Frontiers of Statistical Network Analysis: Inference, Tensors and Beyond

12 May 2025–30 May 2025
National University of Singapore
Institute for Mathematical Sciences

ORGANIZERS

- Jialiang Li (National University of Singapore)
- Dong Xia (Hong Kong University of Science and Technology)
- Yuan Zhang (Ohio State University)

OVERVIEW

Network data present unique structures and distinct analytical challenges. Earlier efforts in statistical network analysis focused on modeling and parameter estimation. The past decade, known as a “golden age” for this field, has witnessed a surge in innovative methods and theoretical developments. The recent boom in data science also gives rise to numerous categories of complex networks where interactions among a set of entities are polyadic or non-linear, which are collectively referred to as tensors, or higher-order networks. Moreover, we witness the increasingly interdisciplinary collaborations between network analysis and applied fields. The goal of this workshop is three-fold: to advance the field by exchanging ideas and deepening understanding; to facilitate future research by discussing challenges and establishing collaboration opportunities; and to benefit the local Singaporean audience and other participants through lectures from leading experts and emerging star scholars.

Book of Abstracts

Speakers

Aaron Potechin (University of Chicago) - 4 -
Annie Qu (University of California Irvine) - 4 -
Anru Zhang (Duke University) - 4 -
Bhaswar B Bhattacharya (University of Pennsylvania) - 5 -
Chenlei Leng (University of Warwick) - 5 -
Christophe Giraud (Université Paris Saclay) - 5 -
David S Choi (Carnegie Mellon University) - 6-
Dong Xia (Hong Kong University of Science and Technology) - 6 -
Emma Jingfei Zhang (Emory University) - 6 -
Garvesh Raskutti (University of Wisconsin-Madison) - 7 -
Harrison Zhou (Yale University) - 7 -
Ji Zhu (University of Michigan) - 7 -
Jinchi Lv (University of Southern California) - 8 -
Jingnan Zhang (University of Science and Technology of China) - 8 -
Jinyuan Chang (Southwestern University of Finance and Economics) - 8 -
Junhui Wang (Chinese University of Hong Kong) - 9 -
Kengo Kato (Cornell University) - 9 -
Lexin Li (University of California, Berkeley) - 10 -
Miaoyan Wang (University of Wisconsin-Madison) - 10 -
Michael Schweinberger (The Pennsylvania State University) - 10 -
Mladen Kolar (University of Southern California) - 11 -
Niels Richard Hansen (University of Copenhagen) - 11 -
Qiwei Yao (London School of Economics) - 12 -
Subhroshekhar Ghosh (National University of Singapore) - 12 -
Wei-Yin Loh (University of Wisconsin-Madison) - 12 -
Wen Zhou (New York University) - 13 -
Will Wei Sun (Purdue University) - 13 -
Yang Feng (New York University) - 13 -
Yingying Fan (University of Southern California) - 14 -
Yuan Zhang (Ohio State University) - 14 -
Yuguo Chen (University of Illinois at Urbana-Champaign) - 14 -
Yuxin Chen (University of Pennsylvania) - 15 -

Aaron Potechin (University of Chicago)

Title: Graph Matrices and Tensor Networks

Abstract: Graph matrices are a type of matrix which plays a key role in proving sum of squares lower bounds on average case problems and is a powerful tool for analyzing problems on random inputs. In this talk, I will describe graph matrices and what is known about them, illustrate how they can be used, and describe how they are related to tensor networks.

Annie Qu (University of California Irvine)

Title: Representation Retrieval Learning for Heterogeneous Data Integration

Abstract: In this presentation, I will showcase advanced statistical machine learning techniques and tools designed for the seamless integration of information from multi-source datasets. These datasets may originate from various sources, encompass distinct studies with different variables, and exhibit unique dependent structures. One of the greatest challenges in investigating research findings is the systematic heterogeneity across individuals, which could significantly undermine the power of existing machine learning methods to identify the underlying true signals. This talk will investigate the advantages and drawbacks of current data integration methods such as multi-task learning, optimal transport, missing data imputations, matrix completions and transfer learning. Additionally, we will introduce a new representation retriever learning aimed at mapping heterogeneous observed data to a latent space, facilitating the extraction of shared information and knowledge, and disentanglement of source-specific information and knowledge. The key idea is to project heterogeneous raw observations to representation retriever library, and the novelty of our method is that we can retrieve partial representations from the library for a target study. The main advantages of the proposed method are that it can increase statistical power through borrowing partially shared representation retrievers from multiple sources of data. This approach ultimately allows one to extract information from heterogeneous data sources and transfer generalizable knowledge beyond observed data and enhance the accuracy of prediction and statistical inference.

Anru Zhang (Duke University)

Title: Theoretical Guarantees for Alternative Least Square Algorithm in Tensor CP Decomposition

Abstract: We introduce a statistical and computational framework for tensor Canonical Polyadic (CP) decomposition, with a focus on statistical theory, convergence, and algorithmic improvements. First, we show that the Alternating Least Squares (ALS)algorithm achieves the desired error rate within three iterations when $R = 1$. Second, for the more general case where $R > 1$, we derive statistical bounds for ALS, showing that the estimation error exhibits an initial phase of quadratic convergence followed by linear convergence until reaching the desired accuracy. Third, we propose a novel warm-start procedure for ALS in the $R > 1$ setting, which integrates tensor Tucker decomposition with simultaneous diagonalization (Jennrich's algorithm) to significantly enhance performance over existing benchmark methods. Numerical experiments support our theoretical findings, demonstrating the practical advantages of our approach.

Bhaswar B Bhattacharya (University of Pennsylvania)

Title: Higher-Order Graphon Theory: Fluctuations, Inference, and Degeneracies

Abstract: Motifs (patterns of subgraphs), such as edges and triangles, encode important structural information about the geometry of a network. Consequently, counting motifs in a large network is an important statistical and computational problem. In this talk we will consider the problem of estimating motif densities and fluctuations of subgraph counts in an inhomogeneous random graph sampled from a graphon. We will show that the limiting distributions of subgraph counts can be Gaussian or non-Gaussian, depending on a notion of regularity of subgraphs with respect to the graphon. Using these results and a novel multiplier bootstrap for graphons, we will construct joint confidence sets for the motif densities. Finally, we will discuss various structure theorems and open questions about degeneracies of the limiting distribution and connections to quasirandom graphs.

Joint work with Anirban Chatterjee, Soham Dan, and Svante Janson.

Chenlei Leng (University of Warwick)

Title: Statistical Analysis of Reciprocity

Abstract: Asymmetric relational data is increasingly prevalent across diverse fields, highlighting the need for directed network models to address the unique challenges posed by their complex structures. Unlike undirected models, directed models can explicitly capture reciprocity—the tendency of nodes to form mutual links. In this talk, we will progressively develop a series of models to examine reciprocity in directed networks. We begin with a simple extension of the Erdős–Rényi model to incorporate reciprocity, addressing the fundamental question of effective sample sizes for reciprocal and non-reciprocal effects. Next, we introduce covariates to refine the modelling of reciprocity. Finally, we present a model that accommodates both node-specific heterogeneity and reciprocity, achieving a fully general framework for modelling reciprocal relationships in directed networks.

Christophe Giraud (Université Paris Saclay)

Title: learning latent features from network data

Abstract. Network structures are often shaped by some historical or latent features of the nodes. For example, communities drive the network properties in Stochastic Block Model, latent features determine the connection patterns in Graphons, and arrival time influences the local properties in random recursive trees. In this talk, we will discuss how to learn these features in various learning frameworks, where probabilities of connection fulfill some shape constraints, should it be tree constraints, or graphon's shape constraints.

David S Choi (Carnegie Mellon University)

Title: Agnostic Characterization of Interference in Randomized Experiments

bstract: In social network settings, the analysis of randomized experiments may be nontrivial due to the presence of interference between units, through diverse mechanisms such as peer influence, transmission of disease, market competition, and sharing of information. We give an approach for characterizing interference by lower bounding the number of units whose outcome depends on selected groups of treated individuals, such as depending on the treatment of others, or others who are at least a certain distance away. The approach is applicable to randomized experiments with binary-valued outcomes. Asymptotically conservative point estimates and one-sided confidence intervals may be constructed with no assumptions beyond the known randomization design, allowing the approach to be used when interference is poorly understood, or when an observed network might only be a crude proxy for the underlying social mechanisms. Point estimates are equal to Hajek-weighted comparisons of units with differing levels of treatment exposure. Empirically, we find that the width of our interval estimates is competitive with (and often smaller than)those of the EATE, an assumption-lean treatment effect, suggesting that the proposed estimands may be intrinsically easier to estimate than treatment effects.

Dong Xia (Hong Kong University of Science and Technology)

Title: Online Decision Making: Algorithm, Regret, Constraints and Uncertainty

Abstract:

We will discuss online decision-making problems in scenarios where covariates are high-dimensional or personalized covariates are unavailable. Our focus is on the ϵ-greedy algorithm for decision making and online gradient descent for estimating model parameters. By carefully balancing exploration and exploitation, we achieve a trade-off between regret performance and estimation accuracy. Additionally, we explore online decision-making under constraints (such as knapsack problems)within a primal-dual framework, demonstrating that sublinear regret is achievable. Finally, we propose an online debiasing approach based on inverse propensity weighting (IPW)for uncertainty quantification. Real data examples will also be discussed.

Emma Jingfei Zhang (Emory University)

Title: Modeling Non-Uniform Hypergraphs Using Determinantal Point Processes

Abstract: Most statistical models for networks focus on pairwise interactions between nodes. However, many real-world networks feature higher-order interactions involving multiple nodes, such as co-authors collaborating on a paper. Hypergraphs provide a natural representation for these networks, with each hyperedge representing a set of nodes. The majority of existing hypergraph models assume uniform hyperedges, that is, edges are of the same size, or are driven by diversity amongst nodes. In this work, we propose a new hypergraph model formulated based on non-symmetric determinantal point processes. The proposed model naturally accommodates non-uniform hyperedges, has tractable probability mass functions, and allows for node similarity or diversity in hyperedges. For model estimation, we maximize the likelihood function under constraints via a computationally efficient projected adaptive gradient descent algorithm and establish the consistency and asymptotic normality of the estimator. Simulation studies confirm the efficacy of the proposed model, and its utility is further demonstrated through predictions on several real-world datasets.

Garvesh Raskutti (University of Wisconsin-Madison)

Title: Context-dependent and model-agnostic network estimation

Abstract: Social networks often present data in the form of events/posts by multiple people over time (e.g. tweets, posts, memes, e.t.c.).One of the underlying challenges is determining who is influencing whom and how the influence is occurring. In this talk, I address this question by posing this problem as learning a context-dependent network structure by posing this problem as convex optimization problem with a tensor network parameter. The model has connections to multinomial auto-regressive models and compositional time series. Our approach is validated with meme-tracker and political tweet data. The main novelty is the scalability, theoretical guarantees and connection to model-agnostic variable importance networks. Building on this connection to model-agnostic variable importance networks, I briefly touch on recent work on scalable variable importance estimation and present some open questions on how this may be connects back to influence network estimation in the model-agnostic setting.

Harrison Zhou (Yale University)

Title: From Score Estimation to Sampling

Abstract: Recent impressive advances in the algorithmic generation of high-fidelity images, audio, and video can be largely attributed to the success of score-based diffusion models. A crucial step in their implementation is score matching, which involves estimating the score function of the forward diffusion process from training data. In this work, we establish the rate-optimal estimation of the score function for smooth, compactly supported densities and explore its applications to estimation of density, transport, and optimal transport.

Ji Zhu (University of Michigan)

Title: Hyperbolic Network Latent Space Model with Learnable Curvature

Abstract: Network data is ubiquitous in various scientific disciplines, including sociology, economics, and neuroscience. Latent space models are often employed in network data analysis, but the geometric effect of latent space curvature remains a significant, unresolved issue. In this work, we propose a hyperbolic network latent space model with a learnable curvature parameter. We theoretically justify that learning the optimal curvature is essential to minimizing the embedding error across all hyperbolic embedding methods beyond network latent space models. A maximum-likelihood estimation strategy, employing manifold gradient optimization, is developed, and we establish the consistency and convergence rates for the maximum-likelihood estimators, both of which are technically challenging due to the non-linearity and non-convexity of the hyperbolic distance metric. We further demonstrate the geometric effect of latent space curvature and the superior performance of the proposed model through extensive simulation studies and an application using a Facebook friendship network.

Jinchi Lv (University of Southern California)

Title: "ATE-GL: Asymptotic Theory of Eigenvectors for Latent Embeddings with Generalized Laplacian Matrices"

Abstract:

Laplacian matrices are commonly employed in many real applications, encoding the underlying latent structural information such as graphs and manifolds. The use of the normalization terms naturally gives rise to random matrices with dependency. It is well-known that dependency is a major bottleneck of new random matrix theory (RMT)developments. To this end, in this paper we formally introduce a class of generalized (regularized) Laplacian matrices, which contains the Laplacian matrix and the random adjacency matrix as a specific case, and suggest the new framework of asymptotic theory of eigenvectors for latent embeddings with generalized Laplacian matrices (ATE-GL).Our new theory is empowered by the tool of generalized quadratic vector equation for dealing with RMT under dependency, and delicate high-order asymptotic expansions of the empirical spiked eigenvectors and eigenvalues based on local laws. The asymptotic normalities established for both spiked eigenvectors and eigenvalues will enable us precise inference and uncertainty quantification for applications involving the generalized Laplacian matrices with flexibility. We discuss some applications of the suggested ATE-GL framework and showcase its validity through some numerical examples. This is a joint work with Jianqing Fan, Yingying Fan, Fan Yang and Xin Diwen Yu.

Jingnan Zhang (University of Science and Technology of China)

Title: A dynamic network autoregressive model for time-varying network-link data

Abstract: Network-linked data, where different units are linked through a network has been extensively studied in literature. However, its extension, specifically time-varying network-link data, has received less investigation. Existing methods for time-varying network-link data only assume that units' attributes change over time, neglecting network evolution. To address this gap, we propose a dynamic network autoregressive model for time-varying network-link data, where both units' attributes and networks are allowed to vary over time. A tensor decomposition method is employed to provide low-dimensional embedding vectors, which are further used to reformulate the traditional network autoregressive model. Interestingly, node-embedding vectors are concentrated around some group centers but are not exactly the same within some groups. Meanwhile, both within-group and global homogeneities are considered for the effect of covariate vectors. To tackle the resultant optimization task, we develop the power update algorithm and an efficient alternative updating algorithm. Furthermore, the asymptotic consistencies of the proposed method are established, irrespective of the presence of the global effect of covariate vector. These consistencies are demonstrated by extensive simulated examples and a real example of time-varying network-linked fund data.

Jinyuan Chang (Southwestern University of Finance and Economics)

Title: Autoregressive Networks with Dependent Edges

Abstract: We propose an autoregressive framework for modelling dynamic networks with dependent edges. It encompasses models that accommodate, for example, transitivity, density-dependence and other stylized features often observed in real network data. By assuming the edges of networks at each

time are independent conditionally on their lagged values, the models, which exhibit a close connection with temporal ERGMs, facilitate both simulation and the maximum likelihood estimation in a straightforward manner. Due to the possibly large number of parameters in the models, the natural MLEs may suffer from slow convergence rates. An improved estimator for each component parameter is proposed based on an iteration employing projection, which mitigates the impact of the other parameters (Chang et al., 2021, 2023). Leveraging a martingale difference structure, the asymptotic distribution of the improved estimator is derived without the assumption of stationarity. The limiting distribution is not normal in general, although it reduces to normal when the underlying process satisfies some mixing conditions. Illustration with a transitivity model was carried out in both simulation and a real network data set.

Junhui Wang (Chinese University of Hong Kong)

Title: Learning nonparametric graphical model on heterogeneous network-linked data

Abstract: Graphical models have been popularly used for capturing conditional independence structure in multivariate data, which are often built upon independent and identically distributed observations, limiting their applicability to complex datasets such as network-linked data. In this talk, we introduce a nonparametric graphical model that addresses these limitations by accommodating heterogeneous graph structures without imposing any specific distributional assumptions. The introduced estimation method effectively integrates network embedding with nonparametric graphical model estimation. It further transforms the graph learning task into solving a finitedimensional linear equation system by leveraging the properties of vectorvalued reproducing kernel Hilbert space. We will also discuss theoretical properties of the proposed method in terms of the estimation consistency and exact recovery of the heterogeneous graph structures. Its effectiveness is also demonstrated through a variety of simulated examples and a real application to the statistician coauthorship dataset.

Kengo Kato (Cornell University)

Title: Limit Laws for Gromov-Wasserstein Alignment with Applications to Testing Graph Isomorphisms

Abstract: The Gromov-Wasserstein (GW) distance enables comparing metric measure spaces based solely on their internal structure, making it invariant to isomorphic transformations. This property is particularly useful for comparing datasets that naturally admit isomorphic representations, such as unlabelled graphs or objects embedded in space. However, apart from the recently derived empirical convergence rates for the quadratic GW problem, a statistical theory for valid estimation and inference remains largely obscure. Pushing the frontier of statistical GW further, this work derives the first limit laws for the empirical GW distance across several settings of interest: (i)~discrete,(ii)~semi-discrete,and (iii)~generaldistributions under moment constraints under the entropically regularized GW distance. The derivations rely on a novel stability analysis of the GW functional in the marginal distributions. The limit laws then follow by an adaptation of the functional delta method. As asymptotic normality fails to hold in most cases, we establish the consistency of an efficient estimation procedure for the limiting law in the discrete case, bypassing the need for computationally intensive resampling methods. We apply these findings to testing whether collections of unlabelled graphs are generated from distributions that are isomorphic to each other.

Lexin Li (University of California, Berkeley)

Title: Tensor Data Analysis and Some Applications in Neuroscience

Abstract: Multidimensional arrays, or tensors, are becoming increasingly prevalent in a wide range of scientific applications. In this talk, I will present two case studies from neuroscience, where tensor decomposition proves particularly useful. The first study is a cross-area neuronal spike trains analysis, which we formulate as the problem of regressing a multivariate point process on another multivariate point process. We model the predictor effects through the conditional intensities using a set of basis transferring functions in a convolutional fashion. We then organize the corresponding transferring coefficients in the form of a three-way tensor, and impose the low-rank, sparsity, and subgroup structures on this coefficient tensor. The second study is a multimodal neuroimaging analysis for Alzheimer's disease, which we formulate as the problem of modeling the correlations of two sets of variables conditioning on the third set of variables. We propose a generalized liquid association analysis method to study such three-way associations. We establish a population dimension reduction model, and transform the problem to sparse decomposition of a three-way tensor.

Miaoyan Wang (University of Wisconsin-Madison)

Title: Application and Methods for Structured Tensor Learning

Abstract: High-order tensor datasets pose common challenges in applications such as recommendation systems, neuroimaging, and social networks. In this work, we introduce two approaches for learning with structured tensors: block models for higher-order clustering and sign-series models for tensor denoising. These approaches provide lens into the unique properties of tensor analysis. We establish statistical and computational efficiency results for each method. Additionally, we present polynomial-time algorithms with guaranteed efficiency. The effectiveness of our methods is demonstrated through applications to neuroimaging data analysis and social network analysis.

Michael Schweinberger (The Pennsylvania State University)

Title: A regression framework for studying relationships among attributes under network interference

Abstract: To understand how the interconnected and interdependent world of the twenty-first century operates and make model-based predictions, joint probability models for networks and interdependent outcomes are needed. We propose a comprehensive regression framework for networks and interdependent outcomes with multiple advantages, including interpretability, scalability, and provable theoretical guarantees. The regression framework can be used for studying relationships among attributes of connected units and captures complex dependencies among connections and attributes, while retaining the virtues of linear regression, logistic regression, and other regression models by being interpretable and widely applicable. On the computational side, we show that the regression framework is amenable to scalable statistical computing based on convex optimization of pseudo-likelihoods using minorization-maximization methods. On the theoretical side, we establish convergence rates for pseudo-likelihood estimators based on a single observation of dependent connections and attributes. We

demonstrate the regression framework using simulations and an application to hate speech on the social media platform X in the six months preceding the insurrection at the U.S. Capitol on January 6, 2021.

Mladen Kolar (University of Southern California)

Title: Trans-Glasso: A Transfer Learning Approach to Precision Matrix Estimation

Abstract: Many real-world systems—ranging from gene regulatory interactions in biology to financial asset dependencies—can be represented by networks, whose edges correspond to conditional relationships among variables. These relationships are succinctly captured by the precision matrix of a multivariate distribution. Estimating the precision matrix is thus fundamental to uncovering the underlying network structure. However, this task can be challenging when the available data for the target domain are limited, undermining accurate inference.

In this talk, I will present Trans-Glasso, a novel two-step transfer learning framework for precision matrix estimation that leverages data from source studies to improve estimates in the target study. First, Trans-Glasso identifies shared and unique features across studies via a multi-task learning objective. Then, it refines these initial estimates through differential network estimation to account for structural differences between the target and source precision matrices. Assuming that most entries of the target precision matrix are shared with at least one source matrix, we derive non-asymptotic error bounds and show that Trans-Glasso achieves minimax optimality under certain conditions.

Through extensive simulations, Trans-Glasso demonstrates improved performance over standard methods, especially in small-sample settings. Applications to gene regulatory networks across multiple brain tissues and protein networks in various cancer subtypes confirm its practical effectiveness in biological contexts, where understanding network structures can provide insights into disease mechanisms and potential interventions. Beyond biology, these techniques are broadly applicable wherever precision matrix estimation and network inference play a crucial role, including neuroscience, finance, and social science.

This is joint work with Boxin Zhao and Cong Ma.

Niels Richard Hansen (University of Copenhagen)

Title: Identification and inference from cross-sectional data via higher order cumulants

Abstract: Drawing inference from single cell data about the dynamics of the chemical reactions in the cell is the epitome of the problem I will address in the talk. The cell is killed during measurement, and we are thus only able to see a snapshot of the cell constituents at a single timepoint. This is an example of cross-sectional data from a multivariate dynamical system, which is obtained with the purpose of inferring properties of the dynamics of the process. This is only possible by making modeling assumptions and/or by obtaining data under various perturbations. Within a framework of linear (non-

Gaussian) steady-state models, I will present a new characterization of all cumulant tensors as solutions of higher order Lyapunov equations. I will outline how this result can be used for practical estimation and inference with applications to single cell gene expression data. I will also present recent theoretical results on generic identifiability of model parameters related to the dynamics of the system from the third order cumulant tensor in the non-Gaussian setting.

Qiwei Yao (London School of Economics)

Title: Network Inference via Edge-Jittering

Abstract: We explore jittered bootstrap for various network inference problems including (i) edge differentially private estimation for beta-models, (ii) network model selection, and (iii) testing and interval estimation for the balance of signed networks.

Subhroshekhar Ghosh (National University of Singapore)

Title: Learning with latent group sparsity via diffusions on networks

Abstract: Group or cluster structure on explanatory variables in machine learning problems is a very general phenomenon, which has attracted broad interest from practitioners and theoreticians alike. In this work we contribute an approach to learning under such group structure, that does not require prior information on the group identities. Our paradigm is motivated by the Laplacian geometry of an underlying network with a related community structure, and proceeds by directly incorporating this into a penalty that is effectively computed via a heat flow-based local network dynamics. In fact, we demonstrate a procedure to construct such a network based on the available data. Notably, we dispense with computationally intensive pre-processing involving clustering of variables, spectral or otherwise. Our technique is underpinned by rigorous theorems that guarantee its effective performance and provide bounds on its sample complexity. In particular, in a wide range of settings, it provably suffices to run the diffusion for time that is only logarithmic in the problem dimensions. We explore in detail the interfaces of our approach with key statistical physics models in network science, such as the Gaussian Free Field and the Stochastic Block Model. We validate our approach by successful applications to real-world data from a wide array of application domains, including computer science, genetics, climatology and economics. Our work raises the possibility of applying similar diffusion-based techniques to classical learning tasks, exploiting the interplay between geometric, dynamical and stochastic structures underlying the data.

Wei-Yin Loh (University of Wisconsin-Madison)

Title: A regression tree approach to missing data and explainable AI

Abstract: The problem of dealing with missing values in data is arguably the most difficult one in statistics. Although imputation is a popular solution, there are everyday situations where imputation makes no sense. Worse yet, unless the variables are well understood, it may be impossible to know whether imputation makes sense or not. In the first part of this talk, we introduce a regression tree algorithm called GUIDE that can fit classification and regression models to data without requiring imputation of missing values in the

predictor variables. Unlike all other regression tree methods that perform imputation implicitly, GUIDE will identify the variables with informative missing values by highlighting them explicitly in the tree structures. In the second part of the talk, we show how GUIDE can be used to explain in simple terms the predictions of any machine learning model.

Wen Zhou (New York University)

Title: Nonparametric inference on network effects with dependent edges: optimality, two-sample, multiple strata

Abstract: Testing network effects in weighted directed networks is a foundational problem in econometrics, sociology, and psychology. Yet, the prevalent edge dependency poses a significant methodological challenge. Most existing methods are model-based and come with stringent assumptions, limiting their applicability. In response, we introduce a novel, fully nonparametric framework that requires only minimal regularity assumptions. While inspired by recent developments in U-statistic literature, our approach notably broadens their scopes. Specifically, we identified and carefully addressed the challenge of indeterminate degeneracy in the test statistics -- a problem that aforementioned tools do not handle. We established Berry-Esseen type bounds for the accuracy of type-I error rate control. With original analysis, we also proved the minimax power optimality of our test. Simulations underscore the superiority of our method in computation speed, accuracy, and numerical robustness compared to competing methods.

Will Wei Sun (Purdue University)

Title: Online Statistical Inference for Low-Rank Reinforcement Learning

Abstract: Reinforcement learning (RL) enables intelligent agents to make data-driven decisions in uncertain environments by leveraging contextual information to maximize cumulative rewards. Modern applications often involve high-dimensional tensor contexts, requiring low-rank structures for sample efficient RL models. While most RL algorithms focus on minimizing regret or selecting actions based on oracle policies, statistical inference for adaptively collected RL data remains underexplored. Such inference is crucial in domains like personalized medicine, mobile health, and autonomous driving, where understanding the statistical uncertainty of policy evaluations is essential. This talk presents online inferential tools designed for low-rank RL models, providing provable measures of uncertainty for safer and more reliable decision-making.

Yang Feng (New York University)

Title: Semiparametric Modeling and Analysis for Longitudinal Network Data

Abstract: We introduce a semiparametric latent space model for analyzing longitudinal network data. The model consists of a static latent space component and a time-varying node-specific baseline component. We develop a semiparametric efficient score equation for the latent space parameter by adjusting for the baseline nuisance component. Estimation is accomplished through a one-step update estimator and an appropriately penalized maximum likelihood estimator. We derive oracle error bounds for the two

estimators and address identifiability concerns from a quotient manifold perspective. Our approach is demonstrated using the New York Citi Bike Dataset.

Yingying Fan (University of Southern California)

Title: HNCI: high-dimensional network causal inference

Abstract: The problem of evaluating the effectiveness of a treatment or policy commonly appears in causal inference applications under network interference. In this paper, we suggest the new method of high-dimensional network causal inference (HNCI)that provides both valid confidence interval on the average direct treatment effect on the treated (ADET)and valid confidence set for the neighborhood size for interference effect. We exploit the model setting in Belloni et al. (2022)and allow certain type of heterogeneity in node interference neighborhood sizes. We propose a linear regression formulation of potential outcomes, where the regression coefficients correspond to the underlying true interference function values of nodes and exhibit a latent homogeneous structure. Such a formulation allows us to leverage existing literature from linear regression and homogeneity pursuit to conduct valid statistical inferences with theoretical guarantees. The resulting confidence intervals for the ADET are formally justified through asymptotic normalities with estimable variances. We further provide the confidence set for the neighborhood size with theoretical guarantees exploiting the repro samples approach. The practical utilities of the newly suggested methods are demonstrated through simulation and real data examples.

Yuan Zhang (Ohio State University)

Title: Higher-order accurate two-sample network inference and network hashing

Abstract: Two-sample hypothesis testing for network comparison presents many significant challenges, including: leveraging repeated network observations and known node registration, but without requiring them to operate; relaxing strong structural assumptions; achieving finite-sample higher-order accuracy; handling different network sizes and sparsity levels; fast computation and memory parsimony; controlling false discovery rate (FDR)in multiple testing; and theoretical understandings, particularly regarding finite-sample accuracy and minimax optimality. In this paper, we develop a comprehensive toolbox, featuring a novel main method and its variants, all accompanied by strong theoretical guarantees, to address these challenges. Our method outperforms existing tools in speed and accuracy, and it is proved power-optimal. Our algorithms are user-friendly and versatile in handling various data structures (single or repeated network observations; known or unknown node registration).We also develop an innovative framework for offline hashing and fast querying as a very useful tool for large network databases. We showcase the effectiveness of our method through comprehensive simulations and applications to two real-world datasets, which revealed intriguing new structures.

Yuguo Chen (University of Illinois at Urbana-Champaign)

Title: Subsampling Based Inference for Large Networks

Abstract: Large networks are increasingly prevalent in many scientific applications. Statistical analysis of such large networks become prohibitive due to exorbitant computation cost and high memory requirements. We develop a subsampling based divide-and-conquer algorithm for statistical inference in large networks. This method saves both memory and computation costs significantly as one needs to store and process only the smaller subnetworks. This method is also parallelizable which makes it even faster. We derive theoretical properties of the algorithm and demonstrate the effectiveness of the algorithm on simulated and real networks.

Yuxin Chen (University of Pennsylvania)

Title: Heteroskedastic Tensor Clustering

Abstract: Tensor clustering, which seeks to extract underlying cluster structures from noisy tensor observations, has gained increasing attention. One extensively studied model for tensor clustering is the tensor block model, which postulates the existence of clustering structures along each mode and has found broad applications in areas like multi-tissue gene expression analysis and multilayer network analysis. However, currently available computationally feasible methods for tensor clustering either are limited to handling i.i.d. sub-Gaussian noise or suffer from suboptimal statistical performance, which restrains their utility in applications that have to deal with heteroskedastic data and/or low signal-to-noise-ratio (SNR).To overcome these challenges, we propose a two-stage method, named High-order HeteroClustering (HHC), which starts by performing tensor subspace estimation via a novel spectral algorithm called Thresholded Deflated-HeteroPCA, followed by approximate k-means to obtain cluster nodes. Encouragingly, our algorithm provably achieves exact clustering as long as the SNR exceeds the computational limit (ignoring logarithmic factors);here, the SNR refers to the ratio of the pairwise disparity between nodes to the noise level, and the computational limit indicates the lowest SNR that enables exact clustering with polynomial runtime. Comprehensive simulation and real-data experiments suggest that our algorithm outperforms existing algorithms across various settings, delivering more reliable clustering performance.