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 focuses on developing foundation models at the intersection of artificial intelligence and biology. I aim to build methods that capture patterns across scales of biological organization, from genes and cells to tissues and organs. Through these models, I seek to advance our understanding of complex developmental processes and the mechanisms of disease.

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*

- Developing a multiscale foundation model of mouse embryogenesis by combining single-cell and spatial transcriptomic data. The model enables *in silico* genetic, cellular, and drug perturbations to study developmental mechanisms and dissect the factors contributing to embryonic defects.
- Applying neural ordinary differential equations to simulate temporal dynamics in embryonic development, with the goal of producing interpretable models that connect computation with biology.
- Contributing to the development of PantheonOS, a multi-agent platform that automates processing, integration, and visualization of single-cell and spatial genomics data, enabling reproducible and scalable analyses.

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

- Applied state-of-the-art machine learning models—including graph neural networks [1, 4, 8], transformers [6, 8, 11], and generative models [13]—to analyze single-cell data. A transformer-based method [8] helped our DANCE team win a Kaggle silver medal in multimodal data integration.
- Co-developed the DANCE toolkit [1, 12], which standardizes preprocessing and benchmarking and supports end-to-end analysis for key biological tasks such as cell type annotation, perturbation prediction, and imputation.
- Pretrained large-scale single-cell foundation models to capture genetic and cellular representations [6]. This work represents one of the first efforts to create universal representations that leverage both single-cell and spatial data and generalize across tissues and experimental platforms.
- Investigated opportunities and challenges of foundation models on graphs [7], and advanced cross-domain pretraining on graphs with a diffusion model [5]. The pretrained diffusion model supports plug-and-play use across diverse downstream tasks by leveraging structure-aware augmentation..

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

- Reframed biological problems such as imputation and perturbation prediction as posterior estimation tasks and solved them with diffusion-based generative models [10].
- Explored how large language models and graph machine learning can complement each other, identifying approaches that benefit both representation learning and biological applications [3].

OPEN-SOURCE PROJECTS

DANCE (core developer)
An extensive deep learning toolkit for single-cell (multi-)omics data with 300+ GitHub stars.

PantheonOS (developer)
A multi-agent AI platform for analyzing single-cell and spatial genomics with 200+ GitHub stars.

PUBLICATIONS

*indicates equal contribution

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.

Conference Papers

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

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

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

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

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

Preprints

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

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

[12] 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.

[13] 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
Biological Data Analysis	scRNA-seq, MERFISH, scATAC-seq, CITE-seq, NGS analysis
Bioinformatics Tools	Scanpy, Seurat, Squidpy, Cell Ranger, Monocle, Bioconductor
Programming	Python, R, C/C++, MATLAB; Docker, CMake, Git/GitHub