

Kehan Pang

Target Position: Algorithm Researcher | Data Governance / Machine Learning / Large Language Models

☎ +86 132 6451 8388 ✉ pangkehan@buaa.edu.cn 🏠 Beijing, China

🎓 Education

Beihang University Software Engineering Ph.D. Candidate 2022.09 – 2027.04

- Advisors: Wenfei Fan and Ping Lu
- Research Interests: Graph Data Mining, Graph Data Cleaning, Explainability of GNNs

Beijing University of Posts and Telecommunications Computer Science B.Eng. 2018.09 – 2022.07

- GPA: 3.7/4.0 Rank: Top 7.6%

🔧 Work Experience

SICS Fundamental Research Department Research Intern 2023.07 – 2024.01

- Conducted research on graph data governance, graph data quality, graph data augmentation, explainability of graph neural networks, graph computing system architecture, and the adaptation of large language models to low-resource settings on large-scale graph data. Contributed to multiple research papers published at CCF A venues, including TODS and ICDE.

Meituan Core Local Commerce, M17 Algorithm Research Intern 2024.09 – 2025.01

- Developed methods for cleaning and evaluating web data used in foundation-model pretraining, and deployed and fully fine-tuned a local 7B language model. On a company-provided, human-annotated web dataset, the model achieved approximately 96% of the GPT-4 baseline in ROUGE-L and improved upon conventional data-cleaning methods by approximately 6%.

🏛️ Research Experience

Making It Tractable to Detect and Correct Errors in Graphs

- Wenfei Fan, Kehan Pang, Ping Lu, and Chao Tian. **TODS, CCF A**, 2025.
- Addressed the limited semantics of conventional graph association rules when applied to temporal graphs. Proposed a model for ranking nodes by timeliness and a method for adapting static graph rules to dynamic graphs, improving the detection of outdated and erroneous data under continuous graph updates.

Explaining GNN Negatives Globally and Locally (Accepted, to appear)

- Kehan Pang, Wenfei Fan, Dandan Lin, and Min Xie. **ICDE, CCF A**, 2026.
- Proposed a unified rule-learning framework for global and local explanations of negative GNN predictions, modeling explanations through the joint characterization of graph structure and features. Designed a reinforcement-learning-based rule mining method and an LLM-based predicate ranking strategy, with results outperforming existing explanation methods on multiple real-world datasets.

Imputing Sparse and Noisy Labels for GNNs

- Wenfei Fan, Kehan Pang, and Chao Tian. **ICDE, CCF A**, 2025.
- Developed a data augmentation and label governance method for GNNs under sparse and noisy labels. Used an LLM to extract task-relevant features from unstructured and semi-structured attributes, generated structure-aware pseudo-labels with Revised-1-WL, and improved downstream classification through soft-label inference and label cleaning.

OpenMind: Query-Time Diversification via Logical Rules

- Wenfei Fan, Kehan Pang, and Dandan Lin. **Under Revision**.
- Proposed a query-time diversification framework to mitigate information narrowing in recommender systems. Designed joint reasoning and pruning algorithms for CTR and Top- k recommendation, enabling viewpoint-diverse recommendations on large-scale interaction graphs without retraining the underlying recommendation model.

ComCE: Community-Aware Counterfactual Explanations for Graph Neural Networks

- Kehan Pang and Dandan Lin. **Under Review**.
- Proposed a community-aware counterfactual explanation framework for GNNs to improve semantic consistency and computational efficiency. Proved that the minimum counterfactual explanation problem is NP-hard and designed a two-stage optimization method combining hierarchical search and reinforcement learning, achieving better results than existing methods on multiple real-world graph datasets.

Efficient Mixture of Experts Based on Large Language Models for Low-Resource Data Preprocessing

- Mengyi Yan, Yaoshu Wang, Kehan Pang, Min Xie, and Jianxin Li. **SIGKDD, CCF A**, 2024.
- Studied information loss and redundancy arising from the integration of multiple experts in mixture-of-experts models. Proved that meta-path selection for data augmentation satisfies the Church–Rosser property and designed an information-bottleneck architecture for the RAG module, improving downstream performance relative to individual expert models.

Linking Entities across Relations and Graphs

- Wenfei Fan, Ping Lu, Kehan Pang, Ruochun Jin, and Wenyuan Yu. **TODS, CCF A**, 2024.
- Developed methods for entity alignment across heterogeneous graphs and relational tables under inconsistent semantics. Designed an entity alignment algorithm, a partition-based indexing method for growing datasets, and an incremental alignment algorithm, supporting efficient processing on large-scale data.

♥ Awards and Academic Service

- First Prize, Beijing College Student Mathematics Competition, 2019 and 2020
- First Prize, National College Student Mathematics Competition, 2019 and 2020
- Honorable Mention, Mathematical Contest in Modeling, 2020
- BUPT Undergraduate Scholarship, 2019–2021
- Beihang University Academic Scholarship, 2022–2025
- External Reviewer, ICDE, 2024–2026

⚙ Research and Technical Expertise

- **Academic Research:** Independent research, academic writing, and cross-team collaboration.
- **Data Governance:** Graph, relational, and unstructured data governance, including entity alignment, error correction, label cleaning, feature extraction, and incremental processing.
- **Graph Computing and Machine Learning:** Graph rules, graph mining, GNNs, and explainability, with experience in indexing, incremental computation, pruning, and distributed graph processing.
- **Large Language Models:** LLM deployment, LoRA and full-parameter fine-tuning, RAG, information retrieval, semantic representation, and model evaluation.
- **Implementation and Experimentation:** Proficient in C++ and Python, with experience in Linux-based system development, large-scale experiments, performance analysis, and parallel computing.
- **Language Proficiency:** CET-6 English; proficient in reading and writing academic papers, technical documents, and research reports.