JUNJIE XU

+1(814) 996-8596 ◇ University Park, PA
◇ junjiexu@psu.edu ◇ Homepage ◇ Google Scholar ◇ LinkedIn

SUMMARY

- **Ph.D.** candidate in Machine Learning with 4+ years of R&D experience in the field.
- **Research Interests**: Machine Learning, AI for Science, GNN, LLM, Geometric Deep Learning.
- **16 peer-reviewed papers** published or in submission, with **600+ citations** as of 05/2025.

EDUCATION

The Pennsylvania State University, University Park, USA 08/2021 - Present
Ph.D. candidate in Informatics
Advisor: Dr. Suhang Wang & Dr. Xiang Zhang

Huazhong University of Science and Technology, Wuhan, China 09/2017 - 06/2021
B.E. in Software Engineering
GPA: 3.91/4.00

University of California, Berkeley, Berkeley, USA 01/2020 - 06/2020
Exchange student in Computer Science

PEER-REVIEWED PAPERS

Preprints

[1] **LLM and GNN are Complementary: Distilling LLM for Multimodal Graph Learning**
Junjie Xu, Zongyu Wu, Minhua Lin, Xiang Zhang, Suhang Wang

[2] **LanP: Rethinking the Impact of Language Priors in Large Vision-Language Models**
Zongyu Wu, Yuwei Niu, Hongcheng Gao, Minhua Lin, Zhiwei Zhang, Zhifang Zhang, Qi Shi, Yilong Wang, Sike Fu, **Junjie Xu**, Junjie Ao, Enyan Dai, Lei Feng, Xiang Zhang, Suhang Wang

[3] **A Comprehensive Survey of Small Language Models in the Era of Large Language Models: Techniques, Enhancements, Applications, Collaboration with LLMs, and Trustworthiness**
Fali Wang, Zhiwei Zhang, Xianren Zhang, Zongyu Wu, Tzuhao Mo, Qiuhao Lu, Wanjing Wang, Rui Li, **Junjie Xu**, Xianfeng Tang, Qi He, Yao Ma, Ming Huang, Suhang Wang

[4] **Self-Explainable Graph Neural Networks for Link Prediction**
Huaisheng Zhu, Dongsheng Luo, Xianfeng Tang, **Junjie Xu**, Hui Liu, Suhang Wang

Peer-Reviewed Papers

[5] **Geometric Hyena Networks for Large-scale Equivariant Learning**
Artem Moskalev, Mangal Prakash, **Junjie Xu**, Tianyu Cui, Rui Liao, Tommaso Mansi
ICML 2025 (Spotlight), The International Conference on Machine Learning

[6] **HARMONY: A Multi-Representation Framework for RNA Property Prediction**
Junjie Xu, Artem Moskalev, Tommaso Mansi, Mangal Prakash, Rui Liao
AI4NA @ ICLR 2025 (Oral), ICLR 2025 Workshop on AI for Nucleic Acids

[7] **Beyond Sequence: Impact of Geometric Context for RNA Property Prediction**
Junjie Xu, Artem Moskalev, Tommaso Mansi, Mangal Prakash, Rui Liao
ICLR 2025, The International Conference on Learning Representations
also appears in AIDrugX @ NeurIPS 2024, NeurIPS 2024 Workshop on AI for New Drug Modalities

[8] **Robustness-Inspired Defense Against Backdoor Attacks on Graph Neural Networks**
Zhiwei Zhang, Minhua Lin, **Junjie Xu**, Zongyu Wu, Enyan Dai, Suhang Wang
ICLR 2025 (Oral), The International Conference on Learning Representations

[9] **Stealing Training Graphs from Graph Neural Networks**
Minhua Lin, Enyan Dai, **Junjie Xu**, Jinyuan Jia, Xiang Zhang, Suhang Wang
KDD 2025, 31th SIGKDD Conference on Knowledge Discovery and Data Mining

[10] **Shape-aware Graph Spectral Learning**
Junjie Xu, Enyan Dai, Dongsheng Luo, Xiang Zhang, Suhang Wang
CIKM 2024, ACM International Conference on Information and Knowledge Management

[11] **Enhancing GNNs with Limited Labeled Data by Actively Distilling Knowledge from LLMs**
Quan Li, Tianxiang Zhao, Lingwei Chen, **Junjie Xu**, and Suhang Wang
BigData 2024, 2024 IEEE International Conference on Big Data

[12] **HC-GST: Heterophily-aware Distribution Consistency based Graph Self-training**
Fali Wang, Tianxiang Zhao, **Junjie Xu**, Suhang Wang
CIKM 2024, ACM International Conference on Information and Knowledge Management

[13] **HP-GMN: Graph Memory Networks for Heterophilous Graphs**
Junjie Xu, Enyan Dai, Xiang Zhang, Suhang Wang
ICDM 2022, IEEE International Conference on Data Mining

[14] **A Comprehensive Survey on Trustworthy Graph Neural Networks: Privacy, Robustness, Fairness, and Explainability**
Enyan Dai, Tianxiang Zhao, Huaisheng Zhu, **Junjie Xu**, Zhimeng Guo, Hui Liu, Jiliang Tang, Suhang Wang
Machine Intelligence Research

[15] **Revisiting Time Series Outlier Detection: Definitions and Benchmarks**
Kwei-Herng Lai, Daochen Zha, **Junjie Xu**, Yue Zhao, Guanchu Wang, Xia Hu
NIPS 2021, Neural Information Processing Systems, Datasets and Benchmarks Track

[16] **TODS: An Automated Time Series Outlier Detection System**
Kwei-Herng Lai, Daochen Zha, Guanchu Wang, **Junjie Xu**, Yue Zhao, Devesh Kumar, Yile Chen, Purav Zumkhawaka, Mingyang Wan, Diego Martinez, Xia Hu
AAAI 2021, AAAI Conference on Artificial Intelligence, Demo track

EXPERIENCE

Johnson & Johnson 05/2024 - 11/2024
Research Intern; *Advisor: Mangal Prakash* New Brunswick, NJ, US

Geometric Deep Learning for RNA Prediction

- Generated and refined RNA datasets encompassing **1D, 2D, and 3D** structures; constructed graphs and geometric graphs based on 2D and 3D structures.
- Conducted extensive benchmarking of state-of-the-art models across 1D, 2D, and 3D methods, evaluating model **scalability, robustness to noise, and generalization** under real-world challenges.
- Developed innovative **3D geometric RNA modeling** methods utilizing **hierarchical multidimensional GNNs**, integrating RNA's geometric complexity with multi-scale structural representations, achieving state-of-the-art overall performance across all real-world settings.

Penn State University 08/2021 - Present
Research Assistant; *Advisors: Suhang Wang & Xiang Zhang* University Park, PA, US

Multi-modal Knowledge Distillation with LLM for Molecule Property Prediction

- Explored the impact of **multi-modal** inputs, including textual descriptions, diagrams, node features, and graph structures, on molecule property prediction, and assessed the LLM's capability to capture multi-modal information.
- **Fine-tuned** language models to enhance the encoding of LLM outputs, and **augmented the input space** for superior representation learning performance.

- Implemented **knowledge distillation** from Large Language Models by aligning their outputs with smaller models, like GNNs and MLPs, boosting **efficiency** and reducing **costs** during reasoning.

Heterophily-aware GNNs and Trustworthy GNNs

- Developed Graph Memory networks to tackle heterophily in graphs, enhancing prediction with local and global pattern learning, and introduced regularization for improved global information capture, achieving top performance.
- Investigated the impact of filter frequency in **Spectral GNNs**, using theoretical and empirical analyses.
- Introduced shape-aware regularization and a Newton Interpolation-based spectral filter in GNN, customizing filters to match homophily levels and enhancing performance on various datasets.
- Surveyed trustworthy graph neural networks to improve **privacy, robustness, fairness, and explainability**.

Rice University 05/2020 - 06/2021
Research Assistant; *Advisor: Xia "Ben" Hu* TX, US

Automated Time-series Outlier Detection System

- Developed a **full stack and automated** machine learning system with preprocessing, feature extraction, detection algorithms, and human-in-the-loop interfaces.
- Integrated a wide range of algorithms including PyOD. Revisited the definition of the time-series anomalies and proposed a taxonomy for point-wise, piece-wise, and pattern-wise anomalies.
- Implemented Automated Machine Learning **(AutoML)** for knowledge-free pipeline construction and automatic optimization of module combinations. Developed graphical user interfaces **(GUI)** to improve usability.
- [Code] (Github 1.5k+ stars, 200 forks); [Website]; [Video].

SERVICE

Reviewer
NeurIPS (2024, 2023, 2022); ICLR (2025, 2024); ICML (2025, 2024); AAAI (2025); ICDM (2024, 2023, 2022); LoG (2023); KDD (2023, 2022); WSDM (2023); CIKM (2023); WWW (2022); TKDD; PAKDD; ICWSM

TEACHING EXPERIENCE

Teaching Assistant, PSU

DS 310: Machine Learning for Data Analytics	Spring 2025
IST 597: Machine Learning on Graphs	Spring 2024
HCDD 364W: Methods for Studying Users	Spring 2024
DS 305: Algorithmic Methods & Tools	Fall 2023

HONORS & AWARDS

IST Travel Award	IST, PSU, 2022
ICDM Student Travel Award	ICDM, 2022
Graham Endowed Fellowship	PSU, 2021
Mitacs Globalink Research Scholarship	China Scholarship Council, 2020
Scholarship for Academic Excellence	HUST, 2018-2019
Scholarship for Academic Excellence	HUST, 2017-2018

SKILLS

Languages	English (Fluent), Mandarin (Native)
Programming	Python, Java, Matlab, C
Deep Learning	PyTorch, PyTorch Geometric, DGL, Tensorflow, PyTorch Lightning, e3nn