

Jifan Zhang

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

Education

Northwestern University , PhD Candidate in Statistics and Data Science	Sept 2021 – Sept 2026
Advisor: Miklós Rácz	Cumulative GPA: 3.96
Tsinghua University , Bachelor of Science in Mathematics	Sept 2017 – Sept 2021
Advisor: Qian Lin	Major GPA: 3.79

Research Interest

Boardly interested in statistics, probabilistic combinatorics and computer science, with a focus on probabilistic generative network models and empirical study of utilizing graph structure in machine learning and causal inference.

Research Experience

Graduate Research , Northwestern University, IL	Feb 2023 – present
Advisor: Prof. Miklós Rácz	

- Characterize the tight information-theoretic threshold for exact community recovery given a constant number of correlated stochastic block models. Quantifies the value of each additional correlated graph for community recovery and showcase the subtle interplay between graph matching and community recovery.
- Derive the information -theoretical threshold for Erdős–Rényi random graph when 1 - neighbourhoods are distinct. The critical regime $p \asymp n^{-161/302}$ is such that the induced subgraph on the neighbors of a typical vertex contains copies of all trees on at most eight vertices but does not contain trees on nine vertices with high probability.
- Study the threshold for community recovery by integrating $O(\log n)$ of correlated regular stochastic block model in sparse regime. Characterize the region where weak recovery is unattainable in a single graph, but exact recovery becomes feasible by integrating information from $O(\log n)$ many graphs via graph matching.

Summer Research , The Institute for Data, Econometrics, Algorithms, and Learning	June 2024 - Oct 2024
Collaborator: Michelle M. Li, Prof. Elena Zheleva	

- Enhance the causal disentanglement framework by integrating graph structures into the autoencoding variational Bayes model, to improve the prediction of unseen combinatorial perturbation effects in genomics.
- The proposed model demonstrated significant improvements, particularly in terms of Maximum Mean Discrepancy loss and the R-squared metric for differentially expressed genes.

Summer Research , Northwestern University, IL	May 2024 – Oct 2024
Collaborator: Fangxin Wang, Prof. Kaize Ding	

- Proposed an Adaptive Conformalized Spatio-temporal Graph Neural Network (ACSGNN) capable of performing multi-step ahead predictions with guaranteed coverage of true values, for time series with spatial dependencies.
- Achieved a 25% reduction in inefficiency, as measured by the mean predicted interval width, compared to the state-of-the-art model on the traffic forecasting dataset.

Publications

Harnessing Multiple Correlated Networks for Exact Community Recovery	Sep 2024
Miklós Rácz*, Jifan Zhang*	
Advances in the 38th Conference on Neural Information Processing Systems(NeurIPS 2024)	

Working Paper

When 1 - neighborhoods become distinct in Erdős–Rényi random graphs	Forthcoming
Miklós Rácz*, Jifan Zhang*	
Adaptive conformalized forecasting for multiple time series with spatial dependencies	Forthcoming
Jifan Zhang*, Fangxin Wang, Kaize Ding	
Improved causal structure learning on genomics	Forthcoming
Jifan Zhang*, Michelle M. Li, Elena Zheleva	

Internship And Projects

Generative Adversarial Networks of galaxy images , Massachusetts Institute of Technology	Jul 2020 - Sep 2020
Advisor: Prof. Mark Vogelsberger	

- Developed a Progressive Generative Adversarial Networks model with the Wasserstein GAN with Gradient Penalty loss function to generate the first high-resolution (512×512) synthetic galaxy images for data augmentation in astronomy, and rigorously evaluated their astronomical distribution properties by calculating the Frechet Inception Distance score to compare them with real cosmological images.

TikTok user's online behavior prediction, ByteDance Company, Tsinghua University

Sep 2020 - Nov 2020

Advisor: Prof. Qian Lin

- Analyzed behavioral data of 100,000 TikTok users over three years to predict online behavior, utilizing clustering methods and Linear Discriminant Analysis to categorize users into distinct groups, and applied the CatBoost algorithm to identify key features, achieving a 20% reduction in Mean Squared Error compared to the baseline model.

Market timing models, Data Analysis Internship for Huatai Securities

Dec 2019 - Mar 2020

- Developed market timing models by simulating different stock price paths using the Monte Carlo method based on 30 days of historical data. Calculated the maximum drawdown rate and selected a buy-in strategy that minimized this rate, resulting in an improved and satisfactory Sharpe Ratio.

Teaching Experience

Teaching Assistance for the courses:	2022-2024
• STAT 457 Applied Bayesian Inference	
• STAT 430-1 Probability for Statistical Inference 1	
• STAT 430-2 Probability for Statistical Inference 2	
• STAT 357 Introduction to Bayesian Statistics	
• STAT 330 Applied Statistics for Research 1	
• STAT 320-3 Statistical Theory & Methods 3	
• STAT 201 Introduction to Statistics and Data Science	

Honors And Additional Experience

Northwestern University Fellowship	2021-2022
First Prize of China undergraduate Mathematical Contest in Modelling , Beijing Region	2019
Honor of Comprehensive Excellence , Department of Mathematics, Tsinghua University	2018
President of Student Science Association , Department of Mathematics, Tsinghua University	2019-2020
Member of the Women's Basketball Team and the Badminton Team , Department of Mathematics, Tsinghua University	2019-2021

Technical Skills

- Programming languages: Python, R, C++, Matlab
- Areas of expertise: Statistical modeling, Network analysis, Machine learning, Time series analysis.