

# Yifei Wang

+1-781-354-5091 | [yifeiwang@brandeis.edu](mailto:yifeiwang@brandeis.edu) | [Homepage](#) | [in LinkedIn](#) | [GitHub](#) | [Google Scholar](#) | Waltham, MA, USA

## RESEARCH INTERESTS

Graph Neural Networks, Multimodal Learning, Generative AI, Large Language Model, Fairness, AI4Science, Healthcare Informatics.

## EDUCATION

### • Brandeis University

Jan 2020 - Current

*Ph.D. student in Computer Science (will master out)*

Massachusetts, USA

- GPA: 4.0/4.0. GSAS Fellowship, Travel and Research Grant.
- Research Focus: Graph Neural Network, Multimodal Learning, Fairness, AI4Science, AI4Healthcare, Generative AI.
- Core Courses: Statistical Machine Learning, Natural Language Processing, Learning on Graphs, Operation System.

### • Tsinghua University

Aug 2015 - Jul 2019

*Bachelor of Pure and Applied Mathematics* advised by Prof. Chenglong Bao and Prof. Chunping Li.

Beijing, China

- GPA: 3.6/4.0. Top 10% Undergraduate Thesis.
- Major Focus: Mathematics, Statistics, Optimization.
- Thesis: An Improved Spectral Clustering Method by Enhancing Sparsity.
- Core Courses: Statistical Inference, Numerical Analysis, Algorithm Analysis and Design, Convex Optimization.

## EXPERIENCE

### • Brigham and Woman's Hospital & Harvard Medical School

Aug 2022 - current

*Research Trainee* part-time advised by Li Zhou, Ph.D., M.D. and Liqin Wang, Ph.D.

Boston, USA

- Studied racial bias in the healthcare system using EHR data.
- Proposed a counterpart-based fairness metric for mortality prediction models.
- Current working on adapting large language model to special medical needs.

### • XtalPi Inc.

July 2019 - Dec 2019

*Algorithm Intern for AI-aided drug discovery*

Shenzhen, China

- Implemented a transformer-like architecture for retrosynthetic reaction path prediction.
- Improved model architectures and training strategies. Refined the pipeline of reaction prediction.

## PUBLICATIONS AND PREPRINTS

- [1] Wang, Yifei, Liqin Wang, Zhengyang Zhou, John Laurentiev, Joshua R. Lakin, Li Zhou, and Pengyu Hong. "Assessing fairness in machine learning models: A study of racial bias using matched counterparts in mortality prediction for patients with chronic diseases." *Journal of Biomedical Informatics* (2024): 104677.
- [2] Wang, Yifei, Shiyang Chen, Guobin Chen, Ethan Shurberg, Hang Liu, and Pengyu Hong. "Motif-Based Graph Representation Learning with Application to Chemical Molecules." *Informatics*, vol. 10, no. 1, p. 8. MDPI, 2023.
- [3] Yang, Tong\*, Yifei Wang\*, Long Sha\*, Jan Engelbrecht, and Pengyu Hong. "Knowledgebra: An Algebraic Learning Framework for Knowledge Graph." *Machine Learning and Knowledge Extraction* 4, no. 2 (2022): 432-445.
- [4] Li, Peizhao, Yifei Wang, Han Zhao, Pengyu Hong, and Hongfu Liu. "On dyadic fairness: Exploring and mitigating bias in graph connections." In *International Conference on Learning Representations (ICLR)*, 2021.
- [5] Hao Xu\*, Yifei Wang\*, Yunrui Li\*, Pengyu Hong. "Asymmetric Contrastive Multimodal Learning for Advancing Chemical Understanding" ([preprint](#)).
- [6] Wang, Yifei\*, Zhengyang Zhou\*, Liqin Wang, John Laurentiev, Peter Hou, Li Zhou, Pengyu Hong. "Counterpart Fairness – Addressing Systematic Between-group Differences in Fairness Evaluation" ([preprint](#))
- [7] Du, Xinsong, Zhengyang Zhou, Yifei Wang, Ya-Wen Chuang, Li Zhou et al. "Generative Large Language Models in Electronic Health Records for Patient Care Since 2023: A Systematic Review." ([preprint](#))
- [8] Wang, Yifei, Hao Peng, Long Sha, Zheyuan Liu, and Pengyu Hong. "State-level COVID-19 Trend Forecasting Using Mobility and Policy Data." ([preprint](#))

\*Equal contribution.

## PROJECTS

---

### View A: Graph Representational Learning, Multimodal Learning, AI for drug discovery.

- **Project 1: An Algebraic Learning Framework for Knowledge Graph** [\[Paper\]](#) [\[Code\]](#)
  - Developed an algebraic framework for learning consistent relation embeddings in knowledge graphs and proposed an algebraic-based instantiation for a knowledge graph embedding model.
  - Proposed simplified variants for learning low-dimensional relation embeddings using shared block settings and regularization-based method to integrate human logic rules.
- **Project 2: Motif-based Graph Representation Learning on Molecules** [\[Paper\]](#) [\[Code\]](#)
  - Designed a novel convolution module for graph representational learning on molecules with an efficient pretraining strategy, enabling the capture of local structural and semantic information from graph motifs.
  - Demonstrated the superiority of the proposed model in molecular learning across multiple dimensions: (1) better results in various molecular property prediction tasks, (2) more stable training, and (3) improved interpretability.
- **Project 3: Explainable Contrastive Multimodal Learning on Molecules** [\[Paper\]](#) [\[Code\]](#)
  - Proposed Asymmetric Contrastive Multimodal Learning, an effective and training-efficient framework tailored for molecules, promoting cross-modality understanding between the molecular graph and other chemical modalities.
  - Demonstrated the effectiveness and interpretability of the proposed framework in various key tasks: (1) large-scale cross-modality retrieval, (2) isomer discrimination, and (3) molecular property prediction.
- **Project 4: Controllable Molecule Generation by Diffusion Language Model** [\[Code\]](#)
  - Proposed a molecule generation framework inspired by the diffusion language model in controllable text generation, showing its superiority in discovering novel and diverse drug molecules.
  - Implemented a light-weight and modular plug-and-play approach to control the generation process, enabling the generation of molecules with optimized chemical properties and controlled substructures.

### View B: Fairness and Bias.

- **Project 5: Exploring and Mitigating Bias in Graph Connections** [\[Paper\]](#)
  - Investigated the disparity of graph embedding for link prediction bridging on intra- and inter-sensitive groups. Proposed a chain of theoretical analyses on how graph connections could affect dyadic fairness.
  - Accordingly introduced an algorithm for fair link prediction by adjusting the adjacency weight matrix to address the fairness-utility trade-off.
- **Project 6: Unveiling Systematic Differences in Group Fairness Evaluation in Medical Contexts** [\[Paper\]](#) [\[Code\]](#)
  - Explored the influence of systematic differences and confounding effects on group fairness estimation and proposed a counterpart-based fairness (CFair) index to improve fairness estimation by mitigating these influences.
  - Studied long-term mortality prediction of patients with chronic diseases, utilizing CFair to assess racial disparities in model predictions and offering insights into the influence of racial bias on model performance.

## SKILLS

---

- **Programming Languages:** Python (familiar with most deep learning packages and platforms such as Sklearn, Pytorch, Diffusion, Transformers, Peft, HuggingFace, etc.), MATLAB, R.
- **Languages:** English (proficient), Chinese (native).
- **Research Skills:** Algorithm Development, Theoretical Proof, Quantitative Analysis, Experiment Design, Interdisciplinary collaboration, Presentation, Writing and Communication.
- **Soft Skills:** Time Management, Problem Solving, Project Management, Collaboration, Mentorship.

## SERVICES

---

- **Conference Reviewer:** NeurIPS, ICLR, ICML, COLING, KDD Workshop, AMIA Annual Symposium.
- **Teaching Assistant:** Computation Theory, Spring 2024 & Statistical Machine Learning, Fall 2020, 2021, 2022.
- **Undergraduate Thesis Advisor:** Supervised a thesis that earned the Kukin Moskowitz Prize for Excellence 2024.