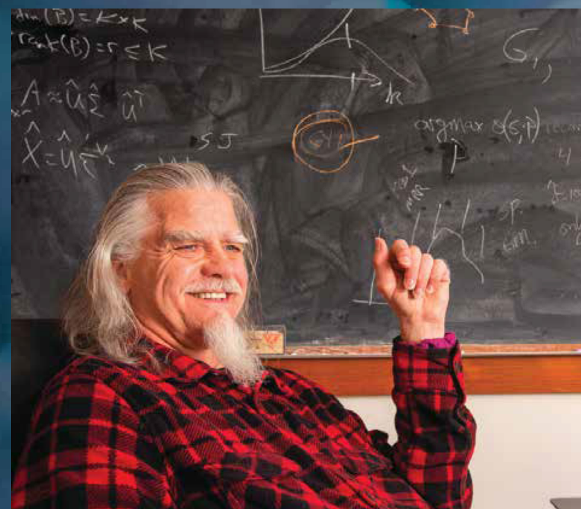
BRIN MATHEMATICS RESEARCH CENTER

Statistical Inference on Networks and High-Dimensional Data

October 18–20, 2023

About the Workshop

The analysis of network-structured and high-dimensional data plays a critical role in many disciplines across the social and natural sciences. The past several years have seen rapid advancements in statistical methodology for such data, spanning classical statistical inference, such as testing and estimation, and modern machine learning, such as neural networks, information retrieval, and prediction. Join us in recognizing Professor Carey E. Priebe's distinguished contributions to statistics on networks at this workshop in honor of his 60th birthday, where colleagues, students, and researchers will meet, collaborate, and tackle important new problems in the field.



A workshop held in honor of Professor Carey E. Priebe's 60th birthday

Organizers

Avanti Athreya, Johns Hopkins University
Vince Lyzinski, University of Maryland
Minh Tang, North Carolina State University

Speakers

Joshua Agterberg, University of Pennsylvania
Jesus Arroyo-Relion, Texas A&M University
Joshua Cape, University of Wisconsin-Madison
Nathaniel Josephs, North Carolina State University
Zheng Tracy Ke, Harvard University
Keith Levin, University of Wisconsin-Madison
Liza Levina, University of Michigan
Tianxi Li, University of Minnesota
Lizhen Lin, University of Maryland
Zachary Lubberts, University of Virginia
Subhadeep Paul, The Ohio State University
Marianna Pensky, University of Central Florida
Karl Rohe, University of Wisconsin-Madison
Patrick Rubin-Delanchy, University of Bristol
Purnamrita Sarkar, University of Texas at Austin
Jonathan Stewart, Florida State University
Daniel Sussman, Boston University
Michael Trosset, Indiana University Bloomington
Joshua Vogelstein, Johns Hopkins University
Harrison Huibin Zhou, Yale University

CSIS Building 4th Floor
8169 Paint Branch Drive
University of Maryland
College Park, MD 20742

brinmrc.umd.edu

DEPARTMENT OF MATHEMATICS

Contents

Workshop Overview

The analysis of network-structured and high-dimensional data plays a critical role across many disciplines across the social and natural sciences. The past several years have seen rapid advancements in statistical methodology for such data, spanning classical statistical inference (such as testing and estimation) and modern machine learning (such as neural networks, information retrieval, and prediction). Join us in recognizing Professor Carey E. Priebe's long and distinguished contributions to statistics on networks at this workshop in honor of his 60th birthday, where colleagues, students, and other researchers will meet, collaborate, and tackle important new problems in this exciting field.

Carey E. Priebe received his BS degree in mathematics from Purdue University in 1984, the MS degree in computer science from San Diego State University in 1988, and his PhD degree in information technology (computational statistics) from George Mason University in 1993. From 1985 to 1994, he worked as a mathematician and scientist in the US Navy research and development laboratory system. Since 1994 he has been a professor in the Department of Applied Mathematics and Statistics, Whiting School of Engineering, Johns Hopkins University, Baltimore, Maryland. At Johns Hopkins, he holds joint appointments in the Department of Computer Science, the Department of Electrical and Computer Engineering, and the Department of Biomedical Engineering, as well as the Center for Imaging Science, the Human Language Technology Center of Excellence, and the Mathematical Institute for Data Science. His research interests include computational statistics, kernel and mixture estimates, statistical pattern recognition, dimensionality reduction, model selection, and statistical inference for high-dimensional and graph data. He is a Senior Member of the IEEE, an Elected Member of the International Statistical Institute, a Fellow of the Institute of Mathematical Statistics, and a Fellow of the American Statistical Association. He won an Office of Naval Research Young Investigator Award in 1995, was named one of six inaugural National Security Science and Engineering Faculty Fellows in 2008, and was recipient of the American Statistical Association Distinguished Achievement Award in 2010. In 2011 he won the McDonald Award for Excellence in Mentoring and Advising.

Organizing Committee

Minh Tang, North Carolina State University

Avanti Athreya, Johns Hopkins University

Vince Lyzinski, University of Maryland

Schedule at a Glance

Time	Wednesday	Thursday	Friday
8:00			
8:30	Breakfast	Breakfast	Breakfast
9:00	Michael Trosset	P. Rubin-Delanchy	Lizhen Lin
9:40	Joshua Agterberg	Nathaniel Josephs	Jesus Arroyo-Relion
10:20	Coffee Break	Coffee Break	Coffee Break
10:50	Karl Rohe	Subhadeep Paul	Marianna Pensky
11:30	Joshua Cape	Daniel Sussman	Purnamrita Sarkar
12:10	Lunch	Lunch	Lunch
13:40	Jonathan Stewart	Keith Levin	
14:20	Zheng Tracy Ke	Joshua Vogelstein	
15:00	Coffee Break	Coffee Break	
15:40	Tianxi Li	Zachary Lubberts	
16:20	Liza Levina	Harrison Huibin Zhou	
17:00	High Tea		
18:00			

Workshop Schedule

Wednesday, October 18, 2023

8:30 - 8:55	Breakfast
8:55 - 9:00	Doron Levy (University of Maryland/Director, Brin MRC) *Opening*
9:00 - 9:40	Michael Trosset (University of Indiana) *Continuous Multidimensional Scaling*
9:40 - 10:20	Joshua Agterberg (University of Pennsylvania) *Estimation and Inference in Tensor Mixed-Membership Blockmodels*
10:20 - 10:50	Coffee Break
10:50 - 11:30	Karl Rohe (University of Wisconsin-Madison) *T-Stochastic Graphs*
11:30 - 12:10	Joshua Cape (University of Wisconsin-Madison) *What Do Those Dots Signify?*
12:10 - 1:40	Lunch
1:40 - 2:20	Jonathan Stewart (Florida State University) *Learning Cross-Layer Dependence Structure in Multilayer Networks*
2:20 - 3:00	Zheng Tracy Ke (Harvard University) *Optimal Network Membership Estimation Under Severe Degree Heterogeneity*
3:00 - 3:40	Coffee Break
3:40 - 4:20	Tianxi Li (University of Minnesota) *Generalized linear models and their inference on noisy network-linked data*

4:20 - 5:00	Liza Levina (University of Michigan) *Latent Space Models for Multiplex Networks with Shared Structure*
5:00 - 6:30	High Tea

Thursday, October 19, 2023

8:30 - 9:00 Breakfast

9:00 - 9:40 P. Rubin-Delanchy (University of Bristol)
A statistical perspective on the manifold hypothesis

9:40 - 10:20 Nathaniel Josephs (North Carolina State University)
Nested stochastic block model for simultaneously clustering networks and nodes

10:20 - 10:50 Coffee Break

10:50 - 11:30 Subhadeep Paul (Ohio State University)
Spectral clustering with dependent excitations for temporal networks

11:30 - 12:10 Daniel Sussman (Boston University)
Matching Embeddings via Shuffled Total Least Squares Regression

12:10 - 1:40 Lunch

1:40 - 2:20 Keith Levin (University of Wisconsin-Madison)
Pretty good, not bad, I can't complain: minimax rates for low-rank network models

2:20 - 3:00 Joshua Vogelstein (Johns Hopkins University)
Prospective Learning: Principled Extrapolation to the Future

3:00 - 3:40 Coffee Break

3:40 - 4:20 Zachary Lubberts (University of Virginia)
Beyond the adjacency matrix: Random line graphs and inference for networks with edge attributes

4:20 - 5:00 Harrison Huibin Zhou (Yale University)
Optimal Estimation of Gaussian Mixtures

7:00 - 9:00 Dinner

Friday, October 20, 2023

8:30 - 9:00	Breakfast
9:00 - 9:40	Lizhen Lin (University of Maryland) *Statistical Foundations of Deep Generative Models*
9:40 - 10:20	Jesus Arroyo-Relion (Texas A&M University) *Learning Joint and Individual Structure in Network Data with Covariates*
10:20 - 10:50	Coffee Break
10:50 - 11:30	Marianna Pensky (University of Central Florida) *Clustering of Diverse Multiplex Networks*
11:30 - 12:10	Purnamrita Sarkar (University of Texas at Austin) *Streaming PCA for Markovian Data*
12:10 - 12:15	Workshop Closing
12:15 - 1:45	Lunch

Abstracts of talks

Continuous Multidimensional Scaling

Michael Trosset

University of Indiana

Wednesday, October 18, 2023 @ 9:00 AM

Multidimensional scaling (MDS) is the act of embedding proximity information about a set of n objects in d-dimensional Euclidean space. As originally conceived by the psychometric community, MDS was concerned with embedding a fixed set of proximities associated with a fixed set of objects. Modern concerns, e.g., that arise in developing asymptotic theories for statistical inference on random graphs, more typically involve studying the limiting behavior of a sequence of proximities associated with an increasing set of objects. Standard results from the theory of point-to-set maps imply that, if n is fixed, then the limit of the embedded structures is the embedded structure of the limiting proximities. But what if n increases? It then becomes necessary to reformulate MDS so that the entire sequence of embedding problems can be viewed as a sequence of optimization problems in a fixed space. We present such a reformulation and derive some consequences.

Estimation and Inference in Tensor Mixed-Membership Blockmodels

Joshua Agterberg

University of Pennsylvania

Wednesday, October 18, 2023 @ 9:40 AM

Higher-order multiway data is ubiquitous in machine learning and statistics and often exhibits community-like structures, where each component (node) along each different mode has a community membership associated with it. In this talk we propose the tensor mixed-membership blockmodel, a generalization of the tensor blockmodel positing that memberships need not be discrete, but instead are convex combinations of latent communities. We first study the problem of estimating community memberships, and we show that a tensor generalization of a matrix algorithm can consistently estimate communities at a rate that improves relative to the matrix setting, provided one takes the tensor structure into account. Next, we study the problem of testing whether two nodes have the same community memberships, and we show that a tensor analogue of a matrix test statistic can yield consistent testing with a tighter local power guarantee relative to the matrix setting. If time permits, we will also examine the performance of our estimation procedure on flight data. This talk is based on two recent works with Anru Zhang.

T-Stochastic Graphs

Karl Rohe

University of Wisconsin-Madison

Wednesday, October 18, 2023 @ 10:50 AM

Previous statistical approaches to hierarchical clustering for social network analysis all construct an "ultrametric" hierarchy. While the assumption of ultrametricity has been discussed and studied in the phylogenetics literature, it has not yet been acknowledged in the social network literature. We show that "non-ultrametric structure" in the network introduces significant instabilities in the existing top-down recovery algorithms. To address this issue, we introduce an instability diagnostic plot and use it to examine a collection of empirical networks. These networks appear to violate the "ultrametric" assumption. We propose a deceptively simple and yet general class of probabilistic models called T-Stochastic Graphs which impose no topological restrictions on the latent hierarchy. To illustrate this model, we propose six alternative forms of hierarchical network models and then show that all six are equivalent to the T-Stochastic Graph model. These alternative models motivate a novel approach to hierarchical clustering that combines spectral techniques with the well-known Neighbor-Joining algorithm from phylogenetic reconstruction. We prove this spectral approach is statistically consistent.

What Do Those Dots Signify?

Joshua Cape

University of Wisconsin-Madison

Wednesday, October 18, 2023 @ 11:30 AM

We address the question raised in the title.

Learning Cross-Layer Dependence Structure in Multilayer Networks

Jonathan Stewart

Florida State University

Wednesday, October 18, 2023 @ 1:40 PM

Multilayer networks are a network data structure in which elements in a population of interest have multiple modes of interaction or relation, represented by multiple networks called layers. We propose a novel class of models for cross-layer dependence in multilayer networks, aiming to learn how interactions in one or more layers may influence interactions in other layers of the multilayer network, by developing a class of network separable models which separate the network formation process from the layer formation process. In our framework, we are able to extend existing single layer network models to a multilayer network model with cross-layer dependence. We establish non-asymptotic bounds on the error of estimators and demonstrate rates of convergence for both maximum likelihood estimators and maximum pseudolikelihood estimators in scenarios of increasing parameter dimension. We additionally establish non-asymptotic error bounds on the multivariate normal approximation and elaborate a method for model selection which controls the false discovery rate. We conduct simulation studies which demonstrate that our framework and method work well in realistic settings which might be encountered in applications. Lastly, we illustrate the utility of our method through an application to the Lazega lawyers network.

Optimal Network Membership Estimation Under Severe Degree Heterogeneity

Zheng Tracy Ke

Harvard University

Wednesday, October 18, 2023 @ 2:20 PM

Real networks often have severe degree heterogeneity, with maximum, average, and minimum node degrees differing significantly. This paper examines the impact of degree heterogeneity on statistical limits of network data analysis. Introducing the empirical heterogeneity distribution (EHD) under a degree-corrected mixed membership model, we demonstrate that the optimal rate of mixed membership estimation is directly linked to EHD. Surprisingly, severe degree heterogeneity can decelerate the error rate, even when the overall sparsity remains unchanged. To develop a spectral algorithm that is rate-optimal for arbitrary EHD, we propose the "two normalizations" approach to enhance the performance of existing spectral algorithms, which yields a rate-optimal for networks with arbitrary degree heterogeneity. Our findings have two significant implications: (1) Degree heterogeneity indeed influences the fundamental statistical limits; and (2) Achieving optimal adaptivity to degree heterogeneity is possible, provided that the algorithm is thoughtfully designed.

Generalized linear models and their inference on noisy network-linked data

Tianxi Li

University of Minnesota

Wednesday, October 18, 2023 @ 3:40 PM

Generalized linear models on network-linked observations have been essential in modeling the relationships between responses and covariates with additional network structures. Many approaches either lack inference tools or rely on restrictive assumptions of social effects. More importantly, these methods usually assume that networks are error-free. I introduce a class of generalized linear models with nonparametric network effects based on subspace assumptions. We do not assume the network structure to be precisely observed and is provably robust to network observational errors. An inference framework is established under the general requirement of network observational errors, and corresponding robustness is studied in detail when observational errors arise from random network models. Results reveal a phase-transition phenomenon of inference validity in relation to network density when no prior knowledge of the network model is available, while significant improvements can be achieved when the network model is known. I then briefly discuss an ensemble network estimation strategy, network mixing, which can improve the adaptivity of the proposed method. The proposed model is applied to investigate social impacts on students' perceptions of school safety based on observed friendship relations. It enables reliable analysis thanks to the nonparametric network effects and the robustness to network observational errors.

Latent space models for multiplex networks with shared structure

Liza Levina

University of Michigan

Wednesday, October 18, 2023 @ 4:20 PM

Statistical tools for analysis of a single network are now widely available, but many practical settings involve multiple networks. These can arise as a sample of networks (for example, brain connectivity networks for a sample of patients), a single network with multiple types of edges (for example, trade between countries in many different commodities), or a single network evolving over time. The term multiplex networks refers to multiple and generally heterogeneous networks observed on the same shared node set; the two examples above are both multiplex networks. We propose a new latent space model for multiplex networks which answers a key question: what part of the underlying structure is shared between all the networks, and what is unique to each one? Our model learns this from data and pools information adaptively. We establish identifiability, develop a fitting procedure using convex optimization in combination with a nuclear norm penalty, and prove a guarantee of recovery for the latent positions as long as there is sufficient separation between the shared and the individual latent subspaces. We compare the model to competing methods in the literature on simulated networks and on a multiplex network describing the worldwide trade of agricultural products.
This is joint work with Peter MacDonald and Ji Zhu.

A statistical perspective on the manifold hypothesis

P. Rubin-Delanchy

University of Bristol

Thursday, October 19, 2023 @ 9:00 AM

Complex topological and geometric patterns often appear embedded in high-dimensional data, and can often be related to some notion of ground truth, after distortion. We show that this can be explained by a generic and remarkably simple statistical model, demonstrating that manifold structure in data can emerge from elementary statistical ideas of correlation and latent variables. The Latent Metric model consists of a collection of random fields, evaluated at locations specified by latent variables and observed in noise. Driven by high dimensionality, principal component scores associated with data from this model are uniformly concentrated around a topological manifold, homeomorphic to the latent metric space. Under further assumptions this relation may be a diffeomorphism, a Riemannian metric structure appears, and the geometry of the manifold reflects that of the latent metric space. This provides statistical justification for manifold assumptions which underlie methods ranging from clustering and topological data analysis, to nonlinear dimension reduction, regression and classification, and explains the efficacy of Principal Component Analysis as a preprocessing tool for reduction from high to moderate dimension.

Nested stochastic block model for simultaneously clustering networks and nodes

Nathaniel Josephs

North Carolina State University

Thursday, October 19, 2023 @ 9:40 AM

We introduce the nested stochastic block model (NSBM) to cluster a collection of networks while simultaneously detecting communities within each network. NSBM has several appealing features including the capacity to work on unlabeled networks with potentially different node sets, the flexibility to model heterogeneous communities, and the ability to automatically select the number of classes for the networks and the number of communities within each network. This is accomplished via a Bayesian model that uses a novel employment of the nested Dirichlet process (NDP) as a prior to jointly model the between-network and within-network cluster information. Introducing dependency through the network data produces non-trivial challenges to the NDP especially related to the development of efficient samplers. We propose several Markov chain Monte Carlo algorithms including a standard Gibbs sampler, a collapsed Gibbs sampler, and two blocked Gibbs samplers for posterior inference that ultimately return two levels of clustering labels from both within and across the networks.

Spectral clustering with dependent excitations for temporal networks

Subhadeep Paul

Ohio State University

Thursday, October 19, 2023 @ 10:50 AM

Continuous-time or temporal networks observed through timestamped relational events data are commonly encountered in application settings including online social media communications, human mobility, financial transactions, and international relations. Such datasets consist of directed interaction events among entities at specific time points. Temporal networks often exhibit community structure and strong dependence patterns among node pairs. This dependence can be modeled through mutual excitations, whereby an interaction event from a sender node to a receiver node increases the possibility of future events among other node pairs. We provide statistical results for a class of models that combines high-dimensional mutually-exciting Hawkes processes with the stochastic block model for modeling both community structure and dependence among node pairs. We obtain an upper bound on the misclustering error of spectral clustering of the event count matrix as a function of the number of nodes and communities, time duration, and a quantity measuring the amount of dependence in the model. The theoretical results provide insights into the effects of dependencies in the mutually-exciting Hawkes processes on the accuracy of spectral clustering. For a particular model case that includes self and reciprocal excitation, we propose a highly scalable estimation methodology that also involves a Generalized Method of Moments (GMM) estimator and study the consistency property of the GMM estimator.

Matching Embeddings via Shuffled Total Least Squares Regression

Daniel Sussman

Boston University

Thursday, October 19, 2023 @ 11:30 AM

A frequently used approach for graph matching is first to embed the networks as points in Euclidean space and then match the embeddings. We consider the case that the two graphs have related but not identical distributions that necessitate a more complex alignment in the matching step. This is related to the problem known as shuffled linear regression. We consider a modified shuffled regression setting where there is noise in both the response and the predictor variables. This setting better matches the graph matching problem and we provide convergence rates for a shuffled total least squares method in terms of the normalized Procrustes quadratic loss.

Pretty good, not bad, I can't complain: minimax rates for low-rank network models

Keith Levin

University of Wisconsin-Madison

Thursday, October 19, 2023 @ 1:40 PM

Latent space network models, in which network structure is driven by low-dimensional geometric structure, are the workhorses of statistical network analysis. These models, in the form of the random dot product graph (RDPG) and extensions thereof, have been central to Carey Priebe's work. In this talk, I will discuss recent results establishing minimax rates for recovering the low-dimensional structure in a class of network models that includes the RDPG as a special case. Our results confirm what we all suspected: existing approaches such as the adjacency spectral embedding recover low-rank geometry at the minimax rate, up to logarithmic factors. Our minimax results also suggest a few gaps in the literature that warrant future work, which I will discuss. Time permitting, I will also discuss connections to recent projects in which we have deployed low-rank network models to address problems in causal inference and multiple-network analysis.

Prospective Learning: Principled Extrapolation to the Future

Joshua Vogelstein

Johns Hopkins University

Thursday, October 19, 2023 @ 2:20 PM

Learning is a process which can update decision rules, based on past experience, such that future performance improves. Traditionally, machine learning is often evaluated under the assumption that the future will be identical to the past in distribution or change adversarially. But these assumptions can be either too optimistic or pessimistic for many problems in the real world. Real world scenarios evolve over multiple spatiotemporal scales with partially predictable dynamics. Here we reformulate the learning problem to one that centers around this idea of dynamic futures that are partially learnable. We conjecture that certain sequences of tasks are not retrospectively learnable (in which the data distribution is fixed), but are prospectively learnable (in which distributions may be dynamic), suggesting that prospective learning is more difficult in kind than retrospective learning. We argue that prospective learning more accurately characterizes many real world problems that (1) currently stymie existing artificial intelligence solutions and/or (2) lack adequate explanations for how natural intelligences solve them. Thus, studying prospective learning will lead to deeper insights and solutions to currently vexing challenges in both natural and artificial intelligences.

Beyond the adjacency matrix: Random line graphs and inference for networks with edge attributes

Zachary Lubberts

University of Virginia

Thursday, October 19, 2023 @ 3:40 PM

Any modern network inference paradigm must incorporate multiple aspects of network structure, including information often encoded in vertices and edges. Methodology for handling vertex attributes has been developed for a number of network models, but comparable techniques for edge-related attributes remain largely unavailable. This gap is addressed in the literature by extending the latent position random graph model to the line graph of a random graph, which is formed by creating a vertex for each edge in the original random graph and connecting each pair of edges incident to a common vertex in the original graph. Concentration inequalities are proved for the spectrum of a line graph and then establish that although naive spectral decompositions can fail to extract the necessary signal for edge clustering, there exist signal-preserving singular subspaces of the line graph that can be recovered through a carefully-chosen projection. Moreover, edge latent positions can be consistently estimated in a random line graph, even though such graphs are of random size, typically have a high rank, and possess no spectral gap. The results also demonstrate that the line graph of a stochastic block model exhibits underlying block structure, and the methods are synthesized and tested in simulations for cluster recovery and edge covariate inference in stochastic block model graphs.

Optimal Estimation of Gaussian Mixtures

Harrison Huibin Zhou

Yale University

Thursday, October 19, 2023 @ 4:20 PM

This talk discusses the optimal rate of estimation in a finite Gaussian location mixture model in high dimensions without separation conditions. We assume that the number of components k is bounded and that the centers lie in a ball of bounded radius, while allowing the dimension d to be as large as the sample size n. Extending the one-dimensional result of Heinrich and Kahn (2015), we show that the minimax rate of estimating the mixing distribution in Wasserstein distance is $\Theta((d/n)^{1/4}+n^{-1/(4k-2)})$, achieved by an estimator computable in time $O(nd^2+n^{5/4})$. Furthermore, we show that the mixture density can be estimated at the optimal parametric rate $\Theta(\sqrt{d/n})$ in Hellinger distance and provide a computationally efficient algorithm to achieve this rate in the special case of $k=2$.

Both the theoretical and methodological development rely on a careful application of the method of moments. Central to our results is the observation that the information geometry of finite Gaussian mixtures is characterized by the moment tensors of the mixing distribution, whose low-rank structure can be exploited to obtain a sharp local entropy bound.

Statistical Foundations of Deep Generative Models

Lizhen Lin

University of Maryland

Friday, October 20, 2023 @ 9:00 AM

Deep generative models are probabilistic generative models where the generator is parameterized by a deep neural network. They are popular models for modeling high-dimensional data such as texts, images and speeches, and have achieved impressive empirical success. Despite demonstrated success in empirical performance, theoretical understanding of such models is largely lacking. We investigate statistical properties of deep generative models from a nonparametric distribution estimation viewpoint. In the considered model, data are assumed to be observed in some high-dimensional ambient space but concentrate around some low-dimensional structure such as a lower-dimensional manifold. Estimating the distribution supported on this low-dimensional structure is challenging due to its singularity with respect to the Lebesgue measure in the ambient space. We obtain convergence rates with respect to the Wasserstein metric of distribution estimators based on two methods: a sieve MLE based on the perturbed data and a GAN type estimator. Such an analysis provides insights into i) how deep generative models can avoid the curse of dimensionality and outperform classical nonparametric estimates, and ii) how likelihood approaches work for singular distribution estimation, especially in adapting to the intrinsic geometry of the data.

Learning Joint and Individual Structure in Network Data with Covariates

Jesus Arroyo-Relion

Texas A&M University

Friday, October 20, 2023 @ 9:40 AM

Datasets consisting of a network and covariates associated with its vertices have become ubiquitous. One problem pertaining to this type of data is to identify information unique to the network, information unique to the vertex covariates, and information that is shared between the network and the vertex covariates. Existing methods for network data often focus on capturing structure that is shared between a network and the vertex covariates but are not able to differentiate structure that is unique to each. This work formulates a solution via a low-rank model and a two-step estimation procedure composed of an efficient spectral method to obtain an initial estimate for the joint structure, followed by an optimization method that minimizes a nonconvex loss function associated with the model. We study the consistency of the initial estimate and evaluate the performance on simulated and real data.

Clustering of Diverse Multiplex Networks

Marianna Pensky

University of Central Florida

Friday, October 20, 2023 @ 10:50 AM

The talk considers the multilayer network model, where all layers of the network have the same collection of nodes and are equipped with the Generalized Dot Product Graph (GDPG) Models. In addition, all layers can be partitioned into groups with the same ambient subspace embedding, although the layers in the same group may have different matrices of connection p robabilities. The model generalizes a multitude of papers that study multilayer networks, where layers are equipped with the GDPG or various block models.

Streaming PCA for Markovian Data

Purnamrita Sarkar

University of Texas at Austin

Friday, October 20, 2023 @ 11:30 AM

Since its inception in 1982, Oja's algorithm has become an established method for streaming principle component analysis (PCA). We study the problem of streaming PCA, where the data-points are sampled from an irreducible, aperiodic, and reversible Markov chain. Our goal is to estimate the top eigenvector of the unknown covariance matrix of the stationary distribution. This setting has implications in scenarios where data can solely be sampled from a Markov Chain Monte Carlo (MCMC) type algorithm, and the objective is to perform inference on parameters of the stationary distribution. Most convergence guarantees for Oja's algorithm in the literature assume that the data-points are sampled IID. For data streams with Markovian dependence, one typically downsamples the data to get a "nearly" independent data stream. In this paper, we obtain the first sharp rate for Oja's algorithm on the entire data, where we remove the logarithmic dependence on the sample size, n, resulting from throwing data away in downsampling strategies.

The Brin Mathematics Research Center

The Brin Mathematics Research Center (Brin MRC) is a research center that sponsors activity in all areas of pure and applied mathematics and statistics. The Brin MRC was funded in 2022 through a generous gift from the Brin Family. The Brin MRC is part of the Department of Mathematics at the University of Maryland, College Park.

Activities sponsored by the Brin MRC include long programs, conferences and workshops, special lecture series, and summer schools. The Brin MRC provides ample opportunities for short-term and long-term visitors that are interested in interacting with the faculty at the University of Maryland and in experiencing the metropolitan Washington DC area.

The mission of the Brin MRC is to promote excellence in mathematical sciences. The Brin MRC is home to educational and research activities in all areas of mathematics. The Brin MRC provides opportunities to the global mathematical community to interact with researchers at the University of Maryland. The center allows the University of Maryland to expand and showcase its mathematics and statistics research excellence nationally and internationally.

List of Participants

Joshua Agterberg, University of Pennsylvania
Jesus Arroyo-Relion, Texas A&M University
Avanti Athreya, Johns Hopkins University
Joshua Cape, University of Wisconsin-Madison
Harrison Huibin Zhou, Yale University
Nathaniel Josephs, North Carolina State University
Zheng Tracy Ke, Harvard University
Keith Levin, University of Wisconsin-Madison
Liza Levina, University of Michigan
Doron Levy, University of Maryland/Director, Brin MRC
Tianxi Li, University of Minnesota
Lizhen Lin, University of Maryland
Zachary Lubberts, University of Virginia
Vince Lyzinski, University of Maryland
Subhadeep Paul, Ohio State University
Marianna Pensky, University of Central Florida
Karl Rohe, University of Wisconsin-Madison
P. Rubin-Delanchy, University of Bristol
Purnamrita Sarkar, University of Texas at Austin
Jonathan Stewart, Florida State University
Daniel Sussman, Boston University
Minh Tang, North Carolina State University
Michael Trosset, University of Indiana
Joshua Vogelstein, Johns Hopkins University