UNIVERSITY OF CENTRAL FLORIDA

COLLOQUIUM

Feature Selection via Knockoffs in the Presence of Unmeasured Confounding

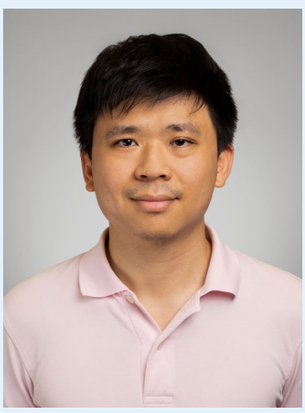
Speaker: Dr. Jiaqi Gu

Affiliation: University of South Florida

Abstract

Identifying features with causal effects is a fundamental problem in observational studies across biological, medical and social sciences. When there exists unmeasured confounding, existing causal effect estimation approaches only provide asymptotic error rate guarantees under specific parametric models and are vulnerable to model misspecification. To address this limitation, we propose a framework for causal feature selection using knockoffs. Specifically, we build our methods upon the concept of backdoor blockers to eliminate confounding effects and develop strategies to construct synthetic backdoor blockers leveraging non-causal features (either provided in a priori or learned from data). Within the framework of knockoffs, we adopt the splitting-and-recycling approach to boost statistical power with provable false discovery rate (FDR) control. In addition, as no parametric assumption is imposed on the causal effects, our methods are robust to model misspecification. Simulations show that our methods can provide valid FDR control and powerfully identify causal features. In analysis of a single-cell RNA sequencing data, we identify 154 causal genes of Alzheimer's disease, many of which are supported by existing literature.
This is joint work with Zihuai He at Stanford University.

Speaker Bio



Dr. Jiaqi Gu is an Assistant Professor in the Department of Mathematics and Statistics at the University of South Florida. He earned his Ph.D. in Statistics from the University of Hong Kong under the supervision of Philip L.H. Yu and Guosheng Yin, and his bachelor's degree in Statistics from Renmin University of China. He subsequently completed postdoctoral training in the Department of Neurology and Neurological Sciences at Stanford University under the mentorship of Zihuai He.

His research interests include statistical genetics, causal inference, high-dimensional statistical inference, Bayesian methods, public health data analysis, survival analysis, preference learning, and network data modeling.

Event Details

- **Date:** Monday, November 16, 2026
- **Time:** 3:30 PM – 4:30 PM
- **Location:** MSB 318

Refreshments will be served from 2:30pm–3:30pm in MSB 201

Dr. Yvette Kanouff School of Data, Mathematical, and Statistical Sciences