MIT PRIMES

Photo: MIT News

SEVENTH ANNUAL CONFERENCE
MAY 2017

PRIMES 2017 Conference Schedule

Saturday, May 20

Mathematics

8:20 am Welcoming Remarks

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

8:50 am Session 1

Franklyn Wang, *Monodromy groups of indecomposable rational functions* (mentor Prof. Michael Zieve, University of Michigan)
Michael Ren, *On quasi-invariant polynomials* (mentor Xiaomeng Xu)
Swapnil Garg, *Hilbert series of the representation of Cherednik algebras* (mentor Alexey Pakharev, Northeastern University)
Megan Joshi, *Maximal self-intersection number of curves on surfaces* (mentor Prof. Moira Chas, SUNY at Stony Brook)

9:50 am: coffee break

10:10 am Session 2

Anlin Zhang, *Modelling epidemics on networks with cliques* (mentor Prof. Laura Schaposnik, University of Illinois at Chicago)
Kyle Gatesman, *An algorithmic and computational approach to optimizing gerrymandering* (mentor Prof. James Unwin, University of Illinois at Chicago)
Richard Xu, *Graph theory and tessellations* (mentor Prof. Sergiy Merenkov, CCNY – CUNY)
Kaiying Hou and Byung Yeon Rhee, *Continuum modelling of traffic system with autonomous vehicles* (mentor Andrew Rzeznik)

11:15 am: break

11:25 am Session 3

Zoe Levitt (PRIMES Circle), *An introduction to group theory and braids* (mentor Lara Booth)
Joshua Lee, *Coin games and 5-way scales* (mentor Dr. Tanya Khovanova)
Hyunjun Ahn, Benjamin Chen, Richard Chen, Ezra Erives, Jeremy Fleming, Michael Gerovitch, Tejas Gopalakrishna, Neil Malur, Nastia Polina, and Poonam Sahoo (PRIMES STEP students)
We are the CHOMPians (mentor Dr. Tanya Khovanova)
Pratik Alladi, Neel Bhalla, Nathan Sheffield, Tiancheng Song, Will Sun, Andrew The, Alan

Wang, Naor Wiesel, Kevin Zhang, Kevin Zhao (PRIMES STEP students) *Impartial combinatorial games* (mentor Dr. Tanya Khovanova)

12:30 pm: lunch break

1:30 pm <u>Session 4</u>

Ayush Agarwal, *Maps between critical groups of group representations* (mentor Christian Gaetz)
Michael Gintz, *Classifying graph Lie algebras* (mentor Dr. Tanya Khovanova)
Aaron Kaufer, *Low-dimensional d-Algebras* (mentor Lucas Mason-Brown)
Daniel Liu and Nathan Ramesh, *Verma modules of the Virasoro algebra* (mentor Siddharth Venkatesh)

2:35 pm: break

2:45 pm <u>Session 5</u>

Richard Zhou, *Pattern avoidance classes invariant under the modified Foata-Strehl action* (mentor Yan Zhuang, Brandeis University)
Dylan Pentland, *Coefficients of q-binomial coefficients modulo N* (mentor Younhun Kim)
Jason Chen, *Arithmetic properties of weighted Catalan numbers* (mentor Dmitry Kubrak)
Mihir Singhal, *Generalizations of Hall-Littlewood polynomials* (mentor Christopher Ryba)

3:45 pm: coffee break

4:00 pm <u>Session 6</u>

William Zhang, *Pattern avoidance on binary matrices* (mentor Jesse Geneson)
Louis Golowich, *Set-sequential trees* (mentor Chiheon Kim)
Michael Ma, *A generalization of Erdös-Szekeres to permutation pattern replacement* (mentor William Kuszmaul, Stanford University)
Wendy Wu and Andy Xu, *Second gonality of Erdös-Rényi random graphs* (mentor Guangyi Yue)

5:05 pm: break

5:15 pm <u>Session 7</u>

Sam Cohen and Aaditya Singh, *Generalization of some properties of discrete curve shortening flow* (mentor Ao Sun)
Grace Tian, *Multi-crossing numbers for knots* (mentor Jesse Freeman)
Gopal Goel, *Limits of interlacing eigenvalues in the tridiagonal β-Hermite matrix model* (mentor Andrew Ahn)
August Chen, *Folding, jamming, and random walks* (mentor Prof. Jayadev Athreya, University of Washington)

6:20 pm: break

6:30 pm <u>Session 8</u>

Jeffery Yu, *Jacobian groups of biconnected graphs* (mentor Dr. Dhruv Ranganathan)
Kaan Dokmeci, *On denesting radicals* (mentor Yongyi Chen)
David Darrow, *A near-optimal spectral method for simulating fluids in a cylinder* (mentors Prof. Alex Townsend, Cornell University, and Prof. Grady Wright, Boise State University)
Peter Rowley, *Random walks on a grid with a periodic boundary condition* (mentor Boya Song)

<u>Sunday, May 22</u>

Computer Science and Computational Biology

9:00 am <u>Welcoming Remarks</u>

Prof. Srini Devadas, Department of Electrical Engineering and Computer Science
Dr. Slava Gerovitch, PRIMES Program Director

9:15 am <u>Session 9</u>

Vivek Bhupatiraju, *Towards append-only authenticated dictionaries* (mentor Alin Tomescu)
Robert Chen, John Kuszmaul, and Yiming Zheng, *Keychat: Secure messaging via Bitcoin* (mentor Alin Tomescu)
Sanjit Bhat and David Lu, *Analyzing Tor's anonymity with machine learning* (mentor Albert Kwon)
Theodor Lukin Yelin, *AnonReddit: A strongly anonymous public forum* (mentor Albert Kwon)

10:30 am: coffee break

10:45 am <u>Session 10</u>

Anjali Saini, *Investigating the consensus algorithm* (mentor Ling Ren)
Nihar Sheth, *Investigating the scalability of Go's garbage collector in multicore environments* (mentor Cody Cutler)
Robert Cunningham, *An analysis of a directory entry cache in a high level language* (mentor Cody Cutler)

11:30 am: break

11:40 am <u>Session 11</u>

Zachary Steinberg, *Development of a new method for multicolor image segmentation of neuronal tissue in 20x expanded hydrogels* (mentor Daniel Goodwin)
Chao Cheng, *1.00 for life: Real-time analysis of computational thinking* (mentor Prof. John Williams)

Mayank Mali, Caleb Trotz, and Justin Yu, *Automated calibration and a real-time web-based control interface for fiber lasers* (mentor Michael Plotkin, IPG Photonics)

12:35 pm: Lunch break

1:35 pm <u>Session 12</u>

Harshal Sheth and Andrew Sun, *Tarpan: A router that supports evolvability* (mentor Dr. Raja Sambasivan, Boston University)
Henry Heffan and Shashvat Srivastava, *Mo-nero, Mo-problems: Defending Monero against temporal analysis* (mentors Ethan Hellman, Dr. Jason Hennessey, Kyle Hogan, and Dr. Mayank Varia, Boston University)
Vinjai Vale, *Vision as inverse graphics: Machine learning techniques towards a program-based model for scene understanding* (mentor Kevin Ellis)

2:35 pm: break

2:50 pm <u>Session 13</u>

Dr. Gil Alterovitz, Harvard-MIT Division of Health Sciences and Technology, Introductory remarks
Adithya Vellal, *Enrichment and analysis of sequence motifs in genomic variant calls* (mentor Dr. Gil Alterovitz)
Andrew Gritsevskiy, *Structural gene expression* (mentor Dr. Gil Alterovitz)
Eric You, *Analyzing compression: Using gene sets to find patterns in genomic compression* (mentor Dr. Gil Alterovitz)

3:45 pm: break

3:55 pm <u>Session 14</u>

Makiah Bennett and Jack Flahive, *Precision cancer medicine web application* (mentor Dr. Gil Alterovitz)
James Jusuf, *A versatile algorithm for finding patterns in large cancer cell line data sets* (mentor Dr. Mahmoud Ghandi, Broad Institute)
Kevin Hu, *An analysis of MDM4 alternative splicing and effects across cancer cell lines* (mentor Dr. Mahmoud Ghandi, Broad Institute)
Kalyan Palepu and Andrew Zhang, *Compression of genomic variants using convolutional autoencoders* (mentor Dr. Gil Alterovitz)

SESSION 1

Franklyn Wang

The classification of indecomposable rational functions by ramification analysis

Mentor: Prof. Michael Zieve, University of Michigan

Project suggested by Prof. Michael Zieve

The most important geometric invariant of a degree-n complex rational function $f(X)$ is its *monodromy group*, which is a certain group of permutations of n objects. Many properties of $f(X)$ can be determined once one knows this monodromy group. We determine all degree-n rational functions whose monodromy group is not A_n or S_n, under some hypotheses. These are the rational functions which do not behave similarly to random rational functions in various contexts. This refines and extends previous work by Zariski, Thompson, Guralnick, Aschbacher, and others.

Michael Ren

On quasi-invariant polynomials

Mentor: Xiaomeng Xu

Project suggested by Prof. Pavel Etingof

We extend certain results of Braverman, Etingof, and Finkelberg and study spaces of twisted quasi-invariant polynomials, which are generalizations of symmetric polynomials. These spaces are modules over the ring of symmetric polynomials, and we determine their structure in certain cases. In particular, we show that these modules are always free and compute the corresponding Hilbert series in those cases.

Swapnil Garg

Hilbert series of the representation of Cherednik algebras

Mentor: Alexey Pakharev, Northeastern University

Project suggested by Prof. Pavel Etingof

We investigate the polynomial representation of the type A_{n-1} rational Cherednik algebra over a field of positive characteristic p. We look at the representation as a graded algebra and investigate the Hilbert series of its irreducible quotient. Recently, Devadas and Sun determined the Hilbert series for the generic case with $p \mid n$. We look at $p \mid n$ in some specific cases, and also at the $c = 0$ case.

Megan Joshi

Maximal self-intersection number of curves on surfaces

Mentor: Prof. Moira Chas, SUNY at Stony Brook

Project suggested by Prof. Moira Chas

A surface S with boundary can be associated with a word in a certain alphabet as it is cut into a planar model and the two edges corresponding with each cut are labeled with corresponding letters. These labels then read out what is defined as the surface word. The sequence of edges through which a curve (up to deformation on the surface S) passes is defined as the curve word. Two numbers can be associated with a deformation class of curves: word length and self-intersection number. Word length is defined as the number of letters within the curve word, and the self-intersection number is the smallest number of times that a curve within its deformation class intersects itself. Using the output of a known algorithm to compute self-intersection number, this project establishes the following conjecture: the maximal self-intersection number of classes of curves of word length L approaches a quadratic polynomial in L as L approaches infinity. This conjecture has already been proven for two surfaces of small topological complexity.

Session 2

Anlin Zhang

Modelling epidemics on networks with cliques

Mentor: Prof. Laura Schaposnik, University of Illinois at Chicago

Project suggested by Prof. Laura Schaposnik

Infection outbreaks have been studied by mathematicians for more than a century. Mathematical models on regular tree graphs are particularly effective for simple networks, as the symmetry gives elegant results. After reviewing the existing models, we

shall show the first steps towards extending the results of the current model on regular trees to include cliques. Specifically, we shall consider bounds for the duration of the outbreaks, and how the placement of the root affects probabilities of infection and recovery. Finally, we shall consider the removal of infection paths, or edges, to describe safe zones in which individuals cannot be infected.

Kyle Gatesman

An algorithmic and computational approach to optimizing gerrymandering

Mentor: Prof. James Unwin, University of Illinois at Chicago

Project suggested by Prof. James Unwin

In a representative democracy, geographic territories are divided into districts, each of which votes to elect a candidate to political office. Aside from the requirements that all voting districts within a territory must be connected and of equal population, there are usually no formal mandates to regulate the redrawing of district lines. Gerrymandering is the act of constructing voting districts with a specific outcome in mind, most typically the reelection of politicians from the incumbent party. We investigate previously studied models and algorithms for the redistricting process, evaluating their strengths and shortcomings.

In subsequent work we will construct a method for optimal gerrymandering of a territory and, in appeal to the moral stance against gerrymandering, we will develop a computational algorithm to detect gerrymandered voting districts based on territorial demographic and geographic data. This improved model may be used in the future to promote fair elections based on equitable voting districts.

Richard Xu

Graph theory and tessellations

Mentor: Prof. Sergiy Merenkov, CCNY CUNY

Project suggested by Prof. Sergiy Merenkov

In this talk we discuss the classical results regarding the relationships between graphs and tessellations. We then present our extremal problem on planar graphs and discuss how the solution to this problem corresponds to a square tiling of a rectangle, and the implication of this correspondence in both Graph Theory and Tiling Problems. Finally, we discuss an algorithm that calculates the solution to this extremal problem.

Kaiying Hou and Byung Yeon Rhee

Continuum modelling of traffic system with autonomous vehicles

Mentor: Andrew Rzeznik

Project suggested by Andrew Rzeznik

Traffic flow is an important topic in applied mathematics and has been studied by mathematicians for years. A common way to describe traffic flow is through continuum modelling, which analyses the continuous density of cars on the road rather than focusing on each individual car. Although this method does not consider the discreteness of each car, it does reveal important aspects of traffic flow. With the advent of autonomous vehicles (self-driving cars), it is then important to learn how autonomous vehicles will influence traffic flow. In our research, we consider combined traffic systems with both human and autonomous vehicles and attempt to find rules for autonomous vehicles that optimize both efficiency and human safety.

SESSION 3

Zoe Levitt, PRIMES Circle

An introduction to group theory and braids

Mentor: Lara Booth

In this talk, we discuss properties of the braid group. We first lay foundation by defining groups and discussing some of their basic properties. This includes a discussion of subgroups, generators, equivalence relations, homomorphisms, the center of a group, and the permutation group. We then define the braid group and lay out some of its properties, including the homomorphism from the braid group to the permutation group. Finally, we consider the center of the braid group. It is known that the center of the braid group is generated by a specific element, τ^n. In this talk, we prove tau^n is in fact in the center.

Joshua Lee

Coin games and 5-way scales

Mentor: Dr. Tanya Khovanova

Project suggested by Dr. Tanya Khovanova

We will discuss the coin-weighing problem with a 5-way scale, which has two more outcomes than the usual scale. These two outcomes correspond to when there are at least two more fake coins on one side than the other, and thus the 5-way scale is an upgrade from the regular scale. We investigate what improvements this change can bring from the

3-way scale, and attempt to find the optimal strategy for the 5-way scale problem with a strategy and an information theoretical bound. We will then discuss a game involving the counterfeit coin problem. Given one lighter fake coin among a number of otherwise identical coins, two players with full knowledge of the coins take turns weighing coins on a two pan scale, under the conditions that they reveal some information every turn to an observer with no prior knowledge and that the observer cannot determine what the fake coin is at any point. We will present the Grundy numbers, or the value equivalent to the number of stones in a nim heap, associated with every position of this game, as well as for a modified version of the game where the fake coin can either be heavier or lighter.

Hyunjun Ahn, Benjamin Chen, Richard Chen, Ezra Erives, Jeremy Fleming, Michael Gerovitch, Tejas Gopalakrishna, Neil Malur, Nastia Polina, and Poonam Sahoo
(PRIMES STEP students)

We are the CHOMPians

Mentor: Dr. Tanya Khovanova

Project suggested by Dr. Tanya Khovanova

We study two variations of Nim and Chomp which we call *Monotonic Nim* and *Diet Chomp*. In Monotonic Nim the moves are the same as in Nim, but the positions are non-decreasing numbers as in Chomp. Diet-Chomp is a variation of Chomp, where the total number of squares removed is limited.

Pratik Alladi, Neel Bhalla, Nathan Sheffield, Tiancheng Song, Will Sun, Andrew The, Alan Wang, Naor Wiesel, Kevin Zhang, and Kevin Zhao
(PRIMES STEP students)

Impartial combinatorial games

Mentor: Dr. Tanya Khovanova

Project suggested by PRIMES STEP students

PRIMES STEP students invented a variety of impartial combinatorial games. The games are played on graphs, piles, and grids. We will describe the games, their winning positions, optimal strategies, and other interesting properties of the games.

SESSION 4

Ayush Agarwal

Maps between critical groups of group representations

Mentor: Christian Gaetz

Project suggested by Prof. Vic Reiner, University of Minnesota, Twin Cities

Critical groups of graphs are well-studied invariants of finite directed graphs. It is known that covering maps between graphs induce surjections on critical groups, and Reiner and Tseng have given a combinatorial description of the kernel of this induced map in terms of "voltage critical groups". Recently, Benkart, Klivans, and Reiner defined critical groups corresponding to faithful representations of finite groups. In this project we study which operations on representations induce maps on critical groups and study their properties, with the goal of similar combinatorial understanding in special cases. We also hope to use properties of these maps and results of Gaetz to better understand the structure of critical groups for representations of the symmetric group.

Michael Gintz

Classifying graph Lie algebras

Mentor: Dr. Tanya Khovanova

Project suggested by Dr. Tanya Khovanova

We define a Lie algebra on a simple graph and consider the problem of determining its structure. We create a series of alterations of a graph which create a graph with an isomorphic Lie algebra. We also determine a set of conditions which prove that the Lie algebra of a graph is isomorphic to the direct sum of two copies of the Lie algebra of an induced subgraph with a specific vertex removed.

Aaron Kaufer

Low-dimensional d-Algebras

Mentor: Lucas Mason-Brown

Project suggested by Prof. Pavel Etingof

The study of commutative algebras plays a central role throughout many different areas of mathematics, including algebraic number theory and algebraic geometry. A commutative algebra is a vector space A over a field F, together with an associative bilinear multiplication operation which satisfies the commutativity law $ab = ba$. A d-algebra is a

generalization of a commutative algebra in a different category. Specifically, it is an associative algebra A with a linear derivation $d : A \to A$ such that $d^2 = 0$ and $ab = ba + d(b)d(a)$. In this presentation, we classify noncommutative d-algebras up to dimension 7.

<div align="center">

Daniel Liu and Nathan Ramesh

Verma modules of the Virasoro algebra

Mentor: Siddharth Venkatesh

Project suggested by Prof. Andrei Negut

</div>

The representation theory of Lie algebras and their quantizations is a fundamental subject in modern algebra. These algebras have strong connections to quantum mechanics and algebraic geometry. In this paper, we study specific representations of some quantizations of the Virasoro algebra. These representations, called Verma modules, are building blocks of the category of all representations. Irreducible factors of the Verma modules correspond to singular vectors inside these modules. We compute explicit formulas for these singular vectors in the PBW basis.

<div align="center">

SESSION 5

Richard Zhou

Pattern avoidance classes invariant under the modified Foata-Strehl action

Mentor: Yan Zhuang, Brandeis University

Project suggested by Prof. Olivier Bernardi, Brandeis University

</div>

For any pattern set $\Sigma = \{\sigma_1, \sigma_2, \ldots \sigma_n\}$, let $Av(\Sigma)$ denote the set of all permutations of any length that avoids every pattern in Σ. We say that $Av(\Sigma)$ is invariant under valley hopping if for any permutation $\pi \in Av(\Sigma)$, any valley hop $\pi' = Hop_j(\pi) \in Av(\Sigma)$. We provide a complete classification of all permutations σ for which $Av(Hop(\sigma))$ is invariant under valley hopping, where $Hop(\sigma)$ denotes the hop equivalence class of Σ. For alternating permutations σ, we bound the size of any pattern set Σ containing alternating permutations of the same length as σ for which $Av(\Sigma)$ is invariant under valley hopping and show that Σ must contain subsets with certain properties. For any σ, we provide a construction for a nontrivial Σ such that $\sigma \in \Sigma$ and $Av(\Sigma)$ is invariant under valley hopping.

Dylan Pentland

Coefficients of q-binomial coefficients modulo N

Mentor: Younhun Kim

Project suggested by Prof. Richard Stanley

Let $\begin{bmatrix} n \\ k \end{bmatrix}_q$ be a q-binomial coefficient. Stanley conjectured that the function

$$f_{k,R}(n) = \left| \left\{ \alpha : [q^\alpha] \begin{bmatrix} n \\ k \end{bmatrix}_q \equiv R \pmod{N} \right\} \right|$$

is quasi-polynomial for $N = 2$. We show that this is fact true for any integer N with a quasi-period explicitly given by a function $\pi'_N(k)$. Additionally, we find a form for the generating function and show that the denominator is a product of cyclotomic polynomials $\Phi_d(x)$. As a byproduct, we prove some results about partitions with at most k parts modulo N and also derive asymptotic estimates of the quasi-period $\pi'_N(k)$ and conjectured minimal period.

Jason Chen

Arithmetic properties of weighted Catalan numbers

Mentor: Dmitry Kubrak

Project suggested by Prof. Alexander Postnikov

The weighted Catalan numbers are a generalization of the Catalan numbers that enumerate many combinatorial objects. For instance, if we use the weights $\mathbf{b} = (1^2, 3^2, 5^2, \dots)$, then we get the number of plane Morse links of order n. In this paper, we study the properties of weighted Catalan numbers modulo a prime p. We provide sufficient and necessary conditions for the weights in order for the sequence to be periodic modulo p, and show that the period depends on the roots of a certain polynomial in $\mathbb{F}_p[x]$.

Mihir Singhal

Generalizations of Hall-Littlewood polynomials

Mentor: Christopher Ryba

Project suggested by Prof. Alexei Borodin

Hall-Littlewood polynomials are important functions in various fields of math, and can be defined combinatorially using a model of path ensembles. Wheeler and Zinn-Justin applied a reflection construction to this model to obtain an expression for type-BC Hall-Littlewood polynomials. Borodin applied a single-parameter deformation to the model

and obtained a formula for generalized Hall-Littlewood polynomials. In the project proposal, Borodin asked whether a similar generalization could be applied to type-BC Hall-Littlewood polynomials. We present the model and expressions for previously obtained polynomials, in addition to an expression for generalized type-BC Hall-Littlewood polynomials.

SESSION 6

William Zhang

Pattern avoidance on binary matrices

Mentor: Jesse Geneson

Project suggested by Jesse Geneson

Pattern avoidance is an important topic in graph theory and combinatorics, and has applications in bounding the Stanley-Wilf limits of permutations and in robot motion planning. Given binary matrices A and B, if some submatrix of A can be transformed into B by changing some ones to zeros, then A contains B. Otherwise, A avoids B. The extremal function $\mathrm{ex}Bn$ denotes the maximum possible number of ones in an $n \times n$ binary matrix that avoids B.

Given a permutation matrix P, it was proven by Marcus and Tardos that $\lim_{n \to \infty} \frac{\mathrm{ex}Pn}{n}$ exists, which we denote $c(P)$.

Using the probabilistic method, we extend a result from Geneson and Tian, and we find asymptotic lower bounds on $\mathrm{ex}Bn$ for all binary matrices B. We also extend a result from Cibulka and Kyncl to d-dimensional permutations.

Louis Golowich

Set-sequential trees

Mentor: Chiheon Kim

Project suggested by Chiheon Kim

A graph is called *set-sequential* if its vertices can be labeled with distinct nonzero vectors in $\mathbb{F}_2^n$ such that when each edge is labeled with the sum $(\mathrm{mod}\ 2)$ of its vertices, every nonzero vector in $\mathbb{F}_2^n$ is the label for either a single vertex or a single edge. We resolve certain cases of a conjecture of Balister, Győri, and Schelp in order to show many classes of trees to be set-sequential. We show that all caterpillars T of diameter k such that $k \leq 18$ or $|V(T)| \geq 2^{k-1}$ are set-sequential, where T has only odd-degree vertices and $|T| = 2^{n-1}$ for some positive integer n. We also present a new method of recursively constructing set-sequential trees.

Michael Ma

A generalization of Erdős-Szekeres to permutation pattern replacement

Mentor: William Kuszmaul, Stanford University

Project suggested by William Kuszmaul

The Erdős-Szekeres Theorem is a classic combinatorial result in pattern avoidance that is often used to demonstrate induction. However here a generalization of it to the area of pattern replacement is given along with an introduction to the field of pattern replacement. This project was proposed by William Kuszmaul who has done past research in the area of permutation pattern replacement equivalence.

Wendy Wu and Andy Xu

Second gonality of Erdős-Rényi random graphs

Mentor: Guangyi Yue

Project suggested by Dr. Dhruv Ranganathan

In this talk, we consider the asymptotic behavior of the second gonality of an Erdős-Rényi random graph and provide upper bounds for the second gonality with a simple probabilistic method, and a more sophisticated approach using the Riemann-Roch Theorem. Our results suggest that the asymptotic behavior of the second gonality of an Erdős-Rényi random graph is equal to the number of vertices under a certain restriction on the probability.

Session 7

Sam Cohen, and Aaditya Singh

Generalization of some properties of discrete curve shortening flow

Mentor: Ao Sun

Project suggested by Ao Sun

Discrete curve shortening flow is the discrete version of curve shortening flow. Some properties of curve shortening flow still holds for discrete version, some other properties don't. In this presentation we will talk about two generalizations of famous results of curve shortening flow to discrete version. One concerns the motion of infinite graphs, and the other one concerns the convexifying property of closed curves.

Grace Tian

Multi-crossing numbers for knots

Mentor: Jesse Freeman

Project suggested by Prof. Colin Adams, Williams College

We study the projections of a knot K that have only n-crossings. The n-crossing number of K is the minimum number of n-crossings among all possible projections of K with only n-crossings. We obtain new results on the relation between n-crossing number and $(2n - 1)$-crossing number for every positive even integer n.

Gopal Goel

Limits of interlacing eigenvalues in the tridiagonal β-Hermite matrix model

Mentor: Andrew Ahn

Project suggested by Prof. Vadim Gorin

A random matrix is a matrix whose entries are given by random variables. Given a random Hermitian matrix, one can interpret it as a random stepped function whose graph is a continuous function with slopes ± 1. For a class of random matrices, known as Wigner matrices, it is known that the associated stepped functions converge, after rescaling, to the Logan-Shepp curve which comes up in asymptotic representation theory. We will provide some background for the connection between random matrices and random stepped functions. We will then provide some of our results in this direction for the random tridiagonal β-Hermite ensemble.

August Chen

Folding, jamming, and random walks

Mentor: Prof. Jayadev Athreya, University of Washington

Project suggested by Prof. Jayadev Athreya

We consider the limiting case of multiple ball systems in two dimensions, or jammed billiard ball configurations, where all the balls have equal mass and are externally tangent to adjacent balls. Suppose for simplicity that the collisions are perfectly elastic and so no energy is lost in collisions. We wish to describe the end state of such configurations. First, we analyze this end state in some special cases and outline approaches to describing the general end state. Furthermore, describing the end state of such billiard ball configurations naturally gives way to a random walk we denote as Folding. We characterize the orbit and transition matrix of Folding in two dimensions and find formulas for its stationary measure in special cases, while analyzing potential simplifications, analysis and approaches in arbitrary dimensions.

SESSION 8

Jeffery Yu

Jacobian groups of biconnected graphs

Mentor: Dr. Dhruv Ranganathan

Project suggested by Dr. Dhruv Ranganathan

The Jacobian construction associates a finite abelian group to every finite graph. We study the corresponding inverse problem of determining which groups can be obtained in this fashion. Given a finite graph, we consider integer linear combinations of the vertices. Known as chip configurations, they form equivalence classes under chip-firing moves, and the set of equivalence classes of degree 0 form the Jacobian group. Since every finite graph can be decomposed into biconnected components, we focus on which groups are attainable as Jacobians of biconnected graphs.

Kaan Dokmeci

On denesting radicals

Mentor: Yongyi Chen

Project suggested by Dr. Stefan Wehmeier, Mathworks

The talk gives an introduction to radical denesting. It explains how denesting radicals works. Then, it covers what a field is and how field extensions are defined. Using these definitions, theorems are presented on radical denesting. These theorems are then used to prove a statement about radicals in radical denesting, and how they lead to future research. In particular, they suggest that radicals can be denested in all cases using Diophantine equations.

David Darrow

A near-optimal spectral method for simulating fluids in a cylinder

Mentors: Prof. Alex Townsend, Cornell University, and
Prof. Grady Wright, Boise State University

Project suggested by Prof. Alex Townsend and Prof. Grady Wright

A spectral method is described for approximating solutions to the incompressible Navier-Stokes equations in the cylinder. The method described addresses three major limitations of current methods: the clustering of discretization points near the centerline of the cylinder, non-optimal computational cost, and a low order of accuracy. The new method

evenly resolves all parts of the cylinder, dramatically reduces computation time by requiring at most $\mathcal{O}(N \log N)$ operations per time step (where N is the total number of discretization points), and has spectral accuracy.

Peter Rowley

Random walks on a grid with a periodic boundary condition

Mentor: Boya Song

Project suggested by Prof. Jörn Dunkel

In this project we analyze the Markov chain formed by the random walk on a $2n + 1 \times 2n + 1$ grid with a periodic boundary condition, for $n > 0$. We examine the eigenvalues of the transition matrix for this process, specifically showing that there are exactly $\binom{n+2}{n}$ distinct eigenvalues for the matrix. We also examine the implications of removing one point from the grid. More specifically, the probability of moving from one point to all available neighboring points are still same and sum up to 1, whereas the probability of moving to or away from a removed point is 0.

SUNDAY, MAY 21
COMPUTER SCIENCE AND COMPUTATIONAL BIOLOGY

SESSION 9

Vivek Bhupatiraju

Towards append-only authenticated dictionaries

Mentor: Alin Tomescu

Project suggested by Alin Tomescu

The distribution of accurate public keys is becoming increasingly important in an age of certificates and secure messaging. However, trusting a directory to accurately publish and retrieve public keys can be unsafe, as the directory can collude with an adversary or get compromised. One way to abstract an auditable and secure directory is as an append-only dictionary, which can provide proofs of partial membership, complete membership, and consistency. In this talk, we outline the meaning of these proofs and discuss two previous attempts at a full AAD. We also discuss a new implementation of an AAD and its related complexities and tradeoffs.

Robert Chen, John Kuszmaul, and Yiming Zheng

Keychat: Secure messaging via Bitcoin

Mentor: Alin Tomescu

Project suggested by Alin Tomescu

We present Keychat, a secure messaging system using the Bitcoin blockchain. Secure messaging across a network of many clients still remains a very difficult challenge. If a communication app's central key directory is compromised by a malicious adversary, then this adversary can equivocate about the public key directory to different clients. Thus an adversary can impersonate a client. Keybase, a public key directory, periodically publishes its Merkle Root, cryptographically signed by the Keybase server, to the Bitcoin blockchain. This essentially makes equivocation as difficult as forking the Bitcoin blockchain, which is nearly impossible. With Keychat, we are using the Keybase directory to implement a secure chat application for users with Keybase accounts. We use Catena to make auditing the Keybase directory more efficient, making KeyChat more secure. We show that KeyChat allows for the efficient exchange of secure messages across a network of many clients.

Sanjit Bhat and David Lu

Analyzing Tors anonymity with machine learning

Mentor: Albert Kwon

Project suggested by Albert Kwon

In a time where more and more governments are coming out to restrict their citizens internet access, anonymity is crucial to protecting an individual's freedom of expression. Tor, the most popular anonymous browser, attempts to provide individuals with this freedom. However, Tors anonymity protections can be circumvented using website fingerprinting (WF) attacks. In this talk, we explore the validity of WF attacks, the impact of various features on WF accuracy, and an application of WF to content-based website classification. Our results indicate that Tor remains vulnerable to WF attacks despite several proposed defenses against WF. Moreover, we identify packet ordering as a particularly identifying feature for the adversary. Finally, we achieve a high accuracy for classifying websites into different categories, a particularly useful attack for adversaries who may not need to know the specific website a client is visiting.

Theodor Lukin Yelin

AnonReddit: A strongly anonymous public forum

Mentor: Albert Kwon

Project suggested by Albert Kwon

AnonReddit is an implementation of a strongly anonymous public forum. While current forums provide a very useful and necessary platform for sharing information, several problems arise from the direct and non-optional attribution of a user's identity to their work. AnonReddit aims to solve this problem, while still retaining features necessary for a forum that normally require knowledge of a user's identity.

Through an implementation of a so-called linkable ring signature construct, AnonReddit is able to provide complete anonymity while eliminating issues such as double-voting and ensuring a user is registered, which normally depend on breaking a users anonymity. We show that AnonReddit scales efficiently to a user-base of 100,000 with low network and time overhead.

SESSION 10

Anjali Saini

Investigating the consensus algorithm

Mentor: Ling Ren

Project suggested by Ling Ren

Consensus is fundamental problem that has been investigated in distributed computing. The main goal is to attain agreement on a single data value among a system of processes. However, some processes in the system can experience some sort of faultiness that distorts their ability to send and receive messages, thus making it uncertain for those processes to agree on the data value. Therefore, in such a system, consensus protocols must be resilient and have fault tolerance to these failures.

Many consensus algorithms, such as Paxos and Practical Byzantine Fault Tolerance, have generally been used in smaller scale with usually no more than 10 processes. Our aim is to increase this scalability and move onto a greater level where many processes are used, such as the internet.

Nihar Sheth

Investigating the scalability of Go's garbage collector in multicore environments

Mentor: Cody Cutler

Project suggested by Prof. Frans Kaashoek

Garbage collection is an important feature of many high level programming languages. It enables developers to focus on application development rather than memory management and reduces the risk of hard-to-debug issues such as memory leaks. However, as the trend towards parallel programming continues, it is increasingly important that garbage collectors are scalable in multicore environments. The purpose of this talk is to test the parallel scalability of Go's garbage collector with multiple live data structures. Memory allocated in arrays and binary search trees is benchmarked, and results show that the Go garbage collector is not as scalable as hoped for. It is hypothesized that the issue has to do with contention on Go's mheap and mcentral data structures, which are involved in memory allocation, but further investigation is needed to confirm this.

Robert Cunningham

An analysis of a directory entry cache in a high level language

Mentor: Cody Cutler

Project suggested by Prof. Frans Kaashoek

Operating systems are conventionally written in C, such that they might squeeze maximum performance out of the available hardware. The price for this performance, however, is that management of a number of low level systems, such as memory allocation and deallocation, falls to the programmer, instead of being automatically managed by the language. This additional programmer overhead is often responsible for the creation of subtle bugs, which are unacceptable due to the consistency requirements on modern operating systems. Thus, we examine the suitability of a higher level language for construction of critical parts of operating systems. We find that by using Go, which supports garbage collection, has rich core libraries, and enforces stricter stylistic constraints than C, the implementation of a concurrent directory entry cache can be significantly simplified. Notably, much of the complexity around reference counting, memory management, memory barriers, and pointer math, all of which is necessary in C, is automatically managed by Go's garbage collector or standard library. However, in addition to an unknown performance cost, we pay for these benefits by giving up some of the clever and widely used idioms which C permits through preprocessor macros, but Go disallows in the name of transparency. In sum, significant parts of the implementation of a concurrent directory entry cache can be drastically simplified by using Go.

SESSION 11

Zachary Steinberg

Development of a new method for multicolor image segmentation
of neuronal tissue in $20\times$ expanded hydrogels

Mentor: Daniel Goodwin

Project suggested by Prof. Ed Boyden

In order to explore the brain's functions, one must first be able to identify connections between neighboring neurons. This subfield of neuroscience, known as connectomics, has historically worked for the past 30 years with black-and-white electron microscope data. Now that multicolor data is available at 20nm resolution via Expansion Microscopy (ExM), creating an automated, reliable algorithm that can handle the future petabytes of neural tissue data is an open problem. Here we outline an approach to differentiate between neurons in a $20\times$ expanded hippocampus slice expressing Brainbow fluorescent proteins. We first use a neural network as a mask to filter data, oversegment in color space

to create supervoxels, and finally merge those supervoxels together to reconstruct the 3D volume for an individual neuron.

Chao Cheng

1.00 for life: Real-time analysis of computational thinking

Mentor: Prof. John Williams

Project suggested by Prof. John Williams

The advent of the Internet has given rise to an explosion of ideas, opportunities, and information. Today, educational methods are still adapting and improving according to the latest innovations in web technology. Most models of education now involve face-to-face interactions, as well as online assignments and materials. However, the online technology can be used not only as a teaching utility, but also as a way to improve the quality of the course. By developing a system to grade student work immediately upon completion, and recording data collected from student usage of online resources such as supplementary videos and Q&A forums, we can analyze the effectiveness of course materials and predict student performance. Ultimately, the analysis of student interaction with online learning will allow instructors to fine-tune their courses to improve efficacy and better student education.

Mayank Mali, Caleb Trotz, and Justin Yu

Automated calibration and a real-time web-based control interface for fiber lasers

Mentor: Michael Plotkin, IPG Photonics

Project suggested by Michael Plotkin

Fiber lasers are lasers that generate their beam using doped optical fibers. Currently, fiber lasers are calibrated manually during the manufacturing process, which can lead to inefficient laser operation. Automating this process makes the laser more power efficient. We propose a web-based software framework to handle automatic calibration. This framework will also make it easier to update operational parameters and operate the laser itself.

SESSION 12

Harshal Sheth and Andrew Sun

Tarpan: A router that supports evolvability

Mentor: Dr. Raja Sambasivan, Boston University

Project suggested by Dr. Raja Sambasivan

Tarpan is an extension to the Border Gateway Protocol that allows the Internet to incrementally transition to a new routing protocol to eventually replace BGP. It provides safe transport of the routing data of any number of new routing protocols in-band within the existing BGP advertisement, and an API for easy implementation of new protocols. Tarpan also adds the ability to communicate out-of-band if necessary to transport routing information over networks that do not yet support the protocol. The Wiser protocol was implemented using this framework, and found that Tarpan has a minimal effect on BGP advertisement processing performance.

Henry Heffan and Shashvat Srivastava

Mo-nero, Mo-problems: Defending Monero against temporal analysis

Mentors: Ethan Hellman, Dr. Jason Hennessey, Kyle Hogan,
and Dr. Mayank Varia, Boston University

Project suggested by Ethan Hellman, Dr. Jason Hennessey, Kyle Hogan,
and Dr. Mayank Varia

One of the privacy-focused cryptocurrencies, designed to improve anonymity, is Monero, which attempts to hide who created transaction using several features. Among these features are confidential transactions, which hide the amount that is being transferred, stealth addresses, which make it harder to track to whom money is being payed, and mixins, which make it harder to know the origin of the money being spent in a transaction. In this talk, we focus in particular on mixins. Significant weaknesses currently exist in mixins. Most importantly, temporal attacks, which take advantage of a weakness in the selection of mixins in Monero, are extremely effective, and can greatly erode the privacy offered by mixins. We discuss both theoretical and actual attacks on Monero. We then propose several countermeasures and analyze their effectiveness against the attacks we have discussed.

Vinjai Vale

Vision as inverse graphics: Machine learning techniques towards a program-based model for scene understanding

Mentor: Kevin Ellis

Project suggested by Prof. Armando Solar Lezama

One of the most significant and complex problems in computer vision is scene understanding. Scene understanding involves recognizing and segmenting each object and object component in an image, as well as understanding how the various objects and components interact to yield some broader context about the scene. The conventional approach to scene understanding is to identify various objects in the scene and infer the context of the scene based on the types and arrangements of the objects. However, these approaches lack a compositional representation of the scene where objects are represented by putting together simpler parts and attributes. This research focuses on developing a new method of scene understanding following the paradigm of "vision as inverse graphics," in which vision is formulated as tuning the parameters of a graphics model that, when rendered, closely approximates the original scene. The proposed graphics model takes the form of a computer program written in a graphics programming language. That is, when given a scene, the software will generate a program that, when compiled and rendered, will yield a virtual model that approximates the original scene. Such a model would have an ingrained compositionality inherited from the nature of object-oriented programming. This would enable scene understanding on a much deeper level, which is vital for many applications such as robotics. This research explores the potential for a evolutionary algorithm to tackle a simplified version of this problem, where scenes are black-and-white and are comprised only of lines and circles.

SESSION 13

Adithya Vellal

Enrichment and analysis of sequence motifs in genomic variant calls

Mentor: Dr. Gil Alterovitz

Project suggested by Dr. Gil Alterovitz

The rising popularity of Next Generation Sequencing (NGS) technology has resulted in genomic data becoming continually cheaper and increasingly accessible. Genomic data is becoming especially useful in developing a better understanding of gene regulation, where genomic sequence motifs are crucial to understanding behavior such as transcription factor binding. The wet lab experiments currently necessary to obtain sequence motifs representing transcription factor binding sites (TFBSs) have proven to be expensive

and time-taking. Resultantly, there is a growing need for computational methods which can determine whether new motifs are TFBSs or not without experimental justification.

We focus on using sequence motif identification and enrichment techniques to analyze genomic variant data. Specifically, we identify the sequence motifs present in genomic sequences constructed from genomic variants and then highlight the patterns that we discover through our analyses of the discovered motifs. Next, we discuss development of a convolutional neural network which incorporates genomic variant data to perform the TFBS classification task on sequence motifs. We also explain how visualization techniques can be applied to the trained network to better understand the biological effects of genomic variants on TFBSs.

Andrew Gritsevskiy

Structural gene expression

Mentor: Dr. Gil Alterovitz

Project suggested by Prof. Gil Alterovitz

The advent of next-generation sequencing technologies provides access to genomic and epigenetic data, such as sequence motifs, histone modifications, as well as structural motifs and binding factors. It has been shown that most of these profiles contribute significantly to gene enrichment and expression. In our study, we propose using deep convolutional neural networks for the purpose of analyzing the functional importance of genomic structures, such as protein enrichment based on Hi-C data. We discuss several methods of representing and convolving over the genome's 3-D shape, as well as using techniques such as one-shot learning in order to determine the spatial orientation of bound factors. Further, we suggest a cross-species comparison of structural motifs in an attempt to pinpoint the evolutionary significance of the enrichment of specific genes.

Eric You

Analyzing compression: Using gene sets to find patterns in genomic compression

Mentor: Prof. Gil Alterovitz

Project suggested by Prof. Gil Alterovitz

Obtaining genomic data has become easier and much cheaper with the rise of Next Generation Sequencing (NGS) technologies. However, due to the sheer size of genomic data, genomic compression through machine learning has been investigated as an alternative to conventional methods such as gzip. Another problem that comes with genomic data is the difficulty quantifying similarities and differences. Methods developed in the past, such as Gene Set Enrichment Analysis (GSEA) and variations such as Spectral Gene Set Enrichment (SGSE) have allowed for a robust analysis of gene expression data. Here, I attempt to combine these two fields and explore the use of machine learning to compress genomic data, while trying to extrapolate biological significance from the compressibility

of the data. I propose one method of comparing raw genomic data, by calculating an enrichment score, based on the gene sets within genomic data. Finally, I will outline some potential improvements to genomic compression through machine learning, by taking advantage of the observed biological connections between the data.

SESSION 14

Makiah Bennett and Jack Flahive

Precision cancer medicine web application

Mentor: Dr. Gil Alterovitz

Project suggested by Dr. Gil Alterovitz

The global focus on precision medicine has greatly increased over the past decade after being highlighted in 2015 by the Obama administration. Many applications have been developed in prior years to drive precision medicine research forward, but the acceleration of technology has rendered many of these applications obsolete. One such application created by Vanderbilt University sought to bypass obsoletism by implementing a number of state-of-the-art features including unique visualization techniques, yet was developed explicitly for the iOS platform, thus limiting its audience to a small subset of who would benefit from it. This project seeks to update the functionality of this application while maintaining a similar purpose: the comparison of cancer patients' diagnosis-specific somatic gene mutations to a population-level set of comparable data, both manually and through machine learning. Through these novel comparisons, this EHR-agnostic, clinico-genomic application can more properly diagnose and prescribe drug therapies.

James Jusuf

A versatile algorithm for finding patterns in large cancer cell line data sets

Mentor: Dr. Mahmoud Ghandi, Broad Institute

Project suggested by Dr. Mahmoud Ghandi

Since the advent of The Cancer Genome Atlas (TCGA) in 2005 and the Cancer Cell Line Encyclopedia (CCLE) in 2012, the amount of data regarding cancer cell lines and their genetic and epigenetic characterizations has grown rapidly, while becoming much more accessible. The abundance of cancer cell line data allows for the reliable detection of potentially meaningful correlation patterns between characteristics such as mutations in specific genes, methylation profiles, drug resistance, etc. In this study, we present an algorithm to calculate p-values to quantify the correlations between methylation and mutation of individual genes among cell lines of a given cancer type (e.g. pancreatic, small-cell lung, colorectal, etc.). We demonstrate how the algorithm can be generalized to compare other characteristics of cancer cell lines in which one variable is continuous and

the other is binary. In the future, we hope to expand the algorithm to compare two continuous variables. The discovery of patterns in large cancer cell line datasets using these algorithms will reveal potential candidates for further experimental analysis and aid in the development of a predictive model for interpreting genetic tests in cancer patients.

Kevin Hu

An analysis of MDM4 alternative splicing and effects across cancer cell lines

Mentor: Dr. Mahmoud Ghandi, Broad Institute

Project suggested by Dr. Mahmoud Ghandi

MDM4 is an important regulatory gene that encodes a protein that binds to and disables the p53 tumor suppressor. Previous studies have shown that an absence of MDM4 expression results in rapid animal death due to a p53-induced lack of cell division. Alternative splicing is responsible for two well-known isoforms of MDM4, the normally-expressed MDM4-FL and a shorter variant, MDM4-S, that lacks the commonly-included 6th exon and is unable to properly prevent p53 regulation. MDM4-S is believed to indirectly inhibit normal MDM4-mediated suppression of p53 by replacing the normally-produced MDM4-FL in the cell. In this study, we apply several techniques for identification and quantification of these isoforms in RNA-seq data from 1,019 distinct human cancer cell lines of the Cancer Cell Line Encyclopedia (CCLE). Using Trinity and MAFFT to perform de novo sequence assembly and multiple sequence alignment, respectively, we confirm the expression of these two isoforms as well as others hypothesized in existing literature. We then estimate the expression levels of these isoforms using a nonnegative matrix factorization (NMF) method over counts of splicing junctions. In the future, we plan on exploring the impact of these isoforms on p53 knockdown sensitivity as well as also examining the role of the structurally similar p53-suppressing MDM2.

Kalyan Palepu and Andrew Zhang

Compression of genomic variants using convolutional autoencoders

Mentor: Dr. Gil Alterovitz

Project suggested by Dr. Gil Alterovitz

The next generation High Throughput Sequencing (HTS) machines have reduced the cost of genome sequencing dramatically from over a million to less than $2,000. The falling price of genome sequencing makes it more affordable and readily available for disease diagnosis, prevention and precision medicine. However, the size of an individual genome data is hundreds of gigabytes, which makes storing and transmitting of the data costly and challenging. Thus, the need for effective compression of genomic data

has become quite urgent. We present a system, based on a deep convolutional autoencoder, that can effectively compress variants based genome sequence data. Testing with James Watson's Genome data shows very promising results, when comparing with other methods.