Ta-Yang Wang

(310) 245-2221 | Google Scholar | tayang.w@databricks.com | in/ta-yang | tayangwang.github.io

Professional Summary

Software Engineer at Databricks and Ph.D. in Computer Science, leveraging deep expertise in Machine Learning and Graph Neural Networks (GNNs) to tackle complex challenges. Proven ability to translate advanced research into scalable system solutions, including accelerating large-scale model training and optimizing distributed resources. Passionate about applying ML to real-world applications.

Education

University of Southern California (USC) — Los Angeles, CA

Ph.D. in Computer Science — *Aug. 2019 – Oct. 2025*

- Dissertation — Heterogeneous Graph Representation Learning: Algorithms and Applications
- Reviewer: WWW 2025, NeurIPS 2024, FPGA 2022 & 2023, SC 2022, CLOUD 2022, HPEC 2021, CCGrid 2021

National Taiwan University (NTU) — Taipei, Taiwan

B.S. in Mathematics | Overall GPA: 3.96/4.3 | Rank: 2nd/55 — *Sep. 2014 – June 2018*

- Honorable Mention, 9th S.-T. Yau College Student Mathematics Contest
  - Achieved finalist status (top 15) in the Applied and Computational Mathematics.
- National Taiwan University Presidential Award (Twice)
  - Awarded to students with the top 5% GPA in the department for that semester.

Experience

Databricks — Mountain View, CA

Software Engineer — *Oct. 2025 – Present*

- Build the core infrastructure for the Data Intelligence Platform, specifically focusing on Data Quality Monitoring and Anomaly Detection.
- Develop scalable, automated data quality checks within Unity Catalog to ensure data integrity, completeness, and freshness across petabyte-scale data lakes.
- Design and implement advanced anomaly detection algorithms to proactively identify and alert on irregular data patterns and quality issues in real-time.

Data Science Lab, University of Southern California — Los Angeles, CA

Graduate Research Assistant — *Aug. 2019 – Oct. 2025*

- Developed a novel bandit-sampling approach to address a key challenge in real-world graph representation learning — missing node attributes in Heterogeneous Graph Neural Networks (HGNNs).
- Accelerated large-scale HGNN training by 50% while maintaining state-of-the-art (SOTA) accuracy, publishing an adaptive sampling method with a theoretical convergence guarantee.

Medical Data Analytics Lab, National Center for Theoretical Sciences — Taipei, Taiwan

Research Assistant — *Sep. 2018 – June 2019*

- Collaborated with National Taiwan University Hospital to advance medical imaging for clinical applications.
- Designed a novel model for prostate cancer histopathology image classification using persistent homology.

Taiwan Chapter of the Society of American Baseball Research (SABR) — Taipei, Taiwan

Co-founder — *Sep. 2017 – June 2018*

- Architected an end-to-end data pipeline to ingest and analyze large-scale TrackMan datasets, delivering a novel quantitative player evaluation system for Taiwan's Professional Baseball League (CPBL).
- Developed predictive models using StatCast and PITCHf/x data to create objective, data-driven metrics for Major League Baseball (MLB) pitch quality, informing player assessment.

Skills

Programming Languages: Python, C/C++, R, Matlab, SQL
Frameworks & Libraries: PyTorch, TensorFlow, Keras, NumPy, Pandas, Matplotlib, SciPy
Developer Tools: Git, Docker, VS Code, Visual Studio, PyCharm
Teaching: Mentored 4,000+ grad/undergrad students as a Teaching Assistant for core CS/Math courses:

- **USC CSCI 570: Analysis of Algorithms** Fall 2019 – 2022 & 2024, Spring 2020 & 2024 – 2025
- **USC CSCI 170: Discrete Methods in Computer Science** Fall 2023
- **USC CSCI 270: Introduction to Algorithms and Theory of Computing** Spring 2021, 2023
- **NTU MATH5425: Introduction to Cryptography** Spring 2018

Publications

Graph Neural Networks

- **Ta-Yang Wang**, Rajgopal Kannan, and Viktor Prasanna. "Effective and Generalizable Pre-Trained Heterogeneous Graph Neural Networks with Bandit Samplers." [Under Review]
- **Ta-Yang Wang**, Rajgopal Kannan, and Viktor Prasanna. "TypeBandit: Leveraging Type-Based Bandit Sampling for Effective Attribute Completion in Heterogeneous Graph Neural Networks." *IEEE Transactions on Big Data*. [Under Review]
- **Ta-Yang Wang**, Rajgopal Kannan, and Viktor Prasanna. "Heterogeneous Graph Neural Network based on Bandit Sampling for Attribute Completion." *The 15th IEEE International Conference on Knowledge Graphs (ICKG)*, 2024.
- **Ta-Yang Wang**, Rajgopal Kannan, and Viktor Prasanna. "Training Heterogeneous Graph Neural Networks using Bandit Sampling." *The 32nd ACM International Conference on Information and Knowledge Management (CIKM)*, 2023.
- **Ta-Yang Wang**, Hongkuan Zhou, Rajgopal Kannan, Ananthram Swami, and Viktor Prasanna. "Throughput Optimization in Heterogeneous MIMO Networks: A GNN-based Approach." *The 1st Graph Neural Networking Workshop (GNNet)*, 2022.

Task Mapping

- **Ta-Yang Wang**, William Chang, Ajitesh Srivastava, Rajgopal Kannan, and Viktor Prasanna. "Monte Carlo Tree Search for Task Mapping onto Heterogeneous Platforms." *The 28th IEEE International Conference on High Performance Computing, Data, and Analytics (HiPC)*, 2021.
- **Ta-Yang Wang**, Ajitesh Srivastava, and Viktor Prasanna. "A Framework for Task Mapping onto Heterogeneous Platforms." *IEEE High Performance Extreme Computing Conference (HPEC)*, 2020.

Tensor Decomposition

- Sasindu Wijeratne, **Ta-Yang Wang**, Rajgopal Kannan, and Viktor Prasanna. "Accelerating sparse MTTKRP for tensor decomposition on FPGA." *The 32nd ACM/SIGDA International Symposium on Field-Programmable Gate Arrays (FPGA)*, 2023.
- Sasindu Wijeratne, **Ta-Yang Wang**, Rajgopal Kannan, and Viktor Prasanna. "Towards Programmable Memory Controller for Tensor Decomposition." *The 11th International Conference on Data Science, Technology and Applications (DATA)*, 2022.

Memory Access Prediction

- Pengmiao Zhang, Ajitesh Srivastava, **Ta-Yang Wang**, Cesar AF De Rose, Rajgopal Kannan, and Viktor Prasanna. "C-MemMAP: Clustering-driven Compact, Adaptable, and Generalizable Meta-LSTM Models for Memory Access Prediction." *International Journal of Data Science and Analytics (JDSA)*, 2021.
- Ajitesh Srivastava, **Ta-Yang Wang**, Pengmiao Zhang, Cesar Augusto F De Rose, Rajgopal Kannan, and Viktor Prasanna. "MemMAP: Compact and Generalizable Meta-LSTM Models for Memory Access Prediction." *The 24th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*, 2020.