

Jinrui (Jerry) Gou

Computer Science and Engineering
Tandon School of Engineering
New York University

Tel : (347) 200 5159
Email: jg6226@nyu.edu
HomePage: jg6226.github.io

EDUCATION	New York University Ph.D. in Computer Science Advisor: Professor Torsten Suel Cumulative GPA: 3.97/4.0 M.S. in Computer Science Cumulative GPA: 4.0/4.0	Brooklyn, NY 09/2021-Present 09/2019-05/2021
	Sun Yat-Sen University B.E. in Computer Science & Technology Cumulative GPA: 90/100 (3.8/4.0) Mathematics Courses GPA: 95/100	Guangzhou, China 09/2015-06/2019

KNOWLEDGE

- **Languages:** C/C++, Python, SQL, AMPL/CPLEX, Latex, MATLAB, Bash, HTML5.
 - **Tools:** Scikit-learn, PyTorch, TensorFlow, Numpy, Matplotlib, TBB, AVX/AVX2, CUDA.
 - **Courses Enrolled:** Computer Architecture, Natural Language Processing, Machine Learning (Rank 1st in Class of 43 students), Deep Learning, Web Search Engines, Design & Analysis of Algorithms (Rank 1st in Class of 47 students), Network Design and Algorithms (Rank 2nd in Class of 20 students), Algorithmic Machine Learning and Data Science.

RESEARCH EXPERIENCE

Ph.D. Research Assistant for Prof. Torsten Suel, NYU

09/2021-Present

1. Effective Candidate Generation and Threshold Estimation for Fast Top-K Query Processing
 - Develop efficient and effective early termination algorithms for simple ranking functions.
 - Design specialized prefixes index structures and fine-tune its storage for efficient access.
 - Perform fast lookups using highly optimized index structures to improve the quality of results.
 - Implement efficient C++ code for experiments on different collections and ranking functions.
 2. Graph Based Approximate Near-neighbor Search
 - Propose navigable graph that can perform greedy routing for nearest neighbor search.
 - Design efficient randomized algorithms to create effective sparse navigable graphs.
 - Bound the maximum degree and average degree of navigable graphs using mathematical proof.
 - Develop efficient search algorithms for navigable graphs in different benchmark graph datasets.

3. Sparse Learned Index Structures Optimization

- Expand documents by generating related questions, using LLM Llama 2.
- Use ELECTRA-based relevance model to improve document expansion quality.
- Train BERT-based language model to predict importance of terms in documents.
- Fuse the knowledge from cross-encoder to improve training, using hard negatives and distillation.

Graduate Research Assistant for Prof. Yong Liu, Networked Systems Lab, NYU

09/2019-05/2021

1. Realtime Mobile Network Prediction for High QoS

- Collect mobile 4G/LTE bandwidth traces in metro area of NYC.
- Develop LSTM RNN models for high-accuracy realtime bandwidth prediction.
- Develop Gradient Boosting models for handoffs predictions between 4G and 5G access modes.

2. Joint Traffic Routing and Computation Server Placement in Edge Cloud Networks

- Build quasi-convex optimization model to balance traffic and computation resource allocation.
- Use AMPL/CPLEX and Python optimization packages to obtain optimal feasible solutions.

PUBLICATIONS

1. Yousef Al-Jazzazi*, Haya Diwan*, **Jinrui Gou***, Cameron Musco*, Christopher Musco*, Torsten Suel*, “Distance Adaptive Beam Search for Provably Accurate Graph-Based Nearest Neighbor Search”, *NeurIPS 2025: Advances in Neural Information Processing Systems, December 2025*¹
2. **Jinrui Gou**, Antonio Mallia, Yifan Liu, Minghao Shao, and Torsten Suel, “Fast and Effective Early Termination for Simple Ranking Functions”, *ACM SIGIR Conference on Research and Development in Information Retrieval, July 2025*.
3. Haya Diwan*, **Jinrui Gou***, Cameron Musco*, Christopher Musco*, and Torsten Suel*, “Navigable Graphs for High-Dimensional Nearest Neighbor Search: Constructions and Limits”, *NeurIPS 2024: Conference on Neural Information Processing Systems, December 2024*.²
4. **Jinrui Gou**, Yifan Liu, Minghao Shao, and Torsten Suel, “Beyond Quantile Methods: Improved Top-K Threshold Estimation for Traditional and Learned Sparse Indexes”, *IEEE International Conference on Big Data, December 2024*.
5. Soyuj Basnet, **Jinrui Gou**, Antonio Mallia, and Torsten Suel, “DeeperImpact: Optimizing Sparse Learned Index Structures”, in *ReNeuIR at SIGIR 2024: The Third Workshop on Reaching Efficiency in Neural Information Retrieval, July 2024*.
6. Lifan Mei, **Jinrui Gou**, Jingrui Yang, Yujin Cai, and Yong Liu, “On Routing Optimization in Networks with Embedded Computational Services”, *IEEE Transactions on Network and Service Management, September 2024*.
7. Lifan Mei, **Jinrui Gou**, Yujin Cai, Houwei Cao, and Yong Liu, “Realtime mobile bandwidth and handoff predictions in 4G/5G networks”, *Computer Networks, February 2022*.

¹All authors contributed equally.

²All authors contributed equally.