Responsible AI in Graph Machine Learning and More

Yushun Dong
Department of Computer Science
Florida State University
yd24f@fsu.edu

In today's rapidly evolving digital landscape, artificial intelligence (AI) has become deeply integrated into our daily lives, from social media recommendations to healthcare diagnostics. However, this increasing influence raises critical questions about AI's reliability, fairness, and societal impact. The Responsible AI (RAI) Lab at Florida State University, directed by Dr. Yushun Dong, is dedicated to addressing these fundamental challenges through cutting-edge research across four interconnected areas.

AI Explainability. Imagine using AI to make crucial medical decisions - wouldn't you want to understand why the AI made its recommendations? This question drives our research in AI explainability. We develop innovative methods to make AI systems more transparent and interpretable, particularly focusing on graph neural networks (GNNs), which are powerful tools for analyzing interconnected data. Our groundbreaking work includes techniques for explaining GNN decisions through training node attribution [12], information bottleneck approaches [30], and leveraging large language models for molecular applications [20]. We've also developed advanced frameworks like GIGAMAE [28] and SEESAW [11] to better understand AI models in structural and empirical ways, making AI systems more transparent.

AI Fairness. Just as we expect fairness in human decision-making, AI systems must treat all individuals and groups equitably. Our lab leads groundbreaking research in AI fairness, particularly in graph-based systems [9]. We've developed comprehensive approaches to ensure both individual fairness [5] and group fairness [29], while creating innovative solutions for debiasing AI systems [8, 39]. Our work extends to fair knowledge distillation [15] and rebalancing techniques [24]. To make these advances accessible to the broader community, we've developed PyGDebias [6], a practical toolkit for implementing fair AI systems, and published influential surveys [9] that guide researchers and practitioners.

AI Security. As AI systems become more prevalent in critical applications, ensuring their security is paramount. Our lab investigates both attack mechanisms and defense strategies [34]. We've pioneered research in adversarial attacks on graph fairness [41], developed novel spectral attacks [43], and created certified defense mechanisms [14]. Our innovative work includes contrastive learning for anomaly detection [40] and certified unlearning in neural networks [42, 13]. We've also made significant contributions to federated learning security [18, 17] and adaptive network filtering [4].

AI/ML Applications. Our theoretical advances translate directly into real-world solutions across various domains. In healthcare, we've developed cutting-edge systems for antibiogram pattern prediction [16] and analyzed COVID-19 policy impacts [26]. Our transportation research includes intelligent route planning [38] and advanced pavement performance forecasting [10]. We've also reviewed recommendation systems [7, 21] and made significant strides in language models [37, 25, 23, 36]. Our work extends to environmental studies [2], brain network analysis [35], time series forecasting [27, 22], and outlier detection [1]. We've also developed sophisticated techniques for few-shot learning [31], hierarchical task learning [33, 32], and hierarchical demonstration optimization [19] to foster practical AI.

The RAI Lab stands at the forefront of responsible AI research, as evidenced by Dr. Dong's comprehensive doctoral work [3] and numerous publications in top-tier venues. We welcome students and researchers passionate about developing AI systems that are not only powerful but also fair, explainable, and secure. Our lab provides a collaborative environment where theoretical innovation meets practical impact, supported by state-of-the-art resources and mentorship. Through the above-mentioned works, we're shaping the future of responsible AI development.

Join us in our mission to make AI systems more trustworthy, fair, and beneficial for society. Whether you're interested in theoretical foundations or practical applications, the RAI Lab offers exciting opportunities to contribute to cutting-edge research that matters. We are always open to research interns, whether you're an undergraduate student, graduate student, or early-career researcher. We welcome everyone to reach out and explore potential collaboration opportunities, regardless of your background or experience level. Feel free to contact us to discuss how you can contribute to and grow with our cutting-edge research in responsible AI.

References

[1] Sihan Chen, Zhuangzhuang Qian, Wingchun Siu, Xingcan Hu, Jiaqi Li, Shawn Li, Yuehan Qin, Tiankai Yang, Zhuo Xiao, Wanghao Ye, et al. Pyod 2: A python library for outlier detection with llm-powered model selection. *arXiv preprint arXiv:2412.12154*, 2024.

[2] Rong Ding, Yushun Dong, Daniel P Aldrich, Jundong Li, Kelsey Pieper, and Qi Ryan Wang. Post-disaster private well water contamination with geosocial network: A case study of post-hurricane harvey. In *Computing in Civil Engineering 2023*, pages 194–201.

[3] Yushun Dong. *Algorithmic Fairness in Graph Machine Learning: Explanation, Optimization, and Certification*. PhD thesis, University of Virginia, 2024.

[4] Yushun Dong, Kaize Ding, Brian Jalaian, Shuiwang Ji, and Jundong Li. Adagnn: Graph neural networks with adaptive frequency response filter. In *Proceedings of the 30th ACM international conference on information & knowledge management*, pages 392–401, 2021.

[5] Yushun Dong, Jian Kang, Hanghang Tong, and Jundong Li. Individual fairness for graph neural networks: A ranking based approach. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pages 300–310, 2021.

[6] Yushun Dong, Zhenyu Lei, Zaiyi Zheng, Song Wang, Jing Ma, Alex Jing Huang, Chen Chen, and Jundong Li. Pygdebias: A python library for debiasing in graph learning. In *Companion Proceedings of the ACM on Web Conference 2024*, pages 1019–1022, 2024.

[7] Yushun Dong, Jundong Li, and Tobias Schnabel. When newer is not better: Does deep learning really benefit recommendation from implicit feedback? In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 942–952, 2023.

[8] Yushun Dong, Ninghao Liu, Brian Jalaian, and Jundong Li. Edits: Modeling and mitigating data bias for graph neural networks. In *Proceedings of the ACM web conference 2022*, pages 1259–1269, 2022.

[9] Yushun Dong, Jing Ma, Song Wang, Chen Chen, and Jundong Li. Fairness in graph mining: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 35(10):10583–10602, 2023.

[10] Yushun Dong, Yingxia Shao, Xiaotong Li, Sili Li, Lei Quan, Wei Zhang, and Junping Du. Forecasting pavement performance with a feature fusion lstm-bpnn model. In *Proceedings of the 28th ACM international conference on information and knowledge management*, pages 1953–1962, 2019.

[11] Yushun Dong, William Shiao, Yozen Liu, Jundong Li, Neil Shah, and Tong Zhao. Seesaw: Do graph neural networks improve node representation learning for all?

[12] Yushun Dong, Song Wang, Jing Ma, Ninghao Liu, and Jundong Li. Interpreting unfairness in graph neural networks via training node attribution. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 7441–7449, 2023.

[13] Yushun Dong, Binchi Zhang, Zhenyu Lei, Na Zou, and Jundong Li. Idea: A flexible framework of certified unlearning for graph neural networks. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 621–630, 2024.

[14] Yushun Dong, Binchi Zhang, Hanghang Tong, and Jundong Li. Elegant: Certified defense on the fairness of graph neural networks. *arXiv preprint arXiv:2311.02757*, 2023.

[15] Yushun Dong, Binchi Zhang, Yiling Yuan, Na Zou, Qi Wang, and Jundong Li. Reliant: Fair knowledge distillation for graph neural networks. In *Proceedings of the 2023 SIAM International Conference on Data Mining (SDM)*, pages 154–162. Society for Industrial and Applied Mathematics, 2023.

[16] Xingbo Fu, Chen Chen, Yushun Dong, Anil Vullikanti, Eili Klein, Gregory Madden, and Jundong Li. Spatial-temporal networks for antibiogram pattern prediction. In *2023 IEEE 11th International Conference on Healthcare Informatics (ICHI)*, pages 225–234. IEEE, 2023.

[17] Xingbo Fu, Song Wang, Yushun Dong, Binchi Zhang, Chen Chen, and Jundong Li. Federated graph learning with graphless clients. *arXiv preprint arXiv:2411.08374*, 2024.

[18] Xingbo Fu, Binchi Zhang, Yushun Dong, Chen Chen, and Jundong Li. Federated graph machine learning: A survey of concepts, techniques, and applications. *ACM SIGKDD Explorations Newsletter*, 24(2):32–47, 2022.

[19] Yinhan He, Wendy Zheng, Song Wang, Zaiyi Zheng, Yushun Dong, Yaochen Zhu, and Jundong Li. Hierarchical demonstration order optimization for many-shot in-context learning.

[20] Yinhan He, Zaiyi Zheng, Patrick Soga, Yaochen Zhu, Yushun Dong, and Jundong Li. Explaining graph neural networks with large language models: A counterfactual perspective on molecule graphs. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 7079–7096, 2024.

[21] Zheng Huang, Jing Ma, Yushun Dong, Natasha Zhang Foutz, and Jundong Li. Empowering next poi recommendation with multi-relational modeling. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2034–2038, 2022.

[22] Zhenyu Lei, Yushun Dong, Jundong Li, and Chen Chen. St-fit: Inductive spatial-temporal forecasting with limited training data. *arXiv preprint arXiv:2412.10912*, 2024.

[23] Lincan Li, Jiaqi Li, Catherine Chen, Fred Gui, Hongjia Yang, Chenxiao Yu, Zhengguang Wang, Jianing Cai, Junlong Aaron Zhou, Bolin Shen, et al. Political-llm: Large language models in political science. *arXiv preprint arXiv:2412.06864*, 2024.

[24] Zhixun Li, Yushun Dong, Qiang Liu, and Jeffrey Xu Yu. Rethinking fair graph neural networks from re-balancing. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1736–1745, 2024.

[25] Haochen Liu, Song Wang, Yaochen Zhu, Yushun Dong, and Jundong Li. Knowledge graph-enhanced large language models via path selection. *arXiv preprint arXiv:2406.13862*, 2024.

[26] Jing Ma, Yushun Dong, Zheng Huang, Daniel Mietchen, and Jundong Li. Assessing the causal impact of covid-19 related policies on outbreak dynamics: A case study in the us. In *Proceedings of the ACM Web Conference 2022*, pages 2678–2686, 2022.

[27] Xihao Piao, Zheng Chen, Yushun Dong, Yasuko Matsubara, and Yasushi Sakurai. Frednormer: Frequency domain normalization for non-stationary time series forecasting. *arXiv preprint arXiv:2410.01860*, 2024.

[28] Yucheng Shi, Yushun Dong, Qiaoyu Tan, Jundong Li, and Ninghao Liu. Gigamae: Generalizable graph masked autoencoder via collaborative latent space reconstruction. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, pages 2259–2269, 2023.

[29] Weihao Song, Yushun Dong, Ninghao Liu, and Jundong Li. Guide: Group equality informed individual fairness in graph neural networks. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1625–1634, 2022.

[30] Jihong Wang, Minnan Luo, Jundong Li, Yun Lin, Yushun Dong, Jin Song Dong, and Qinghua Zheng. Empower post-hoc graph explanations with information bottleneck: A pre-training and fine-tuning perspective. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 2349–2360, 2023.

[31] Song Wang, Yushun Dong, Kaize Ding, Chen Chen, and Jundong Li. Few-shot node classification with extremely weak supervision. In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*, pages 276–284, 2023.

[32] Song Wang, Yushun Dong, Xiao Huang, Chen Chen, and Jundong Li. Faith: Few-shot graph classification with hierarchical task graphs. *arXiv preprint arXiv:2205.02435*, 2022.

[33] Song Wang, Yushun Dong, Xiao Huang, Chen Chen, and Jundong Li. Learning hierarchical task structures for few-shot graph classification. *ACM Transactions on Knowledge Discovery from Data*, 18(3):1–20, 2024.

[34] Song Wang, Yushun Dong, Binchi Zhang, Zihan Chen, Xingbo Fu, Yinhan He, Cong Shen, Chuxu Zhang, Nitesh V Chawla, and Jundong Li. Safety in graph machine learning: Threats and safeguards. *arXiv preprint arXiv:2405.11034*, 2024.

[35] Song Wang, Zhenyu Lei, Zhen Tan, Jiaqi Ding, Xinyu Zhao, Yushun Dong, Guorong Wu, Tianlong Chen, Chen Chen, Aiying Zhang, et al. Brainmap: Learning multiple activation pathways in brain networks. *arXiv preprint arXiv:2412.17404*, 2024.

[36] Song Wang, Peng Wang, Yushun Dong, Tong Zhou, Lu Cheng, Yangfeng Ji, and Jundong Li. On demonstration selection for improving fairness in language models. In *Workshop on Socially Responsible Language Modelling Research*.

[37] Song Wang, Peng Wang, Tong Zhou, Yushun Dong, Zhen Tan, and Jundong Li. Ceb: Compositional evaluation benchmark for fairness in large language models. *arXiv preprint arXiv:2407.02408*, 2024.

[38] Xiao Wang, Quan Yuan, Zhihan Liu, Yushun Dong, Xiaojuan Wei, and Jinglin Li. Learning route planning from experienced drivers using generalized value iteration network. In *Internet of Vehicles. Technologies and Services Toward Smart Cities: 6th International Conference, IOV 2019, Kaohsiung, Taiwan, November 18–21, 2019, Proceedings 6*, pages 88–100. Springer International Publishing, 2020.

[39] Yu Wang, Yuying Zhao, Yushun Dong, Huiyuan Chen, Jundong Li, and Tyler Derr. Improving fairness in graph neural networks via mitigating sensitive attribute leakage. In *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining*, pages 1938–1948, 2022.

[40] Zhiming Xu, Xiao Huang, Yue Zhao, Yushun Dong, and Jundong Li. Contrastive attributed network anomaly detection with data augmentation. In *Pacific-Asia conference on knowledge discovery and data mining*, pages 444–457. Springer International Publishing Cham, 2022.

[41] Binchi Zhang, Yushun Dong, Chen Chen, Yada Zhu, Minnan Luo, and Jundong Li. Adversarial attacks on fairness of graph neural networks. *arXiv preprint arXiv:2310.13822*, 2023.

[42] Binchi Zhang, Yushun Dong, Tianhao Wang, and Jundong Li. Towards certified unlearning for deep neural networks. *arXiv preprint arXiv:2408.00920*, 2024.

[43] Xianren Zhang, Jing Ma, Yushun Dong, Chen Chen, Min Gao, and Jundong Li. Sd-attack: Targeted spectral attacks on graphs. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 352–363. Springer Nature Singapore Singapore, 2024.