YUHUA ZHANG

Email: zyuhua@umich.edu | Tel: 203-809-8086 | LinkedIn ID: Yuhua Zhang

SUMMARY

Yuhua Zhang is a final year Ph.D. candidate at the Biostatistics Department, University of Michigan, whose research focuses on

(1) **Statistical Modeling** of graph/network data using ML skills such as **Clustering**, **Bayesian Methods**, **Probabilistic Graphical Models**, and **Topic Models** with an emphasis on the identification of underlying communities and link prediction;

(2) **Causal Inference** using **Instrumental Variable Method** (Mendelian Randomization) in observational studies with application to exome sequencing data.

EDUCATION

University of Michigan, Ann Arbor, Michigan   *2019.9-2023.6(expected)*
Ph.D. in Biostatistics
Dissertation Topic: Statistical Modeling of Large-scale Network Data
Supervised by: Dr. Walter Dempsey & Dr. Sebastian Zoellner

University of Michigan, Ann Arbor, Michigan   *2017.9-2019.4*
M.S. in Biostatistics
Selected Coursework: **L**inear Regression, **S**tatistical Inference, **B**ayesian Modeling, **N**umerical Linear Algebra, **S**tatistical Computing, **S**tochastic Process, **L**inear Programming, **M**achine Learning

Shanghai Jiaotong University, Shanghai, China   *2013.9-2017.6*
B.S. in Biological Science

INTERNSHIP EXPERIENCE

Uber Technology Inc., San Francisco
Applied Scientist Intern   *2022.5 – 2022.8*

· Investigated the offline information consumption pattern of drivers on the mobile platform
· Proposed a new metric to quantify drivers' driving mode
· Implemented the ML ranking model to decide the order of relevant information shown to drivers

RESEARCH EXPERIENCE

University of Michigan, Ann Arbor   2018.9 – present

Statistical Modeling of Network Data (Thesis)
· Proposed novel statistical models to identify latent cluster labels of nodes in sparse networks based on the edge exchangeable model.
· Used variational inference methods to improve the scalability of inferential algorithms.
· Applied proposed methods to TalkLife data, an online peer-to-peer support network, to identify its users' communities.

Probabilistic TWAS Analysis to Identify Causal Genes of Complex Traits
· Proposed a new computational framework based on the instrumental variables method to investigate causal relationships between gene expressions and complex traits. The proposed model confers higher power compared to existing methods.

· Applied the method to eQTL data across 49 tissues from GTEx and GWAS summary statistics from 114 complex traits to identify candidate causal genes.

Identify Potential Risk Factors to Predict Long-term Disease Courses in Bipolar Disorder
· Constructed genetic risk markers in bipolar disorder using multiple state-of-art methods, and identified the best predictor regarding prediction accuracy.
· Analyzed the clinical data collected from 1,159 bipolar patients and investigated the effects of early stage personalities on the long term disease course using Poisson regression model.

PROFESSIONAL SKILLS SUMMARY

Computing Skills
· Computing language: Python, R, SQL
· Able to work under Linux environment

Statistical Modeling
· Machine Learning Models: Clustering, Regression, Decision Trees, Topic Modeling, etc.
· Statistical Inference: Hypothesis Testing, etc.
· Causal Inference in observational data: Instrumental Variable Method, Propensity Score, etc.

Language:
· Chinese (native), English (working)

RESEARCH INTERESTS

Probabilistic Graph Modeling, **C**ausal Inference in Observational Data, **V**ariational Inference, **A**pplication of ML Methods to Health Data, **B**ayesian Nonparametric Method

SELECTED PUBLICATIONS

Zhang, Y., Dempsey, W. (2022). Node-level community detection within edge exchangeable models for interaction processes. arxiv preprint, arxiv: 2208.08539. (JASA, major revision)

Ryan, K. A., Han, P., **Zhang, Y.**, Marshall, D. F., Yocum, A. K., McInnis, M. G., Zöllner, S. (2021). Stability of personality traits in bipolar disorder: Findings from a longitudinal cohort. Journal of Affective Disorders, 283, 1-10.

Zhang, Y., Quick, C., Yu, K., Barbeira, A., Luca, F., Pique-Regi, R., ... Wen, X. (2020). PTWAS: investigating tissue-relevant causal molecular mechanisms of complex traits using probabilistic TWAS analysis. Genome biology, 21(1), 1-26.

Hu, Y., Lu, Q., Liu, W., **Zhang, Y.**, Li, M., & Zhao, H. (2017). Joint modeling of genetically correlated diseases and functional annotations increases accuracy of polygenic risk prediction. PLoS genetics, 13(6), e1006836.

LEADERSHIP AND SERVICE

Department of Biostatistics, University of Michigan, Ann Arbor

- Co-founder and member of the DEI Student Committee, August 2020- April 2021

- Co-founder and member of the Student Research Conflict Committee, August 2020- April 2021

- Departmental DEI committee member, January 2019- June 2020