WENZHUO TANG

Email: tangwen2@msu.edu ◇ Website: wenzhuotang.github.io

RESEARCH STATEMENT

I am a final-year PhD student at Michigan State University, advised by Prof. Yuying Xie and Prof. Jiliang Tang. My research centers on building scalable foundation models for structured data, with expertise in graph neural networks, transformers, diffusion models, and many others. I design models that generalize across domains and tasks by combining advanced representation learning with scalable modeling of massive, heterogeneous datasets. These models have been applied to networks and single-cell biology and achieved top-tier performance on Kaggle benchmarks, demonstrating both methodological innovation and practical impact.

EDUCATION

Michigan State University 09/2021 - 05/2026 (Expected)
Ph.D., Statistics and Computer Science

University of Science and Technology of China 09/2017 - 06/2021
B.S., Statistics

EXPERIENCES

Stanford University 06/2024 – Present
Visiting Researcher — Advisor: *Xiaojie Qiu*

- Developed multiscale foundation models that integrate large-scale spatial and temporal data, and advanced interpretable dynamic modeling through neural ordinary differential equations.
- Contributed to PantheonOS, a multi-agent AI operating system that enables automated data analysis with command line interaction, notebook integration, and domain-specific expertise.
- Applied these frameworks to single-cell and spatial genomics, yielding insights into embryonic development and disease mechanism.

Michigan State University 09/2022 – Present
Research Assistant — Advisors: *Yuying Xie, Jiliang Tang*

- Advanced machine learning methods for structured and multimodal data, including graph neural networks [1, 4], transformers [2, 4, 6], and diffusion and GAN-based models [1, 5, 8].
- Advanced foundation models by (1) pretraining a large-scale single-cell model capturing transferable genetic and cellular representations [2], and (2) articulating a vision for graph foundation models [3] and realizing it with a pretrained diffusion model that improves graph tasks across domains [1].
- Achieved Top 2% (Silver Medal) in the NeurIPS Kaggle multimodal integration challenge by developing transformer methods [4] with the DANCE team.
- Co-developed the open source DANCE toolkit [1, 7], standardizing preprocessing, benchmarking, and end-to-end pipelines for heterogeneous data.

The Hong Kong Polytechnic University 05/2023 – 08/2023
Research Assistant — Advisor: *Wenqi Fan*

- Reframed prediction and imputation problems as posterior estimation tasks and addressed them using a diffusion-based generative model [5].
- Explored complementarities between large language models and graph machine learning, proposing approaches that advance both paradigms [3].

OPEN-SOURCE PROJECTS

DANCE (core developer)
Open-source deep learning toolkit for single-cell analysis (300+ GitHub stars). Provides standardized preprocessing, benchmarking, and end-to-end workflows for tasks such as cell type annotation, imputation, and modality translation.

PantheonOS (developer)
Open-source multi-agent AI operating system for automated data analysis (200+ GitHub stars). Offers command line interaction, multi-language programming, notebook integration, and incorporates PhD-level expertise in single-cell and spatial genomics.

PUBLICATIONS

**indicates equal contribution*

Conference Papers

[1] **W. Tang**, et al. Cross-Domain Graph Data Scaling: A Showcase with Diffusion Models. *NeurIPS*, 2025.

[2] H. Wen*, **W. Tang***, et al. CellPLM: Pre-training of Cell Language Model Beyond Single Cells. *ICLR*, 2024.

[3] H. Mao*, Z. Chen*, **W. Tang**, et al. Position: Graph Foundation Models Are Already Here. *ICML* **Spotlight**, 2024.

[4] **W. Tang***, H. Wen*, R. Liu*, et al. Single-Cell Multimodal Prediction via Transformers. *CIKM*, 2023.

[5] H. Li, W. Jin, G. Skenderi, H. Shomer, **W. Tang**, et al. Sub-graph Based Diffusion Model for Link Prediction. *LoG*, 2024.

Journal Papers

[1] J. Ding*, H. Wen*, **W. Tang***, R. Liu*, et al. DANCE: A Deep Learning Library and Benchmark Platform for Single-cell Analysis. *Genome Biology*, 2024.

[2] D. Molho*, J. Ding*, **W. Tang***, et al. Deep Learning in Single-Cell Analysis. *ACM Transactions on Intelligent Systems and Technology (ACM TIST)*, 2024.

[3] S. Wang, J. Huang, Z. Chen, Y. Song, **W. Tang**, et al. Graph machine learning in the era of large language models (llms). *ACM TIST*, 2025.

[4] J. Ding, J. Venegas, Q. Lu, Y. Wang, W. Jin, H. Wen, R. Liu, **W. Tang**, et al. Spatialctd: A large-scale tumor microenvironment spatial transcriptomic dataset to evaluate cell type deconvolution for immuno-oncology. *Journal of Computational Biology*, 2024.

Preprints

[5] **W. Tang***, R. Liu*, et al. A General Single-Cell Analysis Framework via Conditional Diffusion Generative Models. *bioRxiv*, 2023.

[6] H. Wen, **W. Tang**, et al. Single Cells Are Spatial Tokens: Transformers for Spatial Transcriptomic Data Denoising. *arXiv*, 2023.

[7] J. Ding, Z. Xing, Y. Wang, R. Liu, S. Liu, Z. Huang, **W. Tang**, et al. DANCE 2.0: Transforming single-cell analysis from black box to transparent workflow. *bioRxiv*, 2025.

[8] Y. Wang*, J. Ding*, L. Wu, A. Wardhani, P. Danaher, Q. Lu, H. Wen, **W. Tang**, et al. MEM-GAN: A Pseudo Membrane Generator for Single-cell Imaging in Fluorescent Microscopy. *bioRxiv*, 2023.

HONORS & AWARDS

Silver Medal (Top 2%) in Kaggle, Team DANCE 2022
NeurIPS Multimodal Single-Cell Integration Across Time, Individuals, and Batches.

Copper Award Scholarship (Top 40%) 2018
University of Science and Technology of China.

Excellent New Student Award (Top 25%) 2017
University of Science and Technology of China.

SERVICES

Conference & Workshop Program Chair

- The Web Conference 2024 Graph Foundation Model Workshop 2024

Conference Session Chair

- ACM International Conference on Information & Knowledge Management (CIKM) 2023

Journal and Conference Paper Reviewer

- NeurIPS 2024 - 2025
- ICLR 2025 - 2026
- ICML 2025
- AISTATS 2025 - 2026
- Bioinformatics 2024
- IEEE Transactions on Artificial Intelligence (IEEE TAI) 2024
- Transactions on Knowledge and Data Engineering (TKDE) 2023 - 2025

Teaching

- Teaching Assistant for Statistical Learning and Data Analysis 2024
  Department of Statistics and Probability. Michigan State University.
- Teaching Assistant for Statistical Methods 2021 - 2023
  Department of Statistics and Probability. Michigan State University.
- Teaching Assistant for Random Process B 2021
  Department of Statistics and Finance. University of Science and Technology of China.

TECHNICAL STRENGTHS

Machine Learning Frameworks PyTorch, PyTorch Geometric, HuggingFace, CUDA
Programming & Systems Python, R, C/C++, MATLAB; Docker, CMake, Git/GitHub