

Jifan Zhang

jifanzhang2026@u.northwestern.edu | +1 773 312 1101

Education

Northwestern University | Ph.D. in Statistics & Data Science

Sep 2021 – Jun 2026

Advisor: Miklós Rácz GPA: 3.97/4.0 Focus: AI & Graph Learning

Tsinghua University | B.S. in Mathematics

Sep 2017 – Sep 2021

Advisor: Qian Lin Major GPA: 3.79/4.0 Focus: Pure Math & Applied Math

Professional Experience & Projects

PhD Decision Science Intern | Epsilon

Jun 2025 – Sep 2025

Robust CTR Controller in Large-Scale Digital Advertising

- Developed end-to-end CTR prediction system processing **200M campaigns** across **30K branches**, utilizing CatBoost feature selection, categorical embeddings, and LSTM-based unified deep learning architecture that achieved **60% correlation improvement** over baseline models.
- Deployed production-ready CTR prediction pipeline with threshold-based filtering, delivering **10% average CTR lift** and **100% target goal achievement rate** in real-world campaign simulations for Dairy Queen and CVS.
- Optimized training infrastructure on Databricks Spark cluster, implementing cold-start weekly training and warm-start daily fine-tuning strategy that reduced daily training runtime by **85%** (from 4 hours to 30 minutes).
- Tech stack: Python, PyTorch, CatBoost, Spark/Databricks, LSTM, pandas, scikit-learn.

Lead Researcher | Northwestern University

Jan 2025 – Present

Uncertainty Quantification for Spatio-Temporal Graph Forecasting

- Innovated STACI, a topology-aware, **model-free** conformal uncertainty quantification framework for graph-structured multivariate time series with theoretical analysis on finite-sample coverage and optimization of *ellipsoidal prediction sets* adapting to the high-dimensional manifold structure.
- Integrated STACI with multiple spatio-temporal backbones (AGCRN, ASTGCN, STGODE), achieving **100%** target coverage while reducing prediction-set volume by at least **15%** on PEMS traffic data vs. UQ baselines (DEEPSTUQ, conformal variants), showcasing *SOTA* performance in reliability–efficiency trade-offs.

Theoretical Foundations for Network Inference: Graph Matching, Community Recovery, Isomorphism

- Advanced theoretical foundations for learning on (multiple) networks, deriving sharp phase transitions and algorithms with implications to large-scale network analytics and recommender systems.
- Graph Matching & Community Recovery:** Designed algorithms that achieve *tight thresholds* for community recovery and graph matching across multiple correlated stochastic block models. **Initiated** the study of *regular* sparse SBMs and proved that matching $O(\log n)$ sparse graphs enables *exact* community recovery.
- Graph Isomorphism & Subgraph Counting:** Established *sharp* phase transitions for isomorphic 1-neighbourhoods in random graphs. Established a *local central limit theorem* for sparse-regime subgraph counts.

Lead Researcher | The Institute for Data, Econometrics, Algorithms, and Learning

Oct 2024 – Jun 2025

Causal Representation Learning for Network-Structured Genomics

- Proposed GraCE-VAE, a causal disentanglement framework by integrating graph topology into VAEs, yielding causal latent representations for multivariate genomics data with identifiability guarantees.
- Experimented on *Norman & Replogle* datasets (**300K** samples, **8,000** dimensions), improved **generalization to unseen interventions** with **5% lower MMD** and **3% higher R^2** versus strong baselines (CMVAE, GEARS).

ML Research Collaborator | ByteDance

Sep 2020 – Nov 2020

User Online Trend Prediction

- Predicted **daily online activity** for **100K** TikTok users over **3 years**; achieved **20% higher correlation** on the held-out test set via feature selection and *XGBoost*-derived features, and built a *Factorization Machines* predictor.
- Built user-behavior clusters and trained category-specific models; for hard-to-predict “*middle*” users, introduced a temporal *LSTM* model, improving **F1 score by 15%** compared with single prediction model.
- Constructed an production-ready **online daily prediction pipeline** (feature generation, model inference, monitoring), enabling more precise ad targeting and improved campaign efficiency.

Summer Researcher | Massachusetts Institute of Technology

Jul 2020 – Sep 2020

High-Resolution Astronomical Image Generation with GANs

- Implemented a *Progressive GAN* with Wasserstein loss on *Linux* clusters using **multi-GPU** training, improving throughput and time-to-quality.
- Synthesized **512×512** astronomical images and ran standardized evaluation; achieved 24 **FID score** (decrease by 10% vs. baseline) with fixed seeds and matched splits.

Data Analysis Intern | Huatai Securities

Dec 2019 – Mar 2020

Cointegration Analytics for Equity Pairs

- Built an end-to-end *cointegration stat-arb* pipeline over **1,889** equities (2019–2020), including *ADF/EG/DW* tests, rolling and change-point diagnostics, *VAR/VECM* stability, and *Johansen* multivariate cointegration; pivoted to multi-asset pair modeling and backtesting.
- Executed a z-score threshold + safety-band hedging strategy: on pair (600528.SH, 000008.SZ) achieved annualized return of **27%** (train) and **25%** (test); on (000046.SZ, 600981.SH) achieved **17.8%** (test), illustrating the advantage of cointegration in marketing.

Core Competencies

Deep Learning & AI: Graph Neural Networks, Causal AI, Generative Models, Uncertainty Quantification, Time-Series Modeling

MLOps & Production: Large-scale ML Systems, Real-time Processing, Parallel Computing; Spark/Databricks

Business Impact: CTR Optimization, Recommender Systems, User Behavior Prediction

Programming: Python, R, C++, MATLAB

Selected Publications & Preprints

Harnessing Multiple Correlated Networks for Exact Community Recovery

Sep 2024

Jifan Zhang, Miklós Rácz. *Advances in Neural Information Processing Systems (NeurIPS 2024)*.

Topology-Aware Conformal Prediction for Stream Networks

May 2025

Jifan Zhang, Fangxin Wang, Zihe Song, Kaize Ding, Shixiang Zhu.

Advances in Neural Information Processing Systems (NeurIPS 2025).

Causal Representation Learning from Network Data

Jul 2025

Jifan Zhang, Michelle Li, Elena Zheleva.

Bridging Pretraining and Supervised learning in Knowledge Graph

Sep 2025

Jifan Zhang, Suqi Liu, Miklós Rácz.

When Local Neighbourhoods Become Distinct in Random Graphs

Oct 2025

Jifan Zhang, Miklós Rácz.

Honors & Leadership

Northwestern University Fellowship

2021–2022

First Prize, China Undergraduate Mathematical Contest in Modeling (Beijing Region)

2019

Honor of Comprehensive Excellence, Dept. of Mathematics, Tsinghua University

2018

President, Student Science Association, Dept. of Mathematics, Tsinghua University

2019–2020