Program for Research in Mathematics, Engineering and Science

Stata Center, MIT. Photo: Slava Gerovitch

Fifteenth Annual Fall-Term PRIMES Conference

October 18-19, 2025

Saturday, October 18

Mathematics

Room 4-370, MIT

9:00 am: Welcoming Remarks

- Prof. Michel Goemans, Head of the MIT Mathematics Department
- Prof. Pavel Etingof, PRIMES Chief Research Advisor
- Dr. Slava Gerovitch, PRIMES Program Director

9:15-10:15 am: Session 1

- Leo Hong, "Generalizations of Jarnik's theorem via Total Density of Certain Subspaces" (mentor Vasiliy Nekrasov, Brandeis)
- Bryan Zhu, "Dimension Reduction for Smooth Convex Optimization via Color Refinement" (mentor Dr. Ziang Chen)
- Alyssa Yu, "Spatiotemporal Two-Pathogen Dynamics on Metapopulation Networks" (mentor Prof. Laura Schaposnik, University of Illinois Chicago)
- Aiden Jeong, "Bounds on the Distinguishing (Chromatic) Number of Posets" (mentor Yunseo Choi)

10:35-11:40 am: Session 2

- Jaewoo Park, "On the Periodic Orbits of Refractive Billiard Systems" (mentor Prof. Maxim Arnold, University of Texas Dallas)
- Jessica Hu and Aryan Raj, "Sparse Inference of Nonlinear Laboratory Earthquake Dynamics" (mentor Shijie Zhang)
- Jonathan Du, "Data-Driven Latent State Estimation and Dynamical System Identification" (mentors Shijie Zhang and Dave Darrow)
- Liam Reddy, "Fast Spectral Approximation of Multi-Valued Functions" (mentor David Darrow)

12:00-1:00 pm: Session 3

- Maggie Shen and Advaith Mopuri, "Exploring Saturn’s Narrow Ringlets with Cassini Radio Science" (mentors Dr. Ryan Maguire, MIT, and Prof. Richard French, Wellesley College)
- Bryan Guo and Brandon Xu, "Unveiling the High-Resolution Structure of Saturn's Rings from Cassini Radio Science Occultations" (mentors Dr. Ryan Maguire, MIT, and Dr. Richard French, Wellesley College)
- Evan Ashoori, Peter Bai, and Hansen Shieh, "Benchmarking Algorithms for the Jones Polynomial" (mentor Dr. Ryan Maguire)

2:00-3:10 pm: Session 4

- Sophia Hou, "Field of Definition of Abelian Surfaces with Maximal Picard Rank" (mentor Dr. Sheela Devadas, University of Maine)
- Kyle Wu, "On Number Fields with Unit Group of a Prescribed Reduction" (mentor Dr. Alexander Petrov)
- Victor Chen, Rohan Garg, and Benny Wang, "Efficient Enumeration of Quadratic Lattices" (mentor Dr. Eran Assaf)
- Michael Middlezong and Lucas Qi, "Computing Orbital Integrals as Local Densities" (mentor Dr. Thomas Rüd)

3:25-4:30 pm: Session 5

- Aaryan Vaishya, "SMT Solvers and their Applications" (mentor Brynmor Chapman)
- Luv Udeshi, "Extending CC0 Circuit Upper Bounds Beyond Symmetric Functions" (mentor Brynmor Chapman)
- Anay Aggarwal, "Automated Discovery of Extremal Unit-Distance Graphs" (mentors Andrew Gritsevskiy, Contramont Research, and Prof. Jesse Geneson, San José State University)
- Vishnu Mangipudi, Ethan Song, and Charles Zhang, "Robust Machine Learning-Enabled Spike Inference in Neuronal Voltage Imaging Data" (mentor Prof. Lu Lu, Yale)

4:45-5:45 pm: Session 6

- Omar Graia, Katrina Liu, and Audrey Wang, "AI-Based Histopathology for Lung Cancer Diagnosis" (mentor Dr. Marly Gotti, Arcus Biosciences)
- Shengtao (Alex) Ding, Jason Yang, and Tarun Rapaka, "Comparative Study of Transformer-Based Embeddings for Topic Coherence" (mentor: Dr. Willy Rodriguez, Torus AI)
- Neil Kolekar, Maiya Qiu, and Richard Wang, "On the Distribution of Irreducible Polynomials with Positive Integer Coefficients" (mentor Dr. Harold Polo, UC Irvine)

6:00-7:00 pm: Session 7

- Bofan Liu, Seabert Mao, and Michael Zhao, "Prime Element Stability in Ring Extensions" (mentors Prof. Jim Coykendall and Jared Kettinger, Clemson University)
- Grant Blitz, Darren Han, and Hengrui Liang, "Maximal Common Divisors in Monoids and Rings" (mentor Dr. Felix Gotti)
- Timothy Chen, Guanjie (Tony) Lu, and Alan Yao, "On Polynomial Evaluation at Algebraic Numbers" (mentor Dr. Felix Gotti)

Sunday, October 19

Mathematics

Room 4-370, MIT

8:45-10:05 am: Session 8 (in parallel with CS/Bio sessions in the morning)

- Feodor Yevtushenko, "Applications of Graphical Representations of Temperley-Lieb Types" (mentor Son Nguyen)
- Shihan Kanungo, "Pattern Avoidance in Sequences" (mentor Prof. Jesse Geneson, San José State University)
- Amogh Akella, "Diameter Bounds for Friends-and-Strangers Graphs" (mentor Rupert Li, Stanford)
- Kai Lum, "Extremal Structural Results for Feedback Arc Sets and Graph Inversions" (mentor Dr. Nitya Mani)
- Alexander Wang, "Edge Bounds on k-visibility Hypergraphs" (mentor Prof. Jesse Geneson, San José State University)

10:25-11:40 am: Session 9

- Eddy Li, "On a Curious Observation of the Jones Polynomial and How To Categorify It" (mentor Kenta Suzuki, Princeton)
- Shruti Arun, "On a Quantum Group in Characteristic 2" (mentors Prof. Julia Plavnik and Dr. Héctor Peña Pollastri, Indiana University Bloomington, and Prof. Dirceu Bagio, Universidade Federal de Santa Catarina)
- Amelia Chen, "The Quantum Group Associated with a Pale Block in Odd Characteristic" (mentors Prof. Julia Plavnik and Dr. Héctor Peña Pollastri, Indiana University Bloomington, and Prof. Oscar Francisco Márquez Sosa, Universidade Federal de Santa Catarina)
- Christina Wang, "An Exact Sequence of Hopf Superalgebras Arising from the Endymion Algebra in Odd Characteristic" (mentors Prof. Julia Plavnik and Dr. Héctor Peña Pollastri, Indiana University Bloomington; Prof. Saradia Sturza Della Flora and Prof. Daiana Aparecida Da Silva Flores, Universidade Federal de Santa Maria)
- Lucas Hinds, "Quantum Groups from a Fomin-Kirillov Algebra" (mentors Prof. Julia Plavnik and Dr. Héctor Peña Pollastri, Indiana University Bloomington)

12:00-1:00 pm: Session 10

- Yun (Everett) Guo, "The Center of a Tambara-Yamagami-Like Category" (mentors Prof. Monique Müller, Indiana University Bloomington/Federal University of São João del-Rei, and Prof. Julia Plavnik, Indiana University Bloomington)
- Shaohuan (Nick) Zhang, "Generalizations of Zesting Construction" (Prof. Julia Plavnik and Abigail Watkins, Indiana University Bloomington)
- Karthik Prasad, "Classifying Yetter-Drinfeld Modules over the Pansera Algebras H_{2n^2}" (mentors Prof. Julia Plavnik, Dr. Héctor Peña Pollastri, and Benjamin Spencer, Indiana University Bloomington)
- Thanh Can, "Measures on Oligomorphic Groups" (mentor Dr. Thomas Rüd)

Computer Science and Computational Biology

Room 4-270, MIT

9:30-9:35 am: Welcoming Remarks

- Dr. Slava Gerovitch, PRIMES Program Director
- Prof. Srini Devadas, PRIMES Computer Science Section Coordinator

9:35-11:00 am: Session 11: Computer Science (in parallel with Math sessions in the morning)

- Ashley Yu, "Time Efficient Swap Regret Minimization" (mentor Maxwell Fishelson)
- Michael Han, "Improved Bounds for Novelty Games" (mentor Maxwell Fishelson)
- Albert Lu, "Alcatraz: Secure Remote Computation via Sequestered Encryption in Minimally Trusted Hardware" (mentors Dr. Sacha Servan-Schreiber and Dr. Jules Drean, Tinfoil, Inc.)
- Adam Ge and Aadya Goel, "Unlearning Mechanisms for Document Classification and Fair Decision-Making" (mentor Mayuri Sridhar)
- Eric Archerman and Celine Zhang, "Architecture-Inspired Modifications to Zero-Knowledge Virtual Machines" (mentor Simon Langowski)

11:15 am - 12:45 pm: Session 12: Computational Biology

- Rajarshi Mandal, "Interpretable Epigenetic Clocks for Disease Aging Signatures" (mentor Prof. Gil Alterovitz, Harvard Medical School)
- Jialai She, "Beyond Additivity: Sparse Isotonic Shapley Regression toward Nonlinear Explainability" (mentor Prof. Gil Alterovitz, Harvard Medical School)
- Brian Li, "Transformer-Based Deep Learning for Multiclass Viral Classification in Metagenomic Sequencing" (mentor Prof. Gil Alterovitz, Harvard Medical School)
- Aaryan Arora, "MT-BN: Multi-Scale Topological Bayesian Networks for Tractable, Interpretable Structure Learning in p>>n" (mentors Prof. Gil Alterovitz, Harvard Medical School)
- Jerry Xu, "A Unified Protein Embedding Model with Local and Global Structural Sensitivity" (mentors Prof. Gil Alterovitz, Harvard Medical School, and Dr. Shaojun Pei, Brigham and Women's Hospital)
- Sophia Yan, "Mapping Enhancer-Gene Interactions in the Human Brain Reveals Reduced use of Enhancers for Synaptic Genes in Schizophrenia" (mentors Dr. Nicole Rockweiler and Prof. Steven McCaroll, Broad Institute)

Mathematics

Room 4-370, MIT

2:00-3:00 pm: Session 13

- Rio Schillmoeller, "Dynamical Functionals on Ancient ARF and AC Ricci Flows" (mentor Isaac Lopez, University of Chicago)
- Arjun Agarwal, "Semisimplifications of α_p-equivariant GL_n-modules" (mentor Dr. Arun S. Kannan, DRW)
- Kai Yamashita, "Radon Transform for General Linear Groups over Finite Fields" (mentor Ivan Motorin)
- Rachel Chen, "The Dual Canonical Basis in the Spin Representation via the Temperley-Lieb Algebra" (mentor Dr. Vasily Krylov, Harvard)

3:15-4:30 pm: Session 14

- Emma Li, "On the e-positivity of the Chromatic Symmetric Function of Left-melting Clique Chains" (mentor Dr. Foster Tom, Dartmouth College)
- Aaron Lin, "The Geometry of a Counting Formula for Deformations of the Braid Arrangement" (mentor Neha Goregaokar, Brandeis)
- Jiayu (Jerry) Liu, "Full-Twist Presentations for Fundamental Groups of Complements of Complexified Hyperplane Arrangements" (mentor Prof. Nathan Williams, University of Texas Dallas)
- Sophia Tatar, "Sharp Constructions for the Szemerédi-Trotter Theorem" (mentor Jiahui Yu)
- Michael Iofin, "Quasiconformal Normalization of Random Meromorphic Functions" (mentor Prof. Sergiy Merenkov, City University of New York)

Saturday, October 18

Session 1

Leo Hong

Generalizations of Jarnik's theorem via Total Density of Certain Subspaces

Mentor: Vasiliy Nekrasov, Brandeis

Diophantine approximation is an area of mathematics which studies how well a real number can be approximated by a rational and, more generally — how well we can approximate certain values of a given linear form, provided with some restrictions on the argument. A classical result by Jarnik shows that, under some restrictions, the set of so-called singular (uniformly very well approximable) matrices is uncountable and dense. In a recent work, Kleinbock et al. introduced a general framework of a "diophantine system", generalizing Jarnik's approach and result: they relate singularity with a topological property known as total density.

We show that certain collections of subspaces are totally dense and, as a corollary, that an analog of Jarnik's theorem holds in a wide variety of Diophantine setups. They include, but are not limited by, inhomogeneous Diophantine approximation, approximations of linear forms with arguments restricted on curves, and approximations with "restricted denominators".

Bryan Zhu

Dimension Reduction for Smooth Convex Optimization via Color Refinement

Mentor: Dr. Ziang Chen

We review the concept of graph isomorphism and the color refinement algorithm for distinguishing graphs. We then discuss how color refinement can be extended to matrices to produce equitable partitions. These partitions can be used for dimension reduction for linear programs and convex linearly constrained quadratic programs. Generalizing these methods, we propose a novel dimension reduction technique for smooth optimization problems where the objective and constraints are locally convex over the box domain. For polynomial cases, we further generalize the concept of equitability from matrices to tensors of arbitrary rank. Experiments show that our method achieves substantial reduction in runtime on existing convex quadratically constrained linear programs under rounding to various precisions, as well as synthetic convex quadratically constrained quadratic programs.

Alyssa Yu

Spatiotemporal Two-Pathogen Dynamics on Metapopulation Networks

Mentor: Prof. Laura Schaposnik, University of Illinois Chicago

Understanding interactions between the spread of multiple pathogens during an epidemic is crucial to assessing the severity of infections in living communities. In this talk, we introduce two new Multiplex Bi-Virus Reaction-Diffusion models (MBRD) on multiplex metapopulation networks: the super-infection model (MBRD-SI) and the co-infection model (MBRD-CI). These models capture two-pathogen dynamics with spatial diffusion and cross-diffusion, allowing the prediction of infection clustering and large-scale spatial distributions. Furthermore, we present examples of spatial patterns that arise within these models and discuss applications of our framework outside of epidemiology.

Aiden Jeong

Bounds on the Distinguishing (Chromatic) Number of Posets

Mentor: Yunseo Choi

In 2021, Collins and Trenk introduced the distinguishing number and distinguishing chromatic number of posets as analogs for the distinguishing (chromatic) number of graphs. A coloring c of a poset P is distinguishing if there are no nontrivial automorphisms of P preserving c. The distinguishing number $D(P)$ is the minimum number of colors in a distinguishing coloring. The distinguishing chromatic number $\chi_D(P)$ is the minimum number of colors in a proper distinguishing coloring. We present the bound $D(L) \leq |Q_L| - h(L) + 2$, where L is a lattice with join-irreducible set Q_L and height $h(L)$. For distributive lattices, Collins and Trenk showed that $\chi_D(L) \leq |Q_L| + \chi_D(Q_L) - 1$ when $\chi_D(Q_L) \geq 3$. We improve this bound to $|Q_L| + k$, where $k \leq 3$ is determined by Q_L.

Session 2

Jaewoo Park

On the Periodic Orbits of Refractive Billiard Systems

Mentor: Prof. Maxim Arnold, University of Texas Dallas

This talk introduces refractive inner and outer (dual) billiard systems, a generalization of classical billiard dynamics with Snell-type refraction rules. For the inner model, we establish a correspondence with the weighted-perimeter minimization problem for inscribed polygons: given a convex n-gon $A_1A_2 \ldots A_n$ and weights $\lambda_1, \lambda_2, \ldots \lambda_n > 0$, which $B_1B_2 \ldots B_n$ minimizes the weighted perimeter $\sum_{i=1}^{n} \lambda_i \cdot |B_{i-1}B_i|$? We present a geometric proof in the case $n = 3$ and outline new techniques that arise for $n \geq 4$. In the second part, we define refractive outer billiards and introduce one of our results, a generalization of a result by Gutkin and Simanyi.

Jessica Hu and Aryan Raj

Sparse Inference of Nonlinear Laboratory Earthquake Dynamics

Mentor: Shijie Zhang

The frictional state of faults plays a fundamental role in earthquake nucleation and recurrence. Laboratory earthquake ("labquake") experiments provide controlled conditions for investigating stick-slip dynamics, yielding high-resolution data on shear stress, gouge evolution, and acoustic emissions (AEs). While prior work has shown that AEs can be used to forecast failure times with machine learning, such models provide limited physical interpretability and require large datasets. Moreover, mechanistically discovering the governing equations of labquake systems remains difficult due to their nonlinear and stochastic nature. In contrast, data-driven inference of governing equations offers a transparent and flexible pathway to uncover the physical rules behind fault stress evolution. Here, we introduce a novel stochastic Bayesian inference framework to systematically analyze labquake frictional dynamics. We infer governing stochastic differential equations (SDEs) with inhomogeneous Poisson processes, incorporating both microslip and major slip failures. Our results reveal that the growth function for fault stress follows a nonlinear hyperbolic sine relation with respect to the frictional state, inferred using the ADAM-SINDy framework, which combines sparse nonlinear model discovery with Adam optimization. This approach provides physically interpretable governing equations, bridging the gap between phenomenological labquake data and theoretical friction laws. More broadly, our methodology demonstrates how stochastic inference can mechanistically uncover governing equations in complex natural systems such as landslides and fracture mechanics.

Jonathan Du

Data-Driven Latent State Estimation and Dynamical System Identification

Mentors: Shijie Zhang and Dave Darrow

In many real-world systems, the full state of a dynamical system is not directly observable, posing a challenge for traditional system identification techniques that require complete state information, such as the Sparse Identification of Nonlinear Dynamics (SINDy) algorithm. In this presentation, we demonstrate a novel framework for the inference of latent (unobserved) states and governing differential equations from partial observations. The proposed Latent-SINDy (L-SINDy) algorithm extends the classical SINDy methodology by jointly estimating latent state trajectories and recovering sparse, interpretable dynamics using nonlinear basis functions. To validate this approach, we successfully apply the L-SINDy methodology to several nonlinear benchmark systems under partial observation scenarios.

Liam Reddy

Fast Spectral Approximation of Multi-Valued Functions

Mentor: David Darrow

Multi-valued functions arise in a variety of 2-D physical systems, including fluid flow and electrostatics. In this talk, we introduce a new method for efficiently approximating and working with multi-valued functions in the complex plane, combining the strengths of the AAA algorithm for rational approximation with those of the 'log-lightning' algorithm for functions with fixed branch point singularities. Our proposed method is able to approximate functions over the complex plane, even when branch point locations are not known ahead of time. We provide a brief explanation of previous algorithms and demonstrate the strength of our approach against previous algorithms.

Session 3

Maggie Shen and Advaith Mopuri

Exploring Saturn's Narrow Ringlets with Cassini Radio Science

Mentors: Dr. Ryan Maguire, MIT, and Prof. Richard French, Wellesley College

The Cassini mission performed occultations of Saturn's rings by transmitting three bands of radio waves through the rings towards Earth. Ring particles much smaller than the wavelength of each band do not cause significant diffraction or absorption of light and are therefore essentially undetected. Comparing the wavelength dependence of optical depth (a measure of detected particles) to values predicted by Mie scattering theory allows us to constrain the distribution of particle sizes in the rings. We determine that there are significant regional differences in size distributions between the F ring, the Strange ringlet, and the C ring plateaus. Our results indicate that the F ring properties are different from those of other ring regions, which may be related to the speculated clumpy nature of the F ring itself.

Bryan Guo and Brandon Xu

Unveiling the High-Resolution Structure of Saturn's Rings from Cassini Radio Science Occultations

Mentors: Dr. Ryan Maguire, MIT, and Prof. Richard French, Wellesley College

Saturn's rings display a remarkable range of structures, from broad plateaus to fine spiral waves and ripples. Understanding their origins and dynamics requires profiling at the highest attainable resolutions. Currently, processed Cassini radio occultation profiles in the NASA Planetary Data System (PDS) have a radial resolution of 1 km. We introduce a diffraction inversion technique achieving radial resolutions down to 55 m, nearly 20 times finer than existing profiles. Two interpolation-based variants preserve most accuracy at much lower computational cost while still outperforming the traditional

Fresnel quadratic approximation. High-resolution reconstructions are validated through forward modeling, in which synthetic signals generated from ideal ring profiles under actual occultation geometry are reprocessed through the full diffraction inversion pipeline. We present multi-resolution reconstructions of several features in the C ring, comparing profiles from 1 km resolutions to 55 m. Our advances will allow to study previously inaccessible features in Saturn's rings and provide new opportunities for Kronoseismology, the study of Saturn's interior through its ring structure.

Evan Ashoori, Peter Bai, and Hansen Shieh

Benchmarking Algorithms for the Jones Polynomial

Mentor: Dr. Ryan Maguire

In our talk, we first introduce knots, links, and their invariants. Next, we define the Kauffman bracket polynomial and writhe, which leads us to a definition of the Jones polynomial. This is a now widely used knot invariant that was discovered in the 1980s. We then explore various algorithms, including a tangle algorithm, skein-template algorithm, and load-balanced algorithm, for calculating the Jones polynomial and remark on their performance for small knots. This has immediate applications in tabulation efforts for 21 and higher crossing knots, which use invariants such as the Jones polynomial to remove duplicate entries.

Session 4

Sophia Hou

Field of Definition of Abelian Surfaces with Maximal Picard Rank

Mentor: Dr. Sheela Devadas, University of Maine

An elliptic curve is a smooth projective curve given by an equation of the form $y^2 = x^3 + ax + b$, where (a, b) lie in some field K. These curves have the structure of an abelian group meaning we can add points. The minimal field over which an elliptic curve can be defined is governed by its j-invariant, which determines the curve's isomorphism class over $\overline{K}$. For higher-dimensional analogues—abelian varieties, the problem of determining the field of definition becomes far more difficult.

Kyle Wu

On Number Fields with Unit Group of a Prescribed Reduction

Mentor: Dr. Alexander Petrov

We investigate the attainability of various groups as images of the unit group of a number field under reduction modulo a prime ideal. We recall that the rank of the group of units is given by Dirichlet's unit theorem. For a prime number p inert in a degree n number field the reductions of units defines a subgroup of the multiplicative group $\mathbb{F}_{p^n}^{\times}$ of the residue field. We prove several results about possible images under reduction when fixing a rational prime p and degree n of the number field K and then varying K, fully describing the maximal image for general n and obtaining a complete description of the possible images in the quadratic case. We also consider the analogous problem for unit groups of non-maximal orders of a number field.

Victor Chen, Rohan Garg, and Benny Wang

Efficient Enumeration of Quadratic Lattices

Mentor: Dr. Eran Assaf

In this talk, we present an algorithm that efficiently finds a representative of a genus of quadratic lattices. Our algorithm is the first polynomial-time algorithm to compute maximal overlattices. It is also the fastest known algorithm for computing representatives when paired with existing algorithms. We review basic facts on lattices, bilinear/quadratic forms, and genera. We present bounds on the number of genus symbols for a fixed rank and determinant. We analyze the running time complexities of existing algorithms and compare both their theoretical and practical time complexities with our algorithm.

Michael Middlezong and Lucas Qi

Computing Orbital Integrals as Local Densities

Mentor: Dr. Thomas Rüd

Orbital integrals are central objects in harmonic analysis, but they are difficult to compute. Starting from the simple idea that the characteristic polynomial of a matrix is invariant under basis change, we study orbital integrals over the conjugacy class of a matrix. Broadly speaking, these orbital integrals compute the proportion of bases in which a given linear transformation has integer coefficients. After an introduction to p-adic numbers and how p-adic volumes are computed, we discuss conjugacy of matrices and how matrices can be conjugate in GL_n but not SL_n. In the end, we showcase a finite counting method for computing orbital integrals.

Session 5

Aaryan Vaishya

SMT Solvers and their Applications

Mentor: Brynmor Chapman

The boolean satisfiability (SAT) problem is one of the core problems in computer science. The SAT problem asks whether a formula written in boolean variables, ANDs, ORs, and NOTs has an assignment of its variables that causes the formula to evaluate to TRUE. In his seminal work, Karp proved that SAT is an NP-complete problem, so all problems in NP reduce to it. It became known as one of twenty-one canonically NP-complete problems stemming from Karp's work. In practice, many important problems reduce simply and efficiently to SAT, and while there exists no fast algorithm to solve SAT, we can often find solutions that are practically efficient.

Satisfiability modulo theories (SMT) is a generalization of SAT that deals with many more data types than booleans, more complex formulas, and larger data structures. SMT is far more expressive than SAT, even allowing us to express uncomputable problems in terms of an SMT problem. Yet, heuristic SMT solvers exist that can solve a wide variety of SMT problems in reasonable time. The versatility of the language of SMT allows us to convert many mathematical problems (far further than just NP problems) into cases of SMT that can be checked by computer. We present several examples of the usage of SMT solvers in mathematics and in our own work.

Luv Udeshi

Extending CC0 Circuit Upper Bounds Beyond Symmetric Functions

Mentor: Brynmor Chapman

Boolean circuit complexity seeks to understand how efficiently Boolean functions can be computed by circuits of limited depth and size. A central focus in this area is the class CC^0, consisting of constant-depth circuits with counting gates. While powerful lower bounds are known for prime moduli, much less is understood for composite moduli. In this talk, we discuss some constructions of small-depth MOD_m circuits for broad families of functions, extending known upper bounds beyond symmetric functions.

Anay Aggarwal

Automated Discovery of Extremal Unit-Distance Graphs

Mentors: Andrew Gritsevskiy, Contramont Research, and Prof. Jesse Geneson, San José State University

Unit distance graphs (UDGs), graphs formed by points in space with edges between pairs exactly one unit apart, have long stood at the intersection of geometry, combinatorics, and theoretical physics. In this talk, we explore them through the lens of computational discovery, focusing on strategies for constructing dense examples in low dimensions. We introduce two complementary frameworks for discovery. In the continuous setting, we introduce the notion of (ε, δ)-UDGs and show how they relax exact distance constraints, enabling a reduction of an infinite action space to a finite one. In the discrete setting, we show that restricting the search to lattices in $\mathbb{R}^d$ is sufficient, and we identify particularly fruitful lattices for exploration. Using computational methods such as simulated annealing, we construct new dense UDGs and analyze their structure. Finally, we discuss how these constructions relate to quantum contextuality and the famous Kochen-Specker theorem.

Vishnu Mangipudi, Ethan Song, and Charles Zhang

Robust Machine Learning-Enabled Spike Inference in Neuronal Voltage Imaging Data

Mentor: Prof. Lu Lu, Yale

One of the most useful heuristics in neuroscience is the pointwise location data of spikes from signals generated by neurons in response to environmental stimuli. Identifying spikes is crucial to better understanding the information encoded in neuronal signals; however, due to low spike frequency, signal mean drift, high background noise, and other challenges, current automatic empirical identification methods are inadequate. In this talk, we introduce a robust machine-learning driven approach to infer spike locations in neuronal voltage imaging data, incorporating tailored model architecture design, feature augmentation through pre-processing, and physiologically informed post-processing methods. Our model's strong performance on testing datasets opens pathways toward more accurate spike identification and broader applications in signal-based data analysis.

Session 6

Omar Graia, Katrina Liu, and Audrey Wang

AI-Based Histopathology for Lung Cancer Diagnosis

Mentor: Dr. Marly Gotti, Arcus Biosciences

We extend prior work on the LungHist700 dataset by examining (i) cross-dataset robustness, (ii) multi-scale learning, and (iii) explainability. We compare classical pipelines (e.g., logistic regression, random forest with engineered color/texture features) and a deep model (ResNet50V2) across internal (LungHist700) and external (LC25000) evaluations. We find that class imbalance, staining variation, and limited sample size materially constrain generalization; on LC25000, transfer performance degrades across methods, and calibrated PCA-based pipelines can match or exceed the CNN. A central in-domain result is that our *multi-scale* random-forest model that integrates 20× and 40× magnifications attains **95%** accuracy on the LungHist700 test set, surpassing the ResNet50-based multiclass baseline reported in the original LungHist700 paper (**90%** at 20× and **82%** at 40×). Furthermore, we use SHAP (for a transparent classical model) to reveal morphological cues that drive predictions and to diagnose failure modes under domain shift. Overall, our study shows that (a) multi-scale representations substantively improve in-domain performance and (b) strong classical baselines remain competitive in histopathology, while robust cross-dataset transfer requires careful attention to task definition, preprocessing, and calibration.

Shengtao (Alex) Ding, Jason Yang, and Tarun Rapaka

Comparative Study of Transformer-Based Embeddings for Topic Coherence

Mentor: Dr. Willy Rodriguez, Torus AI

Topic modeling is a branch of Natural Language Processing (NLP) that leverages probability theory and optimization to organize large collections of texts into coherent groups according to word co-occurrence patterns. Among the various models proposed, Latent Dirichlet Allocation (LDA) remains one of the most widely used and interpretable probabilistic approaches. In our work, we use some previously proposed metrics to explore how the quality of topic decomposition is influenced by the use of modern language representations. Specifically, we compare the performance of different transformer-based architectures with different sizes to assess whether larger and more complex models lead to better or more interpretable topics. In this presentation, we will first introduce the main concepts behind topic modeling and LDA, including the probabilistic intuition that underlies them. We will then discuss recent advances in NLP, such as word embedding algorithms and the transformer architecture, and how these can be integrated into topic modeling pipelines. Finally, we will present and discuss our experimental results.

Neil Kolekar, Maiya Qiu, and Richard Wang

On the Distribution of Irreducible Polynomials with Positive Integer Coefficients

Mentor: Dr. Harold Polo, UC Irvine

We investigate the distribution of irreducibles in $\mathbb{N}_0[x]$, the semidomain of polynomials with nonnegative integer coefficients. Our main result shows that, with respect to a natural asymptotic, almost every polynomial in $\mathbb{N}_0[x]$ is irreducible. We next specialize to 0–1 polynomials (coefficients in $\{0,1\}$) and establish analogous estimates. The methods further adapt to $\mathbb{N}_0[\![x]\!]$, yielding a corresponding analysis for power series with nonnegative integer coefficients.

Session 7

Bofan Liu, Seabert Mao, and Michael Zhao

Prime Element Stability in Ring Extensions

Mentors: Prof. Jim Coykendall and Jared Kettinger, Clemson University

The behavior of prime elements under ring extensions is a classical topic in the area of commutative algebra. Given an extension of domains $R \subseteq T$, we identify conditions under which an arbitrary prime $p \in R$ remains prime in all intermediate ring extensions $R \subseteq S \subseteq T$. In such cases, we say the extension exhibits prime stability. First, assuming that p is prime in T, we prove that prime stability holds whenever T is an integral overring of a one-dimensional ring R. Furthermore, we show that if p is coprime to the conductor of the extension $R \subseteq T$, then p remains prime in T and prime stability holds. Next, with the help of a result on prime behavior in minimal extensions, we prove that prime stability holds for any extension satisfying the Finite Chain Property. Finally, we consider the converse problem in certain settings.

Grant Blitz, Darren Han, and Hengrui Liang

Maximal Common Divisors in Monoids and Rings

Mentor: Dr. Felix Gotti

We say that a commutative monoid M has the MCD property if every nonempty finite subset of M has a maximal common divisor (MCD), and we say that M has the MCD-finite property if every nonempty finite subset of M has only finitely many MCDs up to associates. It is well known that every monoid that satisfies the ascending chain condition on principal ideals is an MCD monoid, while every monoid with the finite factorization property is an MCD-finite monoid. In this talk, we study the MCD and MCD-finite properties in the setting of cancellative commutative monoids and both their power monoids and their monoid algebras.

Timothy Chen, Guanjie (Tony) Lu, and Alan Yao

On Polynomial Evaluation at Algebraic Numbers

Mentor: Dr. Felix Gotti

A systematic study of the additive monoid M_α of the subsemiring $\mathbb{N}_0[\alpha] := \{p(\alpha) \mid p(x) \in \mathbb{N}_0[x]\}$ of $\mathbb{C}$ (for algebraic parameters α) was initiated by Correa-Morris and Gotti in 2022. In this talk, we investigate the structure from the antimatter viewpoint. Following Coykendall, Dobbs, and Mullins, we say that a commutative monoid is antimatter if it contains no irreducibles. We find a necessary and sufficient condition for M_α to be antimatter: for positive real algebraic α, the monoid M_α is antimatter if and only if α^{-1} is a Perron number, i.e., an algebraic integer that exceeds each of its conjugates in norm, and has no positive conjugates aside from itself.

SESSION 8

Feodor Yevtushenko

Applications of Graphical Representations of Temperley-Lieb Types

Mentor: Son Nguyen

Temperley-Lieb immanants are generalizations of determinants that are known to be Schur positive when evaluated on Jacobi-Trudi matrices. There is a purely combinatorial rule to compute their generalized Littlewood-Richardson coefficients using shuffle tableaux, a generalization of semistandard Young tableaux. We will use this rule to explore some conjectures and their applications, along with how this framework motivates novel conjectures that are central to Littlewood-Richardson coefficients.

Shihan Kanungo

Pattern Avoidance in Sequences

Mentor: Prof. Jesse Geneson, San José State University

One common theme in mathematics is asking how large or complicated an object can be if it avoids a certain forbidden pattern. For example, suppose we want to build a sequence of symbols, but we require that some particular smaller sequence never appears inside it. Questions like this show up in extremal combinatorics, a field that studies the limits of such constructions. In this talk, I will focus on a function called the saturation function. This function measures the length of the shortest sequence that not only avoids a forbidden pattern, but would immediately contain that pattern if we tried to make it any longer. A recent conjecture proposed that this function always grows at most linearly (that is, roughly proportional to the number of distinct symbols we use). I will explain how we designed an algorithm that builds these "saturated" sequences and how this helps us prove the conjecture for several important cases. I will also describe two large families of forbidden patterns for which we can now completely settle the conjecture.

Amogh Akella

Diameter Bounds for Friends-and-Strangers Graphs

Mentor: Rupert Li, Stanford

Consider two n-vertex graphs X and Y, where we interpret X as a social network with edges representing friendships and Y as a movement graph with edges representing adjacent positions. The friends-and-strangers graph $\mathsf{FS}(X,Y)$ is a graph on the $n!$ permutations $V(X) \to V(Y)$, where two configurations are adjacent if and only if one can be obtained from the other by swapping two friends located on adjacent positions.

Friends-and-strangers graphs were first introduced by Defant and Kravitz, and generalize sliding puzzles as well as token swapping problems. Previous work has largely focused on their connectivity properties. In this talk, we study the diameter of the connected components of $\mathsf{FS}(X,Y)$. Our main result shows that when the underlying friendship graph is a star with n vertices, the friends-and-strangers graph has components of diameter $O(n^4)$. This implies, in particular, that sliding puzzles are always solvable in polynomially many moves. Our work also provides explicit efficient algorithms for finding these solutions.

We then extend our results to general graphs in two ways. First, we show that the diameter is polynomially bounded when both the friendship and the movement graphs have large minimum degree. Second, when both the underlying graphs X and Y are Erdős-Rényi random graphs, we show that the distance between any pair of configurations is almost always polynomially bounded under certain conditions on the edge probabilities.

Kai Lum

Extremal Structural Results for Feedback Arc Sets and Graph Inversions

Mentor: Dr. Nitya Mani

Directed graphs (digraphs) capture one-way dependencies such as flows, priority orderings, or rankings. Directed cycles often signal issues such as circular dependencies or deadlocks. A Feedback Arc Set is a set of edges whose removal can break all cycles, and a standard measure of a graph's complexity is the size of the minimum feedback arc set, denoted by $\beta(G)$. A digraph is r-free if it contains no directed cycles of length r or shorter. In this presentation, we discuss the bounds for $\beta(G)$ for extremal structural results, and present our results for r-free digraphs and for edge inversion, where selected groups of edges are reversed rather than removed.

Alexander Wang

Edge Bounds on k-visibility Hypergraphs

Mentor: Prof. Jesse Geneson, San José State University

Visibility graphs have been used in computational geometry, robot path planning, and very-large-scale integration (VLSI) layout design because they efficiently describe spatial relationships between objects. In this talk, we discuss certain types of visibility graphs called bar, semi-bar, arc, semi-arc, and rectangle visibility graphs, as well as generalizations of these visibility graphs called k-visibility graphs and k-visibility hypergraphs. In each of these visibility graphs, we discuss results on the maximum number of edges in the graph for a fixed number of vertices and how these bounds relate to applications in VLSI layout design.

Session 9

Eddy Li

On a Curious Observation of the Jones Polynomial and How To Categorify It

Mentor: Kenta Suzuki, Princeton

Determining whether two knots can deform into each other is a difficult topological problem. Jones assigned a polynomial $V_K(t)$ to each knot K which remains invariant under all deformations of K. The Jones polynomial can be defined as a $\mathbb{C}(t)$-linear combination of characters of the Temperley-Lieb algebra $\mathsf{TL}_n(t^{1/2}+t^{-1/2})$. For even n the coefficients of the linear combination have a pole at $t=-1$, which corresponds to the vanishing of an alternating sum of characters of $\mathsf{TL}_n(0)$. The character relation categorifies to an exact sequence of standard modules of $\mathsf{TL}_n(0)$, which we construct explicitly. From the exact sequence, we also deduce the structure of all irreducible modules of $\mathsf{TL}_n(0)$.

Shruti Arun

On a Quantum Group in Characteristic 2

Mentors: Prof. Julia Plavnik and Dr. Héctor Peña Pollastri, Indiana University Bloomington, and Prof. Dirceu Bagio, Universidade Federal de Santa Catarina

Quantum groups have many implications in both mathematics and physics. They quantify non-classical symmetry, i.e. one with a braiding. Modules over a quantum group thus represent how such braided symmetries act. In this talk, we work with a Yetter-Drinfeld module in characteristic 2 with a point and a pale block of dimension 3. We compute its Nichols algebra, a structure that encompasses the quantum analog of infinitesimal symmetries. We then use various constructions to calculate and present quantum groups associated with this Nichols Algebra. We have explicit presentations of several simple modules over one of these quantum groups.

Amelia Chen

The Quantum Group Associated with a Pale Block in Odd Characteristic

Mentors: Prof. Julia Plavnik and Dr. Héctor Peña Pollastri, Indiana University Bloomington, and Prof. Oscar Francisco Márquez Sosa, Universidade Federal de Santa Catarina

It is known that applying the Drinfeld double construction to Nichols algebras of diagonal type over a field of characteristic zero yields examples of Drinfeld-Jimbo type quantum groups, objects used to model phase transitions in statistical mechanics. In this talk, we will explore a generalized notion of quantum group obtained by applying the double construction to a recently discovered Nichols algebra over a field of characteristic p, whose vector space decomposes as a pale block and a point. We will define Hopf and Nichols algebras, and introduce the generators and relations of $\mathfrak{C}_*(q)$, the algebra we are working with. We will then demonstrate the computation of the Drinfeld double and examine the structure of the resulting algebra.

Christina Wang

An Exact Sequence of Hopf Superalgebras Arising from the Endymion Algebra in Odd Characteristic

Mentors: Prof. Julia Plavnik and Dr. Héctor Peña Pollastri, Indiana University Bloomington; Prof. Saradia Sturza Della Flora and Prof. Daiana Aparecida Da Silva Flores, Universidade Federal de Santa Maria

Exact sequences provide a framework for understanding how complex algebraic structures can be built from simpler components. Analogously, in the study of Hopf algebras, exact sequences allow us to decompose and construct quantum groups from known Hopf algebras. In this talk, we investigate an exact sequence of Hopf superalgebras which arises naturally from the Drinfeld double of the bosonization of an Endymion algebra in odd characteristic. We then analyze the restricted enveloping algebra of a Lie superalgebra that results from the exact sequence.

Lucas Hinds

Quantum Groups from a Fomin-Kirillov Algebra

Mentors: Prof. Julia Plavnik and Dr. Héctor Peña Pollastri, Indiana University Bloomington

Quantum groups arose in the 1980s as algebraic objects providing deformations of classical Lie algebras and function algebras on groups. They have since found applications in areas such as knot theory and quantum field theory. In this talk, I will review the historical development of quantum groups and their algebraic framework. I will then introduce the Fomin–Kirillov algebras, focusing particularly on $\mathcal{FK}_3$, as well as a related quantum function algebra constructed through codouble bosonization, which is dual to a quantum group studied by Pogorelsky and Vay.

Session 10

Yun (Everett) Guo

The Center of a Tambara-Yamagami-Like Category

Mentors: Prof. Monique Müller, Indiana University Bloomington/Federal University of São João del-Rei, and Prof. Julia Plavnik, Indiana University Bloomington

Fusion categories can be thought of as a generalization of rings in the categorical setting and have applications in topological quantum field theory. The Drinfeld center construction gives a way to construct modular categories from spherical fusion categories. A paper by Gelaki, Naidu, and Nikshych provides a technique to compute the centers of fusion categories faithfully graded by a finite group. In this talk, we will introduce the notion of fusion categories and give a brief exposition of this technique. We will define the Tambara-Yamagami category, which provides a coherent associativity structure for a fusion rule obtained by extending the finite group algebra with one additional noninvertible object. We will introduce the category $\mathsf{GLM}(G,\sigma,\omega,\delta,\epsilon)$ defined in a recent paper by Galindo, Lentner, and Möller, which generalizes the Tambara-Yamagami category for even groups. We will also discuss our results on the computation of the center of this category.

Shaohuan (Nick) Zhang

Generalizations of Zesting Construction

Mentors: Prof. Julia Plavnik and Abigail Watkins, Indiana University Bloomington

Given the usefulness of fusion categories in topology, quantum field theory, representation theory, and much more, there is a need for concrete examples to study their properties. Zesting thus arose as a process to construct new fusion categories by modifying structures of existing ones with cohomological data, and does this in a controlled manner so that invariants are easy to calculate. We introduce two generalizations of the zesting construction that weaken the requirements of the original and show that they still produce new fusion categories. We show that the first generalization produces associative zestings that are equivalent to original ones, and analyze the obstruction to the second generalization and the fusion categories it can produce.

Karthik Prasad

Classifying Yetter-Drinfeld Modules over the Pansera Algebras H_{2n^2}

Mentors: Prof. Julia Plavnik, Dr. Héctor Peña Pollastri, and Benjamin Spencer, Indiana University Bloomington

Hopf algebras are algebraic structures generalizing groups. In the same way that groups can be understood by their normal subgroups, Hopf algebras can be understood by their coradical. However, it is very difficult to classify such Hopf algebras. The Andruskiewitch-Schneider lifting method provides a method to classify all Hopf algebras with a given coradical; but this has been done mostly for the case of the coradical of a group algebra.

In recent years, there has been great interest in generalizations of the Kac-Paljutkin Algebra H_8, the first non-commutative, non-cocommutative Hopf algebra (notably, its coradical is NOT a group algebra). The first such generalization, the Pansera Algebras H_{2n^2}, was constructed in 2017. In this work, we present partial progress towards the first ingredient for the Andruskiewitch-Schneider lifting method: classification of the simple Yetter-Drinfeld modules over the Pansera algebras H_{2n^2}.

Thanh Can

Measures on Oligomorphic Groups

Mentor: Dr. Thomas Rüd

Deligne's theorem says that any "well-behaved" class of finite-dimensional vector spaces with controlled tensor power growth must come from representations of a (super)group. An important goal is to explore classes of finite-dimensional vector spaces with exponential growth. An oligomorphic group is a group G acting on a set Ω such that G has finitely many orbits on Ω^n for all $n \geq 0$. Considering these orbits as relations on a class of finite combinatorial structures (graphs, total orders, etc.), Harman and Snowden introduced in 2022 the notion of measure, which is a rule assigning each embedding of structures a value such that a list of axioms is satisfied. Measures are an essential ingredient for constructing new classes of finite-dimensional vector spaces, yet classifying measures on a given oligomorphic group is a difficult combinatorial problem. In this talk, we survey concrete examples of oligomorphic groups (e.g., S_∞, $\mathrm{Aut}(\mathbb{Q},<)$, automorphism groups of variants of trees) and discuss measures on them.

Ashley Yu

Time Efficient Swap Regret Minimization

Mentor: Maxwell Fishelson

No-regret learning algorithms provide principled frameworks for multi-agent decision-making, with swap regret minimization enabling convergence to correlated equilibria, a stronger solution concept than the coarse correlated equilibria achieved by external regret algorithms. The classical Blum-Mansour (BM) algorithm achieves optimal $O(\sqrt{NT \log N})$ swap regret bounds, but computing the stationary distribution of an $N \times N$ Markov chain at each iteration requires $O(N^3)$ time complexity that severely limits scalability.

We propose a novel approach that replaces exact stationary distribution computation with efficient sampling-based estimation, reducing per-iteration complexity from $O(N^3)$ to $O(N)$ while maintaining the fundamental structure of the original algorithm.

Michael Han

Improved Bounds for Novelty Games

Mentor: Maxwell Fishelson

We study the *novelty* game, introduced by Léchine and Seiller in 2024, where p non-communicating players each receive k integers from $[N]$ and output m integers from $[N]$. The players succeed if some player outputs an integer not among the pk inputs. We improve upper bounds on the minimal N admitting a guaranteed winning strategy. Focusing on oblivious strategies and the case $m = 1$, we develop a framework with six optimizations that sharpen the classical bound. For $(p, k, m) = (3, 2, 1)$ we reduce the best known bound from 1.71×10^8 to 193050. In general $(p, k, 1)$ settings, our analysis yields an upper bound of the form $\prod_{i=0}^{p-1} \binom{D(p,k)}{k^i}$, leading to an exponential reduction of roughly $e^{2k^{p/2}} \prod_{i=0}^{p-1} (k^i)!$ relative to q^r. We also propose two lower-bound conjectures and conjecture that our upper bounds are tight up to subexponential factors.

Albert Lu

Alcatraz: Secure Remote Computation via Sequestered Encryption in Minimally Trusted Hardware

Mentors: Dr. Sacha Servan-Schreiber and Dr. Jules Drean, Tinfoil, Inc.

This talk introduces “Alcatraz,” a new architecture that enables secure remote computation with minimal trust in the hardware. In Alcatraz, sensitive data is always encrypted, except when it is inside a small, trusted circuit, which is composed of an Arithmetic Logic Unit (ALU) gated by a decryption and encryption engine. By design, the internal states of the trusted circuit is inaccessible from any software, and unencrypted data is never exposed outside the trusted circuit. Thus it is extremely difficult for any attacker to gain information about the sensitive data by observing or attacking other parts of the processor and computer (e.g., registers or caches), or by exploiting any microarchitectural side channels.

We implemented Alcatraz on a field-programmable gate array (FPGA), and verified with a formal proof that the circuit is secure at the wire-level, which is stronger than the register-transfer-level (RTL) security proved previously. Wire-level verification has the benefit that it is much closer to the physical reality, i.e., the timing and level of signals on the wire, that may be observed by attackers.

We apply Alcatraz to single-server private information retrieval and estimate based on benchmark that Alcatraz achieves significant improvement when compared with the current state-of-the-art approach for private information retrieval.

Adam Ge and Aadya Goel

Unlearning Mechanisms for Document Classification and Fair Decision-Making

Mentor: Mayuri Sridhar

In this talk, we investigate machine unlearning as a principled way to address two risks in machine learning models: privacy leakage and representational bias. On the privacy side, we study class-level unlearning for document classification. Documents are encoded with TF-IDF and classified by an SVM (hinge loss, L2). When a class must be removed, we apply a Hessian downweight, a fast second-order (inverse-Hessian) parameter subtraction that removes the contribution of the deleted class to the model, followed by a short L-BFGS fine-tuning burst on the reduced objective to mimic full retraining. Rather than randomly reclassifying the affected documents (which hurts utility), we reassign them to the next-best class by similarity. We evaluate privacy with membership-inference AUC on the unlearned class, which scores the ability of an attacker to distinguish whether samples from a class were in training. Across datasets, the starting AUC is high ($\approx$0.71–0.88) before unlearning, illustrating leakage, and drops to $\approx$0.49–0.51, indicating random-chance, after our method. Our classification accuracy on the retained classes remains preserved. On the fairness front, machine learning models are increasingly relied on in high stakes decision making contexts like loan approvals and hiring. However, models trained on data reflecting existing societal inequities often learn and amplify societal biases, which can lead to discriminatory outcomes for underrepresented or marginalized groups. Our primary objective is to address this by removing sensitive attribute information from the neural representation, thereby ensuring predictions rely solely on legitimate qualifications. We have developed a technique and achieved an approximately 50% reduction in the standard bias metrics (Maximum Equalized Odds/Opportunity Difference). Furthermore, the mutual information between the neural representation and sensitive attributes dropped from 14.9 to 7.4 bits on a widely used dataset, all while maintaining nearly the same task prediction accuracy.

Eric Archerman and Celine Zhang

Architecture-Inspired Modifications to Zero-Knowledge Virtual Machines

Mentor: Simon Langowski

Zero Knowledge Virtual Machines (zkVMs) enable the efficient verification of arbitrary computer programs run by an untrusted party without leaking any knowledge of the program's inputs. A program is compiled into assembly instructions, whose correct execution on a virtualized processor is proven using a combination of cryptographic schemes. zkVMs currently find applications in privacy focused and high throughput blockchain projects, but are useful to generate efficient proofs for any program, especially where conditional statements, loops, array indexing, and other non-circuit friendly logic is used. In this talk, we outline computer hardware-inspired modifications to zkVMs, using the RISC-V zkVM Jolt as an example. Jolt currently processes register accesses in the same way as RAM accesses, so we move register accesses to map to the "fast memory" of the Rank-1 Constraint System (R1CS). Additionally, we implement memory batching for RAM accesses to reduce virtual memory traffic, and handle the additional computation associated with this batching again through the R1CS. We find that while in some applications our modifications bring a speedup, in most cases the blowup of the R1CS causes an overall slowdown — which depends on the number and types of instructions used in the program. When combined with a future additional optimization to reduce the trace length, a program with many sequential accesses to memory, for example, will benefit from virtual memory batching.

Rajarshi Mandal

Interpretable Epigenetic Clocks for Disease Aging Signatures

Mentor: Prof. Gil Alterovitz, Harvard Medical School

DNA methylation-based epigenetic clocks are used for estimating biological age and understanding disease-associated aging patterns. Using methylation profiles across 485,512 CpG sites, we implemented a machine learning pipeline to develop accurate and interpretable age prediction models. Feature selection was first performed with Ridge regression, identifying the top 1% most informative CpG sites. Twenty regression algorithms spanning linear, kernel-based, tree-based, boosting, and neural architectures were trained and optimized using Optuna. The four most accurate models (CatBoost, LightGBM, Gradient Boosting, and TabNet) were further analyzed with residual distributions and MAE by age group. Interpretability was ensured with SHAP analysis and feature importance rankings, revealing biologically relevant CpG sites associated with aging pathways. A weighted ensemble leveraging inverse squared MAE produced the most robust estimator and was applied to both control and nine disease groups, including Alzheimer's, Parkinson's, and Type 2 Diabetes. Comparative analyses across diseases demonstrated distinct epigenetic aging signatures, with disease-specific shifts in residual distributions and error profiles. These results highlight the potential of interpretable epigenetic clocks for advancing predictive medicine and deducing the biological basis of age-related disease.

Jialai She

Beyond Additivity: Sparse Isotonic Shapley Regression toward Nonlinear Explainability

Mentor: Prof. Gil Alterovitz, Harvard Medical School

Shapley values, a gold standard for feature attribution in Explainable AI, face two primary challenges. First, the canonical Shapley framework assumes that the worth function is additive, yet real-world payoff constructions—driven by non-Gaussian distributions, heavy tails, feature dependence, or domain-specific loss scales—often violate this assumption, leading to distorted attributions. Secondly, achieving sparse explanations in high-dimensional settings by computing dense Shapley values and then applying ad hoc thresholding is prohibitively costly and risks inconsistency. We introduce Sparse Isotonic Shapley Regression (SISR), a unified nonlinear explanation framework. SISR simultaneously learns a monotonic transformation to restore additivity—obviating the need for a closed-form specification—and enforces an L0 sparsity constraint on the Shapley vector, enhancing computational efficiency in large feature spaces. Its optimization algorithm leverages Pool-Adjacent-Violators for efficient isotonic regression and normalized hard-thresholding for support selection, yielding ease in implementation and global convergence guarantees. Analysis shows that SISR recovers the true transformation in a wide range of scenarios and achieves strong support recovery even in high noise. Moreover, we are the first to demonstrate that irrelevant features and inter-feature dependencies can induce a true payoff transformation that deviates substantially from linearity. Extensive experiments in regression, logistic regression, and tree ensembles demonstrate that SISR stabilizes attributions across payoff schemes, correctly filters irrelevant features; in contrast, standard Shapley values suffer severe rank and sign distortions. By unifying nonlinear transformation estimation with sparsity pursuit, SISR advances the frontier of nonlinear explainability, providing a theoretically grounded and practical attribution framework.

Brian Li

Transformer-Based Deep Learning for Multiclass Viral Classification in Metagenomic Sequencing

Mentor: Prof. Gil Alterovitz, Harvard Medical School

Identifying viruses within metagenomic datasets remains a central challenge in computational biology. With the threat of infectious diseases, such as the recent COVID-19 pandemic, pathogen detection and early viral discovery serves as an increasingly important problem. Alignment-based techniques like BLAST, which are computationally intensive, struggle with highly divergent sequences, high mutation rates within viral species, and the novel discovery of viruses. With machine learning tools, sensitivity and speed are increased, however most of these tools are still limited to binary classification of viruses versus non-viruses, offering little resolution at the level of viral families or classes. Convolutional neural networks (CNNs), such as the one used in VirDetect-AI (Zarate et al, 2025), have made it possible to create finer-grained models, but CNNs tend to struggle with capturing long-range dependencies which are a crucial component of the structure of viral genomes. Here, we introduce a transformer-based architecture that can better model global sequence relationships. This new model demonstrates robust learning capacity and strong performance in evaluations, demonstrating the utility of transformer-based approaches for multiclass viral classification.

Aaryan Arora

MT-BN: Multi-Scale Topological Bayesian Networks for Tractable, Interpretable Structure Learning in p»n

Mentor: Prof. Gil Alterovitz, Harvard Medical School

High-dimensional datasets with far more variables than samples ($p \gg n$) overwhelm classical flat Bayesian-network learners: their search space over DAGs grows super-exponentially and, by treating all nodes at one level, they ignore the modular, multi-scale organization that real systems exhibit, hurting both computational tractability and interpretability. We introduce MT-BN (Multi-Scale Topological Bayesian Network), a structure-learning framework that first infers an adaptive hierarchy of modules, then learns DAGs at L layers: a coarse graph among modules and fine graphs within modules. Connectivity is guided by topological priors, combining the stochastic-block prior that favors dense intra-module and sparse inter-module connections, and a log-normal degree prior that captures heterogeneous hub structure, while a hybrid inference algorithm pipeline yields scalable computation and calibrated edge posteriors. Let the p variables be partitioned into K modules with sizes $m_1, \dots, m_K$ and define $m := \max_i m_i$. Let $\mathcal{G}_s$ denote the set of labeled DAGs on s nodes, with $|\mathcal{G}_s|$ its cardinality. We prove that, under bounded module size m and K modules, MT-BN reduces the effective structure space from $2^{\binom{p}{2}}$ candidates to at most $|\mathcal{G}_K| \times |\mathcal{G}_m|^K$, an exponential shrinkage formalized in our search-space theorem. We demonstrate the effectiveness of MT-BN through ground-truth benchmarking on the DREAM5 network-inference challenge and case studies on tuberculosis gene-expression cohorts and financial markets (sector→industry→stock influences), showing the potential of MT-BN as a tool for structure learning problems in the $p \gg n$ regime while ensuring interpretability.

Jerry Xu

A Unified Protein Embedding Model with Local and Global Structural Sensitivity

Mentors: Prof. Gil Alterovitz, Harvard Medical School, and Dr. Shaojun Pei, Brigham and Women's Hospital

Structural comparison between proteins is key to many research tasks, including evolutionary analysis, peptidomimetics, and functional annotation. Traditional structure alignment tools based on three-dimensional protein structures, such as TM-Align, DALI, or ProBiS, are accurate, but they are computationally expensive and impractical at scale. Existing protein language models (PLMs), such as TM-Vec, improve computational efficiency but only capture global structural similarity, overlooking important motif-level structural details. In this talk, we propose a novel PLM consisting of a Siamese neural network, enabling efficient embedding-based structural comparison while also capturing both global and local structural similarity. Our model was trained on a dual loss function combining TM-score, a global similarity metric, and a variation of lDDT scores, a per-residue similarity metric. We tested against two datasets: a varied TM-score dataset from TM-Vec, and a high TM-score mutant dataset from VIPUR. Against these sets, our model achieved a TM-score MAE of 0.0741 and 0.0583, respectively, and a lDDT-score MAE of 0.0788 and 0.0038, respectively. Our model fulfills two key roles: first, it rapidly detects global structural differences. Second, it supports fine-grained structural assessments, improving sensitivity to subtle but functionally important structural changes.

Sophia Yan

Mapping Enhancer-Gene Interactions in the Human Brain Reveals Reduced Use of Enhancers for Synaptic Genes in Schizophrenia

Mentors: Dr. Nicole Rockweiler and Prof. Steven McCaroll, Broad Institute

Gene regulation is a critical process for maintaining cell type-specific functions. Concerted dysregulation of genes is present in many neurological and neuropsychiatric disorders such as in schizophrenia (SZ). SZ is a polygenic disorder characterized by hallucinations, delusions, and declining cognitive capacities, whose pathophysiological basis remains elusive. Here, we computationally explored one of the primary regulators of gene expression, enhancers, at single-nucleus resolution, to characterize their expression variation across cell types and SZ in the dorsolateral prefrontal cortex (dlPFC) of 180 human postmortem brains (93 control, 87 SZ).

We describe a strategy to use single nucleus RNA-seq (snRNA-seq) data to detect a canonical marker of active enhancers, enhancer RNA (eRNA). Compared to random intergenic RNA from nearby regions (as a control), these putative eRNAs show higher expression and distinctive bidirectional transcription, supporting snRNA-seq's ability to recognize activity of at least some enhancers. Mapping these putative enhancers to their target genes uncovered 66k enhancer-gene-cell-type combinations, 600-fold more than pairs expressed ubiquitously across cell-types, suggesting a highly cell-type-specific enhancer-gene regulation network.

A subset of these enhancer-gene pairs, which we term synaptic neuron and astrocyte program (SNAP) enhancer-gene pairs, showed differential expression changes in SZ. Risk for SZ is highly heritable, and hundreds of common genetic influences on risk have been found, but over 90% of SZ risk variants are located in the non-protein-coding regions of the genome, where they may act by influencing gene expression rather than protein sequence (e.g., by overlapping an enhancer). We found that enhancers recruited by SNAP showed a 1.25-fold enrichment in SZ risk variants ($p = 3.51e\text{-}8$; Chi-squared test) compared to non-SNAP enhancers. In glutamatergic (excitatory pyramidal) neurons, the SNAP enhancer-gene pairs downregulated in SZ were enriched for neuronal functions and include dopamine receptor genes, whereas the strongest pathway enriched in SZ upregulated pairs was negative regulation of amine transport, consistent with findings of reduced hypodopaminergic activity in the dlPFC in SZ.

The single-cell-resolution enhancer-gene pairs we have found provide insight into the complex regulatory networks in the brain, shedding light on how disruption of enhancer activity could be part of the underlying etiology of brain disorders. In addition, our framework for identifying putative active enhancers and enhancer-gene pairs is widely applicable to snRNA-seq analyses across tissues and species.

Session 13

Rio Schillmoeller

Dynamical Functionals on Ancient ARF and AC Ricci Flows

Mentor: Isaac Lopez, University of Chicago

First, we will introduce Riemannian manifolds and the Ricci flow equation. Then, we will briefly discuss the Perelman lambda functional. Finally, we will give a quick summary of our results on the dynamical lambda functional.

Arjun Agarwal

Semisimplifications of α_p-equivariant GL_n-modules

Mentor: Dr. Arun S. Kannan, DRW

Representation theory is a field of mathematics that allows us to represent elements of an algebraic object as linear maps of a vector space. This mapping of elements to linear maps, called a representation, lets us study these objects using the well-understood theory of linear algebra. Such actions arise naturally in physics, where representations can help study the symmetries and quantum states of fundamental particles.

One fundamental algebraic object that we are interested in studying in representation theory is the general linear group, $GL(n)$. The problem of determining the representations of $GL(n)$ has been solved in characteristic zero. But for positive characteristic, p, this remains an open problem and of much interest over the last fifty years. Moreover, the case of algebraic supergroups, $GL(m|n)$, is even harder by virtue of being a generalization. However, for positive characteristic, the Steinberg tensor product theorem helps reduce this problem to studying the restricted representations of the corresponding Lie superalgebra $\mathfrak{gl}(m|n)$.

In this talk, we discuss a process called semisimplification which allows us to transform representations of $\mathfrak{gl}(m+n(p-1))$ into representations for the Lie superalgebra $\mathfrak{gl}(m|n)$, the former being better understood. We provide explicit calculations in low rank and characteristic and observe in all cases that the semisimplification of a simple module is semisimple.

Kai Yamashita

Radon Transform for General Linear Groups over Finite Fields

Mentor: Ivan Motorin

A Radon transform is a linear operator that arises naturally from the representation theory of finite groups. In this talk we will introduce the Radon transform corresponding to a particular family of unipotent subgroups of a general linear group over a finite field. We relate it to Hecke algebras and present other miscellaneous results. Prerequisites include linear algebra and group theory.

Rachel Chen

The Dual Canonical Basis in the Spin Representation via the Temperley-Lieb Algebra

Mentor: Dr. Vasily Krylov, Harvard

The spin representation is an important object in both mathematics and physics that encodes the spin states of multiple-electron systems. It contains a standard basis, but also a dual canonical basis introduced by Lusztig that is important in many areas of algebra and geometry. The spin representation is a representation of the Temperley-Lieb algebra, an object generated by a basis of diagrams. In this presentation, we first introduce the spin representation, the Temperley-Lieb algebra, and the notion of a representation. Then, we introduce our main result, an explicit bijection between the basis of the diagrams in the Temperley-Lieb algebra and the dual canonical basis of the spin representation, a result which builds off of the work of Khovanov. Finally, we describe the significance of this result by outlining our other results on the spin representation and related objects then discussing our future research.

Emma Li

On the e-positivity of the Chromatic Symmetric Function of Left-melting Clique Chains

Mentor: Dr. Foster Tom, Dartmouth College

In 1995, Stanley introduced the chromatic symmetric function $X_G(x)$ as a generalization of the chromatic polynomial. Stanley and Stembridge conjectured that $X_G(x)$ is e-positive for unit interval graphs, which was then proven by Hikita. Shareshian and Wachs defined a generalization $X_G(x;q)$ of the chromatic symmetric function and also conjectured e-positivity for unit interval graphs. In this talk, we show the e-positivity of the chromatic symmetric function of left-melting cliques, and thus left-melting clique chains.

Aaron Lin

The Geometry of a Counting Formula for Deformations of the Braid Arrangement

Mentor: Neha Goregaokar, Brandeis

We consider real hyperplane arrangements whose hyperplanes are of the form $\{x_i - x_j = s\}$ for some integer s, which we call *deformations of the braid arrangement*. In 2018, Bernardi gave a counting formula for the number of regions of any deformation of the braid arrangement $\mathcal{A}$ as a signed sum over some decorated trees. He further showed that each of these decorated trees can be associated to a region R of the arrangement $\mathcal{A}$, and hence we can consider the contribution of a region to the signed sum. Bernardi showed that for *transitive* arrangements, the contribution of any region of the arrangement is 1. We remove the transitivity condition, showing that for *any* deformation of the braid arrangement the contribution of a region is 1. This provides an alternative proof of the original counting formula, further deepening our understanding of the geometry underlying the formula.

Jiayu (Jerry) Liu

Full-Twist Presentations for Fundamental Groups of Complements of Complexified Hyperplane Arrangements

Mentor: Prof. Nathan Williams, University of Texas Dallas

In this talk, we explore new presentations of fundamental groups of complements of complexified hyperplane arrangements by generalizing the classical Artin presentation for pure braid groups. Using Salvetti's combinatorial model, we construct generators from minimal galleries of regions and derive relations governed by rank-two subarrangements. We then introduce the notion of the full twist and conjecture a presentation of the fundamental group by equating all the reduced words for the full twist. Finally, we prove this conjecture for Coxeter arrangements of types A, B, and D.

Sophia Tatar

Sharp Constructions for the Szemerédi-Trotter Theorem

Mentor: Jiahui Yu

We will begin by exploring the Szemerédi-Trotter theorem in its various equivalent formulations and discussing its significance in incidence geometry. The focus will then shift to sharp constructions for the Szemerédi-Trotter theorem, tracing their development from the earliest examples to recent ones based on generalized arithmetic progressions.

Michael Iofin

Quasiconformal Normalization of Random Meromorphic Functions

Mentor: Prof. Sergiy Merenkov, City University of New York

In this talk, we introduce the notion of a Riemann surface spread over the sphere and its conformal type. Each simply connected such surface gives rise to a meromorphic function, whose Nevanlinna characteristic — a measure of its growth — controls much of its behavior, such as the density of zeroes. Our main result establishes parabolicity with probability one and provides almost sure bounds on the Nevanlinna characteristic for functions associated with a certain class of random surfaces having square grid nets.