Anrui Wang

CS student

RESEARCH INTERESTS

AI for Science, Single-cell Representation Learning, Graph Representation Learning

RESEARCH EXPERIENCE

STATE-Based Multi-Omics Modeling for Synthetic-Lethality Prediction
Supervised by Prof. Jie Zheng Nov. 2025 – Present

- Framed and independently led a staged research program toward synthetic-lethal (SL) target discovery, predicting gene-pair SL interactions for cancer-selective drug-target prioritization.
- Integrated multi-omics evidence into SL pair prediction: Perturb-seq perturbation transcriptomics, DepMap functional dependency (CRISPR GeneEffect), together with ESM2 protein-sequence embeddings.
- Modeled perturbation responses as cell-state representations predictive of cancer dependency, developing from variational single-cell representations (scVI) and multiple-instance pooling to a prototype-based distribution-regression model that recovers dependency signal missed by pseudobulk and attention baselines.
- Built cross-cell-line selectivity features for context-specific dependency, and evaluated SL prediction on a gene-pair benchmark with adversarial leakage.

Topology-Constrained Deep Learning for Inductive Interactome Reconstruction
Supervised by Prof. Xuming He Jan. 2026 – Present

- Reframed protein–protein interaction prediction from independent pairwise classification into inductive reconstruction of the interaction network.
- Designed a topology-constrained conditional generator modeling the full adjacency structure of an unseen protein set via a decomposed edge formulation (pairwise compatibility, hub propensity, community membership, set-level density).
- Instilled topological realism with a training-only topology teacher (masked-edge reconstruction distillation) and differentiable graph-level objectives, while keeping the deployed model strictly inductive on intrinsic protein features (ESM).
- Established credible baselines through systematic architecture ablations, and reproduction of SOTA models.

SKILLS

- **Programming:** Python, PyTorch
- **ML / Tooling:** HuggingFace Accelerate, single-cell foundation models, ESM series, graph neural networks
- **Coding Agents:** highly skilled with agentic coding tools, including Claude Code and Codex
- **Domains:** single-cell genomics, cancer dependency (DepMap), synthetic lethality, protein–protein interaction networks

SELECTED COURSES

Machine Learning, Deep Learning, AI for Science and Engineering, Bioinformatics