Xin (Allen) Wang

805-259-5932 | xin.wang.1@vanderbilt.edu | Website

Education

Vanderbilt University — Aug. 2023 – May 2025
M.S Computer Science, GPA: 4.0/4.0, Thesis Track — *Nashville, TN*

University of California, Santa Barbara — Aug. 2019 – Jun 2023
B.S. Statistics and Data Science, Overall GPA: 3.88/4.0, Major GPA: 3.92/4.0 — *Goleta, CA*

Research Experience

Network and Data Science (NDS) Lab — Aug. 2023 – Present
Graduate Research Assistant — *Vanderbilt University*
- Research Interests: AI for Biochemistry, Data-driven ML, Deep Generative Model

Geometric Intelligence Lab — Jan. 2023 – Jun. 2023
Undergraduate Researcher — *UC, Santa Barbara*
- Research Interests: Geometric Machine Learning, Manifold Learning

Caves Lab/Data Science Capstone — Nov. 2022 – Jun. 2023
Undergraduate Research Assistant — *UC, Santa Barbara*
- Research Interests: Image Processing, Computer Vision

Math Directed Reading Program — Jan. 2022 – Jun. 2022
Mentee — *UC, Santa Barbara*
- Research Interests: Universal Approximation Theory

Publications

1. Liu, Yunchao (Lance)*, Ha Dong*, Xin Wang*, Rocco Moretti, Yu Wang, Zhaoqian Su, Jiawei Gu, Bobby Bodenheimer, Charles David Weaver, Jens Meiler, and Tyler Derr. "WelQrate: Defining the Gold Standard in Small Molecule Drug Discovery Benchmarking." In Proceedings of the Neural Information Processing Systems Conference, Datasets and Benchmarks Track (NeurIPS '24).

2. Wang Yu, Nedim Lipka, Ruiyi Zhang, Alexa Siu, Yuying Zhao, Bo Ni, Xin Wang, Ryan Rossi, and Tyler Derr. "Topology-aware Retrieval Augmentation for Text Generation." In Proceedings of the 33rd ACM International Conference on Information and Knowledge Management (CIKM '24), 2442–2452.

Projects

WelQrate: Defining the Gold Standard in Small Molecule Drug Discovery
- Core maintainer of the WelQrate python package
- Main contributor of WelQrate's data collection, evaluation, and benchmarking framework

BioML Challenge 2024: Bits to Binders — *Report*
- Designed binding domains for a Chimeric Antigen Receptor (CAR) to bind the extracellular region of cancer antigen CD20
- Built a binder design pipeline with SOTA generative models (e.g. RFDiffusion, Chai)

Scaffold-aware Augmentation for Molecular Property Prediction via Diffusion Model — *Report*
- Examined the out-of-distribution issue of the molecular datasets split by scaffolds
- Designed a novel learning framework that augments molecules with the awareness of diverse scaffolds, enhancing the robustness and generalizability of the GNNs

Learning Molecules as Cellular Complexes — *Report*
- Built up a data processing pipeline to transform molecules from SMILES strings to cellular complexes
- Rebuilt Cell Attention Network and its training framework in the TopoModelX package for molecular datasets

Identifying and Measuring Visual Acuity Features from 2D Bee Images — *Poster* | *Code*

- Part of the NSF-funded Big Bee Project
- Developed a highly automatic pipeline to measure the diameter of the ommatidia and the interommatidial angles on the eye's surface from 2D images to estimate the bee's visual acuity
- Applied the pipeline to Apidae family and derived patterns with ecological and biological meanings

Proving Universal Approximation Theorem *Poster*

- Conducted a throughout study on Moshe Leshno et al. *"Multilayer feedforward networks with a nonpolynomial activation function can approximate any function"*

Experience

Summer Research Assistant Jun. 2024 – Aug. 2024

Vanderbilt University, Computer Science Department *Nashville, TN*

- Improved graph anomaly detection by augmenting local topology via diffusion model
- Assisted writing the proposal of Amazon Research Awards: *Generative AI for Graph Anomaly Detection*

Undergraduate Teaching Assistant Sep. 2022 – Jun. 2023

UC, Santa Barbara, PSTAT Department *Goleta, CA*

- Assisted teaching in Statistical Machine Learning, Regression Analysis, and Big Data Analytics courses
- Held lab/office hours to help students with homework and coding problems

Algorithm Engineer Intern Jun. 2022 – Aug. 2022

PING AN TECH, Intelligent City Group *Shenzhen, China*

- Mined and analyzed monthly macro-data of cities and trained an LSTM-based model
- Cleaned the Policy Database used for the Policy Recommendation System

Algorithm Intern Jun. 2021 – Aug. 2021

BOSERA FUND, Index and Quantitative Department *Shenzhen, China*

- Collected and modeled the quantitative data on daily stock transactions

Honors & Awards

- **Vanderbilt Graduate School Conference Travel Grant**
- **Vanderbilt Graduate Fellowship**
- **UCSB's Arts and Sciences: Graduation with College Honors**
- **UCSB Math DRP 2022: People Choice Award**

Service

- **Subreviewer**: CIKM 2024
- **PC Member**: WSDM2024-MLoG, GTA3-2024

Technical Skills

Programming Languages: Python, R, C/C++, SAS, SQL
Libraries/Softwares: PyTorch, PyTorch-Geometric, PyMOL, AutoDock, AmberTools,
Relevant Coursework: Computational Structural Biochemistry, Graph ML, Deep Learning, Geometric ML, Advanced ML, Representation Learning, Stochastic Process, Statistical Computing, Real Analysis