wu, Fangzhao Wu, Junxin Liu, Shaojian He, Yongfeng Huang, and Xing Xie. 2019. Neural demographic prediction using search query. In *Proceedings of the 12th ACM International Conference on Web Search and Data Mining*. 654–662.

[32] Qilong Yan, Yufeng Zhang, Qiang Liu, Shu Wu, and Liang Wang. 2021. Relation-aware Heterogeneous Graph for User Profiling. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. 3573–3577.

[33] Liang Yao, Chengsheng Mao, and Yuan Luo. 2019. Graph convolutional networks for text classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 7370–7377.

[34] Rex Ying, Dylan Bourgeois, Jiaxuan You, Marinka Zitnik, and Jure Leskovec. 2019. Gnn explainer: A tool for post-hoc explanation of graph neural networks. *arXiv preprint arXiv:1903.03894* (2019).

[35] Rex Ying, Ruining He, Kaifeng Chen, Pong Eksombatchai, William L Hamilton, and Jure Leskovec. 2018. Graph convolutional neural networks for web-scale recommender systems. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 974–983.

[36] Hao Yuan, Haiyang Yu, Shurui Gui, and Shuiwang Ji. 2022. Explainability in graph neural networks: A taxonomic survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2022).

[37] Chuxu Zhang, Dongjin Song, Chao Huang, Ananthram Swami, and Nitesh V Chawla. 2019. Heterogeneous graph neural network. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 793–803.

[38] Ziwei Zhang, Peng Cui, and Wenwu Zhu. 2022. Deep Learning on Graphs: A Survey. *IEEE Transactions on Knowledge and Data Engineering* 34, 1 (2022), 249–270.