Kaiwen Dong

PhD. Candidate
Department of Computer Science and Engineering
University of Notre Dame

Phone: +1(217)607-9853
Email: kdong2@nd.edu
Homepage: https://barcavin.github.io/

Education

University of Notre Dame, Notre Dame, IN, US — Aug 2021 – **June 2025 (Expected)**
Ph.D. Candidate in Computer Science and Engineering
Advisor: Prof. Nitesh V. Chawla (ACM & IEEE & AAAI & AAAS Fellow)

University of Illinois at Urbana-Champaign, Champaign, IL, US — Aug 2016 – May 2018
Master of Science in Statistics

Sichuan University, Chengdu, China — Aug 2012 – July 2016
Bachelor of Science in Mathematics

Research Interests and Applications

Graph Learning: enable machines to effectively and efficiently capture complex relationships within real-world data.
Deep Relational Learning: advance representation learning of tabular data, especially for relational databases.
AI applications: develop and deploy AI applications, including enterprise text-to-SQL, transaction categorization, and customer service monitoring.

Professional Experience

Research Scientist Intern — April 2024 – July 2024
Intuit Inc.
· Developed a transaction categorization model ([P6]) using relational deep learning for financial data stored in relational databases. Integrated language modeling and heterogeneous graph techniques to model complex relationships within the database. Designed a unified model that performs transaction categorization directly on relational data, eliminating the need for manual feature engineering.
Mentors: Dr. Kamalika Das and Dr. Xiang Gao

Data Scientist — June 2018 – July 2021
Aunalytics Inc.
· Led the development of the text-to-SQL system ([S2]) at enterprise scale, which translates natural language to a machine-executable SQL query and provides a means for non-technical people to interact with relational databases. Built the entire pipeline of the production-ready model from the data simulation, model training/experiments to the scoring service, CI/CD, and quality assurance.
· Developed a Sentiment Index project based on customer service data to find dissatisfied customers in advance. Built model interpretability tools around LIME and Shapley value to empower the model to provide not only predictions but also actionable suggestions. Improved customer relationships for over 50 clients daily.
· Improved the existing Churn project to help the stakeholder easily target customers who are likely to unsubscribe. Achieved 40% accuracy on actionable customer lists and saved over 500 customers each month.
Supervisor: Dr. David Cieslak

Graduate Research Assistant — August 2022 – Present
University of Notre Dame
· Worked on robustness ([S3,P5]) and efficiency ([S4,P3]) of graph neural networks for link prediction tasks. Developed novel techniques to enhance GNN performance, achieving SOTA performance ([S4]) on large-scale graph datasets from Open Graph Benchmark.
Advisor: Prof. Nitesh V. Chawla

Graduate Teaching Assistant — August 2021 – December 2021
University of Notre Dame
· CSE 21312 Data Structures: Review assignments, Q&A during office hours, Grade.
Professor: Matthew Morrison

Publications [Google Scholar]

Selected Peer-Reviewed Publications

[S4] **Pure Message Passing Can Estimate Common Neighbor for Link Prediction**
Kaiwen Dong, Zhichun Guo, Nitesh V. Chawla
The Conference on Neural Information Processing Systems, 2024 ***(Neurips'24)***

[S3] **FakeEdge: Alleviate Dataset Shift in Link Prediction**
Kaiwen Dong, Yijun Tian, Zhichun Guo, Yang Yang, Nitesh V. Chawla
The Learning on Graphs Conference, 2022 ***(LoG'22)***

[S2] **An Optimized NL2SQL System for Enterprise Data Mart**
Kaiwen Dong, Kai Lu, Xin Xia, David Cieslak, Nitesh V. Chawla
The European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, 2021 ***(ECML PKDD'21)***

[S1] **Heterogeneous Graph Masked Autoencoders**
Yijun Tian, **Kaiwen Dong**, Chunhui Zhang, Chuxu Zhang, Nitesh V. Chawla
The AAAI Conference on Artificial Intelligence, 2023 ***(AAAI'23)***

Preprints and Under Submissions

[P6] **Transaction Categorization in QuickBooks with Relational Deep Learning**
Kaiwen Dong, Xiang Gao, Ayan Acharya, Maria Kissa, Hilaf Hasson, Mauricio Flores, Nitesh V. Chawla, Kamalika Das
Under Submission, 2025 ***(KDD'25)***

[P5] **CORE: Data Augmentation for Link Prediction via Information Bottleneck**
Kaiwen Dong, Zhichun Guo, Nitesh V. Chawla
Under Submission ***(TKDD)***

[P4] **You do not have to train Graph Neural Networks at all on text-attributed graphs**
Kaiwen Dong, Zhichun Guo, Nitesh V. Chawla
Under Submission, 2025 ***(TMLR)***

[P3] **Universal Link Predictor By In-Context Learning on Graphs**
Kaiwen Dong, Haitao Mao, Zhichun Guo, Nitesh V. Chawla
Under Submission, 2025 ***(TMLR)***

[P2] **Node Duplication Improves Cold-start Link Prediction**
Zhichun Guo, Tong Zhao, Yozen Liu, **Kaiwen Dong**, William Shiao, Mingxuan Ju, Neil Shah, Nitesh V. Chawla
Under Submission, 2025 ***(ICLR'25)***

[P1] **MolX: Enhancing Large Language Models for Molecular Learning with A Multi-Modal Extension**
Le Huy Khiem, Zhichun Guo, **Kaiwen Dong**, Xiaobao Huang, Bozhao Nan, Roshni Iyer, Xiangliang Zhang, Olaf Wiest, Wei Wang, Nitesh V. Chawla
Under Submission, 2025 ***(AAAI'25)***

Professional Services

Conference Reviewer

International Conference on Learning Representations (**ICLR**)	2025
International Conference on Machine Learning (**ICML**)	2025
Neural Information Processing Systems (**NeurIPS**)	2024
SIGKDD Conference on Knowledge Discovery and Data Mining (**KDD**)	2025, 2024
Conference on Information and Knowledge Management (**CIKM**)	2024
International Conference on Artificial Intelligence and Statistics (**AISTATS**)	2025

Journal Invited Reviewer

ACM Transactions on Knowledge Discovery from Data (**TKDD**)	2023