Beyond-Accuracy Perspectives on Graph Neural Network-Based Models for Behavioural User Profiling

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

ABSTRACT

The presented doctoral research aims to develop a behavioural user profiling framework focusing simultaneously on three *beyond-accuracy* perspectives: *privacy*, to study how to intervene on graph data structures of specific contexts and provide methods to make the data available in a meaningful manner without neither exposing personal user information nor corrupting the profiles creation and system performances; *fairness*, to provide user representations that are free of any inherited discrimination which could affect a downstream recommender by developing debiasing approaches to be applied on state-of-the-art GNN-based user profiling models; *explainability*, to produce understandable descriptions of the framework results, both for user profiles and recommendations, mainly in terms of interaction importance, by designing an adaptive and personalised user interface which provides tailored explanations to the end-users, depending on their specific user profiles.

CCS CONCEPTS

• **Information systems → Recommender systems**; **Personalization**; • **Security and privacy**; • **Human-centered computing → User models**; **User interface programming**; • **Applied computing → Law, social and behavioral sciences**;

KEYWORDS

User Profiling, Graph Neural Networks, Fairness, Explainability, Privacy, User Interfaces

ACM Reference Format:
Erasmo Purificato. 2022. Beyond-Accuracy Perspectives on Graph Neural Network-Based Models for Behavioural User Profiling. In *Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization (UMAP '22), July 4–7, 2022, Barcelona, Spain.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3503252.3534361

1 GENERAL INFORMATION

- **Name and Surname**: Erasmo Purificato;
- **University where conducting Ph.D.**: Otto von Guericke University Magdeburg, Germany;
- **Main advisor**: Prof. Dr.-Ing. Ernesto William De Luca;
- **Affiliations**: Research assistant at Otto von Guericke University Magdeburg, Germany (50%) and at Leibniz Institute for Educational Media | Georg Eckert Institute, Brunswick, Germany (50%);
- **Current year of study**: Third (currently enrolled in the fifth semester);
- **Estimated completion date**: Second half of 2024 (date intended as thesis submission);
- **Regulations of the Ph.D. program**: No specific regulations about the lenght of a Ph.D. program in our university. For a student working simultaneously at the university and a research institute, the typical completion time is 4-6 years. At the end of the sixth year, the group of supervisors (two internal supervisors have to be appointed) can decide, only in case of proven absence of progress, to remove the student from the Ph.D program.

2 BACKGROUND, MOTIVATION AND RELATED WORK

Due to the extensive amount of data provided by web applications and platforms nowadays, *user profiling* has become a key topic in many real-world applications, especially recommender systems [22] and social networks [19]. The main goal of user profiling is to infer an individual's interests, personality traits or behaviours from generated data to create an efficient user representation, i.e. a *user model*, whose construction is particularly important in the context of adaptive and personalised systems, as it is usually considered the main practice adopted to develop recommender systems [23]. Early profiling approaches considered only the analysis of static characteristics (*explicit user profiling*), with data often coming from online forms and surveys [25]. However, these methods have been proved to be ineffective as users are not concerned about providing their information directly. Therefore, modern systems focus more on profiling users' data implicitly based on individuals' actions and interactions (*implicit user profiling*). This approach is also referred to as *behavioural user profiling* [17]. A natural way to model these behaviours is through *graphs*, where edges can easily describe the interactions between users, represented by nodes. In this light, Graph Neural Networks (GNNs) constitute the perfect class of methods to deal with data represented by graph data structures. Recent studies have demonstrated the effectiveness of GNNs in modelling graph data on several domains, such as recommender systems [16], natural language processing [35], text mining [31], as well as user profiling, in particular CatGCN [8] and RHGN [34].

UMAP'22, July 4–7, 2022, Barcelona, Spain
© 2022 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-9207-5/22/07.
https://doi.org/10.1145/3503252.3534361

Generally, existing approaches evaluate user profiling models based on the effectiveness of a classification task at predicting a user's personal characteristics, such as the purchasing level, gender or age. In my thesis, I aim to look *beyond* the usual accuracy-based approaches by developing a **GNN-based behavioural user profiling framework** which takes simultaneously into account the perspectives of *fairness*, *explainability* and *privacy* in order to feed a downstream recommender systems and provide suggestions and tailored explanations to the end-users having different profiles through an adaptive and personalised user interface.

The main challenge I want to address within my Ph.D. project is to leverage GNN models to produce fair user representations for graph structures representing several types of relations between nodes, having furthermore the ability to explain the motivation laying behind each representation, in terms of interaction importance. Adopting GNNs to reach fair results is not a common solution. Even though they are proved to be successful in classifying user profiles, as any machine learning models trained on historical data, GNNs are prone to reproduce in their outcomes the biases learned in such data. This phenomenon is mostly due to the topology of graph data structures and the conventional message-passing process of GNNs, which can amplify discrimination because nodes of the same sensitive attribute are more likely to be linked to each other than those of different values [29]. *Algorithmic fairness* is an increasingly emerging topic for decision-making systems. Many works has been already published about methods to detect and mitigate biases produced by machine learning models [5, 7], but only a few of them are related to fairness on GNNs, especially for user profiling models, such as FairGNN [10].

Another rising topic for automated systems is *explainability*, mainly after the term "Explainable AI" (XAI) being coined in 2017 [13]. It is a key research area having the goal of exposing artificial intelligence (AI) models to humans in an interpretable manner [30], taking care of specific regulations, such as EU GDPR, which explicitly require users to be able to understand why and how a particular result from a system is obtained. Despite the heated debate within the AI community about whether the continuous and persistent search of interpreting how a system works to the end-users is really needed, I consider explainability significant for the framework I am developing for two main reasons, according to Miller [21]: (1) *trust*, because people cannot just get the results, or at most read some statistics about the model performance, and believe a decision is correct; (2) *ethics*, because we should always prove that a developed system is not producing discrimination of any kind. Even though the latter point closely connects the concepts of explainability and fairness, not much work has been published considering both of these aspects at the same time. Just as the development of explainable user models, except for some recent work [4], remains an almost open research challenge to tackle.

Strictly related to the desire of building a valuable, interpretable system, there is the need to develop an understandable and easy-to-use *user interface* (UI), which is currently one of the weakest points of research in the XAI field [1]. UIs play a fundamental role in providing the right explanations to the end-users in order to have a real user-centric experience, even more than implementing the system itself in many cases. Much research about innovative human-centred explainable AI systems has been recently done [20, 21]. However, a common issue in this context is to create UIs following the *one-fits-all* paradigm, meaning that all the users get the same explanations, without considering their different profiles, that are often related to different knowledge, background or expertise.

In addition to the above, the further perspective faced in my thesis considers *privacy* concerns in specific contexts, with the goal of developing a method to protect personal data while at same time allowing a meaningful use of the available data for the intended purpose. Despite privacy issues in personalised systems being under study for a long time [18], it is in the last years that more emphasis has been placed on this question, as users were never really aware of the problem before, especially about what personal data is being used and how securely it is stored [2]. In such a scenario, implementing systems and methodologies that guarantee personal data protection and privacy by design is extremely important, even mandatory in certain circumstances and under specific regulations, and again we can take EU GDPR as an example.

3 RESEARCH GOALS AND METHODOLOGY

The focus of the thesis is to develop a graph neural network-based model for behavioural user profiling whose resulting user models are applied as the input of a recommender system, acting as a bridge with an adaptive and personalised user interface implemented to provide tailored explanations to the end-users depending on their specific user profiles. Once the whole framework is implemented, evaluation of the behavioural user profiling models, as well as of the recommender system, will be carried out mainly on academic graph data, such as the Open Knowledge Research Graph (ORKG) [3]. These kind of sources include information about researchers, scientific papers they published and projects they are involved in, in order to develop specific use cases, such as creation of researcher user profiles or scientific paper or expert recommendation.

The intended framework resulting from this work, whose architecture is shown in Fig.1, deals with three "*beyond-accuracy*" perspectives:

- *Privacy*, to study how to intervene on graph data structures of specific privacy-related contexts and provide methods to make that data available as input without neither exposing personal user information nor corrupting the profiles creation and system performances;
- *Fairness*, to build user representations that are free of any inherited discrimination which could affect the downstream recommendations by developing debiasing approaches to be applied on state-of-the-art GNN-based user profiling models;
- *Explainability*, to produce understandable descriptions of the framework results, both for user profiles and recommendations, mainly in terms of interaction importance; the explanations are not the same for all end-users, but tailored to the different profiles, in order to serve a concrete human-centred experience.

Around the outlined challenges, I formulate the following research questions, which define the main contributions of the doctoral research:

- **RQ1** How can we guarantee personal data protection on a graph data structure while avoiding to affect user models

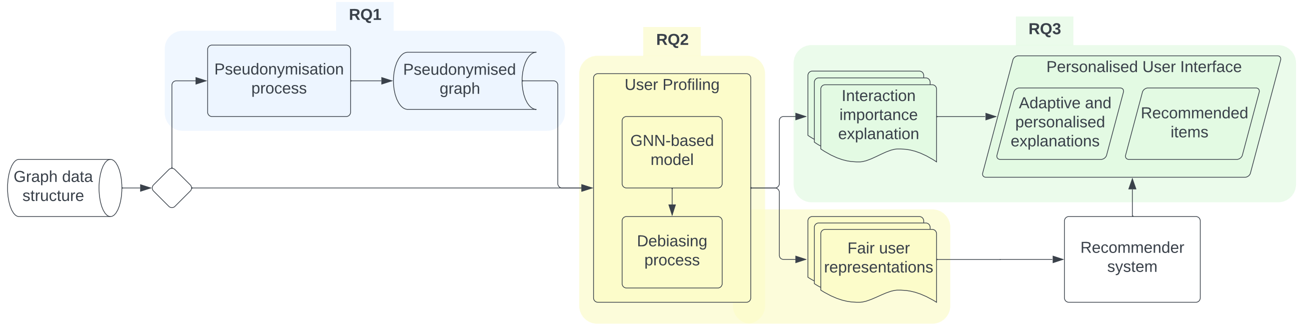

Figure 1: Overall architecture of the framework. Highlighted areas represent the modules developed to address each corresponding RQ.

construction and retaining the performance of the recommender system built upon them?

- **RQ2** How do we build fair user representations from GNN-based user profiling models to keep the input of the downstream recommender debiased?
- **RQ3** How can we personalise user interfaces to adapt the explanations to the needs, demands and requirements of different end-user profiles, considering their distinct knowledge, background and expertise?

To tackle the *privacy* perspectives and thus answering **RQ1**, it is needed to consider the aspects that make it challenging. In particular, user de-identification can be either full or partial, and it can also change dynamically over time. This is due to the fact that users decide whether to provide their explicit consent to the use of personal data, and modify the taken decision at any time. Moreover, as visible in Fig.1, the privacy process is not due for every input dataset. I thus introduce a strategy for *pseudonymisation* with the goal to dynamically transform entities and attributes of the original graph data structure, so that any person processing the data cannot identify individuals but can work with the data at hand in a sensible and meaningful manner. Pseudonymisation refers to the process of de-identifying a data subject (i.e. an individual) from its personal data by replacing personal identifiers (i.e. informative attributes that can allow the identification, such as name and email address) with the so called *pseudonyms* (also referred to as *cryptonyms* or just *nyms*). The choice of pseudonymisation is supported by legal and technical grounds: it is indeed defined within the EU GDPR (Art.4(5)) and recommended by the U.S. National Institute of Standards and Technology (NIST) as the best practice for protecting personal data. A number of pseudonymisation techniques can be applied to de-identify users and their attributes depending on the specific dataset, application and context we are dealing with. So, the main idea is to develop a module for the pseudonymisation process that is flexible and easily adaptable to different environments and domains. One final observation about this perspective: even if it is arguable that publicly available data (e.g. scientific publications in academic domain) are not classified as personal data, and hence not strictly subject to de-identification, several guidelines for the use of pseudonymisation solutions, such as the one published in 2018 by the German Society for Data Protection and Data Security [24], specify that when pseudonymisation is utilised as a technical protective measure, any possible risk of re-identification of an individual must be removed, by decoupling personal information from other data or properly handling those, for instance by generalising attribute values linked to the user (e.g. research interests).

The research on the *fairness* perspective (**RQ2**) starts with an in-depth study of the literature on GNNs. Several different methods and applications has been reviewed in order to find the best possible architectures to apply within the framework. Generally, Graph Neural Networks are deep learning models that capture the dependence of graphs via message passing between the nodes of graphs [38]. Variants of GNNs differ to each other mainly for two aspects: (1) the type of graph structure on which they are built, and (2) the computational modules employed by the neural model. Given the framework being mostly applied in domains where both graph entities and the interaction between them are of different types (e.g. academic data), I consider the usage of GNNs models working on *heterogeneous graphs*. Concerning computational modules, for the same reasons, I select GNNs including convolution or attention operators, which have been proved to be efficient in aggregating information from neighbors, and skip connection operator, used to gather information from historical representations of nodes and mitigate the over-smoothing problem. The final framework design could also consider to include multiple GNNs which work one at a time to provide the best results.

The specific focus of the fairness part of the thesis is on the notion of *disparate impact*. Also known as *adverse impact*, it refers to a form of indirect and often unintentional discrimination that occurs when practices or systems seem to apparently treat people the same way [14]. It concerns with situations where the model disproportionately discriminates certain groups, even if the model does not explicitly employ the sensitive attribute to make predictions but rather on some proxy attributes [32]. This happens with GNNs, where user models are created by aggregating information from neighbours and the sensitive attribute is not explicitly taken into consideration during classification. Assessing disparate impact is beneficial when a linkage in training data between the target label and the sensitive attribute is unclear [37]. Metrics to take into account when evaluating a model's disparate impact are, among others, *statistical parity* [11], *equal opportunity* [15] and *overall accuracy equality* [6]. Furthermore, in scenarios where it is hard to define the correctness of a prediction related to sensitive attribute values, it is worth to argue that assessing *disparate mistreatment* should be required. This concept is particularly significant in contexts where misclassification costs depend on the group affected by the errors and can be evaluated through *treatment equality* metric [6]. Only few work has been done to date to develop fair GNN models, especially for user profiling task, and the most successful approach seems to be the adoption of an additional GNN estimator for the sensitive attribute [10].

The third research question (**RQ3**) addressed in this work relates to the *explainability* perspective. On one hand, the need is to provide interpretation for the results of the user profiling model. Explaining interaction importance is not (yet) a common desired achievement for general GNN models, but some work has been published in the chemical field [9], exploiting a layer-wise relevance propagation method. An interesting and challenging approach is surely to adapt this concept to different GNNs in different domains, while comparing the outcomes with well-established techniques for producing post-hoc explanations for GNNs [36]. On the other hand, the key is the design of an adaptive and personalised user interface (UI) able to show the right explanation to the right user, given his/her profile characteristics. The notion of "right" or "good" explanation is continuously under study in the human-centered research area: results in this direction show, for instance, that different goals and cognitive capabilities affect the perception of explanation [21] and different users require different explanation details [20], while at the same time different individual characteristics can change the perception of transparency [12] or even lead to preferring to have no explanation at all [33]. Through extensive user studies, the aim is to design and implement a UI providing tailored explanations to the end-users which maximises their expectation and satisfaction.

4 RESULTS AND CONTRIBUTIONS TO DATE

In the initial two years of the doctoral studies, I started the research in all the three areas defined by the research questions previously described. In the following, I will list the results and contributions to date:

- For *privacy* perspective (**RQ1**), I developed a first pseudonymisation approach for privacy-preserving recommendations on academic graph data and evaluated the performance preservation in terms of precision. The paper has been published in Computers journal in 2021 [28].
- For *fairness* perspective (**RQ2**), in close collaboration with Dr. Ludovico Boratto of the University of Cagliari, I conducted an assessment of fairness in state-of-the-art GNN-based models (i.e. CatGCN [8] and RHGN [34]) for behavioural user profiling tasks, in terms of both disparate impact and disparate mistreatment. The work will be submitted at CIKM'22.
- Concerning both *fairness* (**RQ2**) and *explainability* (**RQ3**), as a first attempt to embrace both the two perspectives together, I designed and implemented a system with the scope to show how the use of these techniques can lead to the growth of a domain expert's trust and reliance on an AI system. A novel "*Trust&Reliace Scale*" to evaluate XAI systems has been proposed in the paper, which is at present in the rebuttal phase for the publication in IJHCI journal, after being accepted with minor revision.
- Regarding the study and investigation of adaptive and personalised user interfaces (**RQ3**), I carried out a user study among researchers belonging to various research fields to evaluate how two different explainable UIs, providing explanations for the outcomes of a recommender system for scientific papers, are perceived, in terms of understandability, trust and user satisfaction [26]. This work has been presented at IntRS Workshop 2021.
- Following the path of the work at the previous point, in order to extend to the community the line of research on designing adaptive and personalised explainable user interfaces (**RQ3**), we proposed and organised the APEx-UI Workshop at IUI'22 conference [27], in collaboration with Prof. Cataldo Musto and Prof. Pasquale Lops, both from the University of Bari, which has been well welcomed. Held on March 21, 2022, we had around 30 participants and the pleasure to host keynotes by Prof. Katrien Verbert and Prof. Denis Parra.

5 CURRENT STATUS AND FUTURE WORK

As of time of writing this paper, I almost clearly defined what are the key references of the doctoral study to put in the dissertation, and the participation in the doctoral consortium will help me to finally set the research questions.

The future steps of the work, planned for the next 18 months, are illustrated below:

(1) Extend the assessment of fairness of GNN-based user profiling models with a full characterisation of potential biases produced by the analysed state-of-the-art approaches by increasing case studies and considered domains.
(2) Develop the debiasing and explainability modules of the framework, as well as the automated pipeline connecting all the components displayed in the illustrated architecture (Fig. 1).
(3) Continue the research on adaptive and personalised UI design, leveraging the results of APEx-UI Workshop, by performing new user studies in that direction.
(4) Revise the pseudonymisation approach to be applied to the different domains and case studies analysed for the other perspectives of the doctoral project.

In conclusion, at this point in my doctoral journey, I wish to have a successful academic career, and so I am aiming to pursue the Ph.D. as proficiently as possible. However, my previous work experience and current position at the research center, allow me to consider a possible future in the industry, where I would still aim to work in research.

REFERENCES

[1] Ashraf Abdul, Jo Vermeulen, Danding Wang, Brian Y Lim, and Mohan Kankanhalli. 2018. Trends and trajectories for explainable, accountable and intelligible systems: An hci research agenda. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–18.

[2] Erfan Aghasian, Saurabh Garg, and James Montgomery. 2018. User's Privacy in Recommendation Systems Applying Online Social Network Data, A Survey and Taxonomy. In *Big Data Recommender Systems: Recent Trends and Advances*. The Institution of Engineering and Technology, 1–26.

[3] Sören Auer, Markus Stocker, Lars Vogt, Grischa Fraumann, and Alexandra Garatzogianni. 2021. ORKG: Facilitating the Transfer of Research Results with the Open Research Knowledge Graph. *Research Ideas and Outcomes* 7 (2021), e68513.

[4] 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.

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

[6] Richard Berk, Hoda Heidari, Shahin Jabbari, Michael Kearns, and Aaron Roth. 2021. Fairness in criminal justice risk assessments: The state of the art. *Sociological Methods & Research* 50, 1 (2021), 3–44.

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

[8] Chen, Feng, Wang, He, Song, Ling, and Zhang. 2021. CatGCN: Graph Convolutional Networks with Categorical Node Features. *IEEE Trans. Knowl. Data Eng.* (Dec. 2021), 1–1.

[9] Hyeoncheol Cho, Eok Kyun Lee, and Insung S Choi. 2020. Layer-wise relevance propagation of InteractionNet explains protein–ligand interactions at the atom level. *Scientific reports* 10, 1 (2020), 1–11.

[10] 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.

[11] Michael Feldman, Sorelle A Friedler, John Moeller, Carlos Scheidegger, and Suresh Venkatasubramanian. 2015. Certifying and removing disparate impact. In *proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*. 259–268.

[12] Fatih Gedikli, Dietmar Jannach, and Mouzhi Ge. 2014. How should I explain? A comparison of different explanation types for recommender systems. *International Journal of Human-Computer Studies* 72, 4 (2014), 367–382.

[13] David Gunning. 2017. Explainable artificial intelligence (xai). *Defense Advanced Research Projects Agency (DARPA)* 2 (2017).

[14] Sara Hajian, Francesco Bonchi, and Carlos Castillo. 2016. Algorithmic bias: From discrimination discovery to fairness-aware data mining. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 2125–2126.

[15] Moritz Hardt, Eric Price, and Nati Srebro. 2016. Equality of opportunity in supervised learning. *Advances in neural information processing systems* 29 (2016).

[16] 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.

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

[18] Shyong K Lam, Dan Frankowski, and John Riedl. 2006. Do You Trust Your Recommendations? An Exploration of Security and Privacy Issues in Recommender Systems. In *Internation Conference on Emerging Trends in Information and Communication Security*. Springer Berlin Heidelberg, 14–29.

[19] 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.

[20] Martijn Millecamp, Nyi Nyi Htun, Cristina Conati, and Katrien Verbert. 2019. To explain or not to explain: the effects of personal characteristics when explaining music recommendations. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*. Association for Computing Machinery, Los Angeles, CA, USA, 397–407.

[21] Tim Miller. 2019. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence* 267 (Feb. 2019), 1–38.

[22] Cataldo Musto, Fedelucio Narducci, Pasquale Lops, Marco de Gemmis, and Giovanni Semeraro. 2019. Linked open data-based explanations for transparent recommender systems. *International Journal of Human-Computer Studies* 121 (2019), 93–107.

[23] 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.

[24] Data Protection Focus Group of the Digital Summit. 2018. *Requirements for the use of pseudonymisation solutions in compliance with data protection regulations*. Technical Report. German Society for Data Protection and Data Security, Heinrich-Böll-Ring 10, 53119 Bonn, Germany. 17 pages. A working paper of the Data Protection Focus Group of the Platform Security, Protection and Trust for Society and Business, 2018.

[25] 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.

[26] Erasmo Purificato, Baalakrishnan Aiyer Manikandan, Prasanth Vaidya Karanam, Mahantesh Vishvanath Pattadkal, and Ernesto William De Luca. 2021. Evaluating Explainable Interfaces for a Knowledge Graph-Based Recommender System. In *Proceedings of the 8th Joint Workshop on Interfaces and Human Decision Making for Recommender Systems, co-located with RecSys'21* (Amsterdam, Netherlands). 73–88.

[27] Erasmo Purificato, Cataldo Musto, Pasquale Lops, and Ernesto William De Luca. 2022. First Workshop on Adaptive and Personalized Explainable User Interfaces (APEx-UI 2022). In *27th International Conference on Intelligent User Interfaces* (Helsinki, Finland) *(IUI '22 Companion)*. Association for Computing Machinery, New York, NY, USA, 1–3. https://doi.org/10.1145/3490100.3511168

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

[29] 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.

[30] 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.

[31] 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.

[32] Mingyang Wan, Daochen Zha, Ninghao Liu, and Na Zou. 2021. Modeling Techniques for Machine Learning Fairness: A Survey. *arXiv preprint arXiv:2111.03015* (2021).

[33] Clarice Wang, Kathryn Wang, Andrew Bian, Rashidul Islam, Kamrun Naher Keya, James Foulds, and Shimei Pan. 2022. Do Humans Prefer Debiased AI Algorithms? A Case Study in Career Recommendation. In *27th International Conference on Intelligent User Interfaces* (Helsinki, Finland) *(IUI '22)*. Association for Computing Machinery, New York, NY, USA, 134–147. https://doi.org/10.1145/3490099.3511108

[34] 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*. Association for Computing Machinery, New York, NY, USA, 3573–3577.

[35] 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.

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

[37] Muhammad Bilal Zafar, Isabel Valera, Manuel Gomez Rodriguez, and Krishna P Gummadi. 2017. Fairness beyond disparate treatment & disparate impact: Learning classification without disparate mistreatment. In *Proceedings of the 26th international conference on world wide web*. 1171–1180.

[38] Jie Zhou, Ganqu Cui, Shengding Hu, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, and Maosong Sun. 2020. Graph neural networks: A review of methods and applications. *AI Open* 1 (2020), 57–81.