

Bowen Jin

Room 1117, Siebel Center for Computer Science, 201 N. Goodwin Ave, Urbana, IL 61801

Email: bowenj4@illinois.edu | Homepage: <https://peterjin.me> | Phone: 217-819-1796

RESEARCH INTERESTS

Text Mining • Graph Mining • Information Retrieval • Recommender System

My research interests lie in modeling unstructured data (e.g., text), structured data (e.g., table, graph), and their intersection. I strive to answer the following questions.

- What structured knowledge do pretrained language models / large language models contain?
- How to leverage external structure knowledge to better enhance pretrained language models / large language models?
- How to empower structure learning (e.g., graph learning, table learning) with the associated textual semantics signal?

EDUCATION

University of Illinois at Urbana-Champaign

Ph.D. Student in Computer Science. Advisor: Prof. Jiawei Han

2021.08 - Present

Tsinghua University

B.S. in Electrical Engineering & Statistics. Advisor: Prof. Yong Li
Outstanding Graduates (Top 1%).

2017.09 - 2021.07

RESEARCH EXPERIENCE

University of Illinois at Urbana-Champaign

2021.08 - Present

- Research Assistant. Data Mining Group. Advisor: Prof. Jiawei Han
- Project: Representation Learning on Text-rich Networks with Pretrained Language Models (ICLR'23, ACL'23, KDD'23)

Amazon Search

2023.05 - 2023.08

- Research Intern. Query Understanding Group. Mentor: Dr. Xianfeng Tang
- Project: Semantic ID and Generative retrieval.

Microsoft Research

2022.05 - 2022.08

- Research Intern. Information and Data Science Group. Mentors: Dr. Chenyan Xiong and Alec Berntson
- Project: Dense Retrieval for Heterogeneous Data.

Microsoft Research

2020.09 - 2021.03

- Research Intern. Social Computing Group. Mentors: Dr. Zheng Liu and Dr. Xing Xie
- Project: Knowledge-empowered News Recommendation.

University of California Los Angeles

2020.07 - 2020.09

- Research Assistant. Data Mining Group. Mentor: Prof. Yizhou Sun
- Project: Kernel-based Graph Pooling for Graph representation Learning.

Tsinghua University

2018.09 - 2020.07

- Research Assistant. Future Intelligence Lab. Advisor: Prof. Yong Li
- Project: Recommendation with Graph Neural Networks (SIGIR'20)

University of Michigan

2019.07 - 2019.09

- Research Assistant. Michigan Institute for Data Science. Advisor: Prof. Ji Zhu
- Project: Statistical Network Analysis

SELECTED PUBLICATIONS

Bowen Jin, Yu Zhang, Qi Zhu, and Jiawei Han.

Heterformer: Transformer-based Deep Node Representation Learning on Heterogeneous Text-Rich Networks.
The 29th ACM SIGKDD Conference on Knowledge Discovery & Data Mining (KDD 2023).

Bowen Jin, Wentao Zhang, Yu Zhang, Yu Meng, Xinyang Zhang, Qi Zhu, and Jiawei Han.

Patton: Language Model Pretraining on Text-rich Networks.
The 61st Annual Meeting of the Association for Computational Linguistics (ACL 2023).

Bowen Jin, Yu Zhang, Yu Meng, and Jiawei Han.

Edgeformers: Graph-Empowered Transformers for Representation Learning on Textual-Edge Networks.
The 11th International Conference on Learning Representations (ICLR 2023).

Bowen Jin, Chen Gao, Xiangnan He, Depeng Jin, and Yong Li.
Multi-behavior Recommendation with Graph Convolutional Networks.
The 43rd ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2020).

Yu Zhang, **Bowen Jin**, Xiusi Chen, Yanzhen Shen, Yunyi Zhang, Yu Meng, and Jiawei Han.
Weakly-supervised Multi-label Classification of Full-Text Scientific Papers.
The 29th ACM SIGKDD Conference on Knowledge Discovery & Data Mining (KDD 2023).

Yu Zhang, **Bowen Jin**, Qi Zhu, Yu Meng, and Jiawei Han.
The Effect of Metadata on Scientific Literature Tagging: A Cross-Field Cross-Model Study.
The 2023 ACM Web Conference (WWW 2023).

Pengcheng Jiang, Shivam Agarwal, **Bowen Jin**, Xuan Wang, Jimeng Sun, and Jiawei Han.
Text-augmented Open Knowledge Graph Completion via Pretrained Language Models.
The 61st Annual Meeting of the Association for Computational Linguistics (ACL 2023 Findings).

Bowen Jin, Wentao Zhang, Yu Zhang, Yu Meng, Han Zhao, and Jiawei Han.
Learning Multiplex Embeddings on Text-rich Networks with One Text Encoder.
(submitted to ICLR 2023).

Bowen Jin, Hansi Zeng, Guoyin Wang, Xiusi Chen, Tianxin Wei, Ruirui Li, Zhengyang Wang, Zheng Li, Yang Li, Hanqing Lu, Suhang Wang, Jiawei Han, and Xianfeng Tang.
Language Models as Semantic Indexers.
(submitted to ICLR 2023).

Tianxin Wei, **Bowen Jin**, Ruirui Li, Hansi Zeng, Zhengyang Wang, Jianhui Sun, Qingyu Yin, Hanqing Lu, Suhang Wang, Jingrui He, and Xianfeng Tang.
Towards Universal Multi-Modal Personalization: A Language Model Empowered Generative Paradigm.
(submitted to ICLR 2023).

Qi Zhu, Yu Zhang, **Bowen Jin**, and Jiawei Han.
Author Name Disambiguation via Graph-Enhanced Language Model Finetuning.
(submitted to EMNLP 2023).

PROFESSIONAL SERVICES

Conference Program Committee Member

WSDM 2023; KDD 2023; NeurIPS 2023;
ICLR 2024; WWW 2024; SDM 2024;

Journal Reviewer

IEEE Transactions on Knowledge and Data Engineering (TKDE);
ACM Transactions on Information System (TOIS);

Student Volunteer

KDD 2023

Guest Instructor

UIUC CS512 Fall23

AWARDS

2023	KDD 2023 Student Travel Grant
2021	Outstanding Graduates, Tsinghua University (Top 1%)
2021	“Star of Tomorrow” Award, Microsoft Research
2020	National Scholarship (Top 1%)
2019	National Scholarship (Top 1%)
2019	Honorable Mention (top 15.35%), Mathematical Contest in Modeling
2018	National Scholarship (Top 1%)
2017	First Prize, National Olympiad in Mathematics in Provinces

TECHNICAL STRENGTHS

Skills	Machine Learning, Natural Language Processing, Language Modeling, Graph Mining, Weakly Supervised Learning, Unsupervised Learning, Information Retrieval
Programming Languages	Python, C/C++, MATLAB, R, Linux, Markdown, Shell, SQL
Machine Learning Packages	PyTorch, Keras, HuggingFace Transformers, Scikit-learn, PyTorch-Geometric, Deep Graph Library
Tools	Bash, L ^A T _E X, Git