Internship project

Subject: Studying the long-term implications of fairness constraints on dynamic graphs

Possibility to continue as a PhD candidate: YES

Description of the internship

Supervision

Charlotte Laclau (https://laclauc.github.io/)
Jefrey Lijffijt (https://users.ugent.be/ jlijffij/)

Location and dates of the internship

Address: Télécom Paris, 19 Place Marguerite Perey, 91120 Palaiseau
Date of the beginning of the internship: beginning 2023

Team where the thesis will be written Department IDS, Team Signal, Statistique et Apprentissage (S2A)

Keywords machine learning, representation learning, dynamic graphs, fairness constraints

Detailed subject Addressing fairness for graph data has only recently started to draw attention from the scientific community. The challenge behind dealing with this particular data type is due to the fact that existing fairness enhancing approaches originally developed for standard tabular data may not and usually do not work on relational graph data. In particular, graphs are non-iid (independent and identically distributed) and non-euclidean by nature. The first assumption implies that changing or altering information about a given node (attributes or connections) in the graph will impact its neighbors. As a result, evaluating and mitigating potential bias requires proper handling of this important graph property. The second assumption implies that before learning any model solving a particular task such as, classification, ranking or clustering, one should first learn a representation of the graph in the form of a vector. At the node level, this corresponds to node embedding models. Depending on the objective function used in these models, the learned node representation can capture different amount of bias, or make it more difficult to mitigate the bias afterward. Contributions in the domain of fair graph mining differ according to where the bias mitigation happens in the machine learning pipeline: directly on the edge distribution of the graph [6, 5, 13], during the representation learning step [10, 2], notably with GNNs-based models [9], or through the use of adversaries [1, 8, 3]. However, in practical applications, such as social networks, a vast majority of graphs are dynamic meaning that their graph topology (links between individuals) evolves over time. To capture the evolution in this setting, graphs are most commonly represented as a collection of static graphs at different timestamps, and for each graph a new representation is systematically learned to solve the downstream task. While the dynamic graphs were studied previously, there exist no previous work or study on fair graph mining that considers the impact of these temporal dynamics on bias. This is despite the fact that the shifting distribution of population (e.g. the protected group may change over time), or the altering of decisions being made following users feedback may lead to counter-intuitive phenomena (like the Simpson's paradox) leading to bias being amplified. As a result, imposing static fairness criteria at every timestep may actually exacerbate unfairness [4, 7, 12].

The objective of this internship is to explore, in a simulated environment, the impact of fairness constraints on time-evolving graphs. Starting from a generative process for continuous temporal graphs proposed in [11], our first objective is to define a set of dynamic use case scenarios and to implement them. This should allow us to perform the first study on the short term and long term implications of fair interventions on graph machine learning models. In a second step, we are interested in designing an appropriate evaluation framework to compare the graph structure (e.g. with an appropriate graph distance) resulting from the *unbiased* vs *biased* generative process. Overall, we are particularly interested in addressing the three following research questions: (1) Are fairness constraints actually improving fairness on the long term? (2) Do all fairness constraints have a similar effect on the evolution of the graph? (3) Are all nodes impacted in a similar manner by imposing fairness constraints and their potential side effects? For this latter, we are particularly attentive to the properties of the nodes (eg. degree) for which this impact might be significant.

Candidate profile Student having master 2 research

- Statistical learning, bases of probability
- Good level of programming (Python)
- Good command of English
- Strong interest in AI ethics

Application To send on charlotte.laclau@telecom-paris.fr:

- Curriculum Vitae
- Personalized motivation letter that explains interest of the candidate in the subject (can be directly in the body of the email)
- Grade reports for recent years
- Contact of a person willing to give recommendation

Incomplete applications will not be considered.

References

[1] A. Bose and W. Hamilton. Compositional fairness constraints for graph embeddings. In *ICML*, pages 715–724, 2019.

[2] M. Buyl and T. De Bie. The kl-divergence between a graph model and its fair i-projection as a fairness regularizer. In *ECML-PKDD*, 2021.

[3] Enyan Dai and Suhang Wang. Learning fair graph neural networks with limited and private sensitive attribute information. *IEEE Transactions on Knowledge and Data Engineering*, pages 1–14, 2022.

[4] Alexander D'Amour, Hansa Srinivasan, James Atwood, Pallavi Baljekar, D. Sculley, and Yoni Halpern. Fairness is not static: Deeper understanding of long term fairness via simulation studies. In *Proceedings of FAccT*, 2020.

[5] Charlotte Laclau, Ievgen Redko, Manvi Choudhary, and Christine Largeron. All of the fairness for edge prediction with optimal transport. Proceedings of AISTATS, pages 1774–1782, 2021.

[6] Peizhao Li, Yifei Wang, Han Zhao, Pengyu Hong, and Hongfu Liu. On dyadic fairness: Exploring and mitigating bias in graph connections. In *ICLR*, 2021.

[7] Lydia T. Liu, Sarah Dean, Esther Rolf, Max Simchowitz, and Moritz Hardt. Delayed impact of fair machine learning. In *Proceedings of IJCAI*, pages 6196–6200, 7 2019.

[8] F. Masrour, T. Wilson, H. Yan, P.N. Tan, and A. Esfahanian. Bursting the filter bubble: Fairness-aware network link prediction. *AAAI*, 34(01):841–848, 2020.

[9] J. Palowitch and B. Perozzi. Monet: Debiasing graph embeddings via the metadata-orthogonal training unit. In *ASONAM*, 2020.

[10] T. A Rahman, B. Surma, M. Backes, and Y. Zhang. Fairwalk: Towards fair graph embedding. In *IJCAI*, pages 3289–3295, 2019.

[11] Raphaël Romero, Bo Kang, and Tijl De Bie. Graph-survival: A survival analysis framework for machine learning on temporal networks, 2022.

[12] Andrea Simonetto and Ivano Notarnicola. Achievement and fragility of long-term equitability. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, page 675–685, New York, NY, USA, 2022. Association for Computing Machinery.

[13] N. Wang, L Lin, J. Li, and H. Wang. Unbiased graph embedding with biased graph observations. In *WWW*, page 1423–1433, 2022.