MIT PRIMES
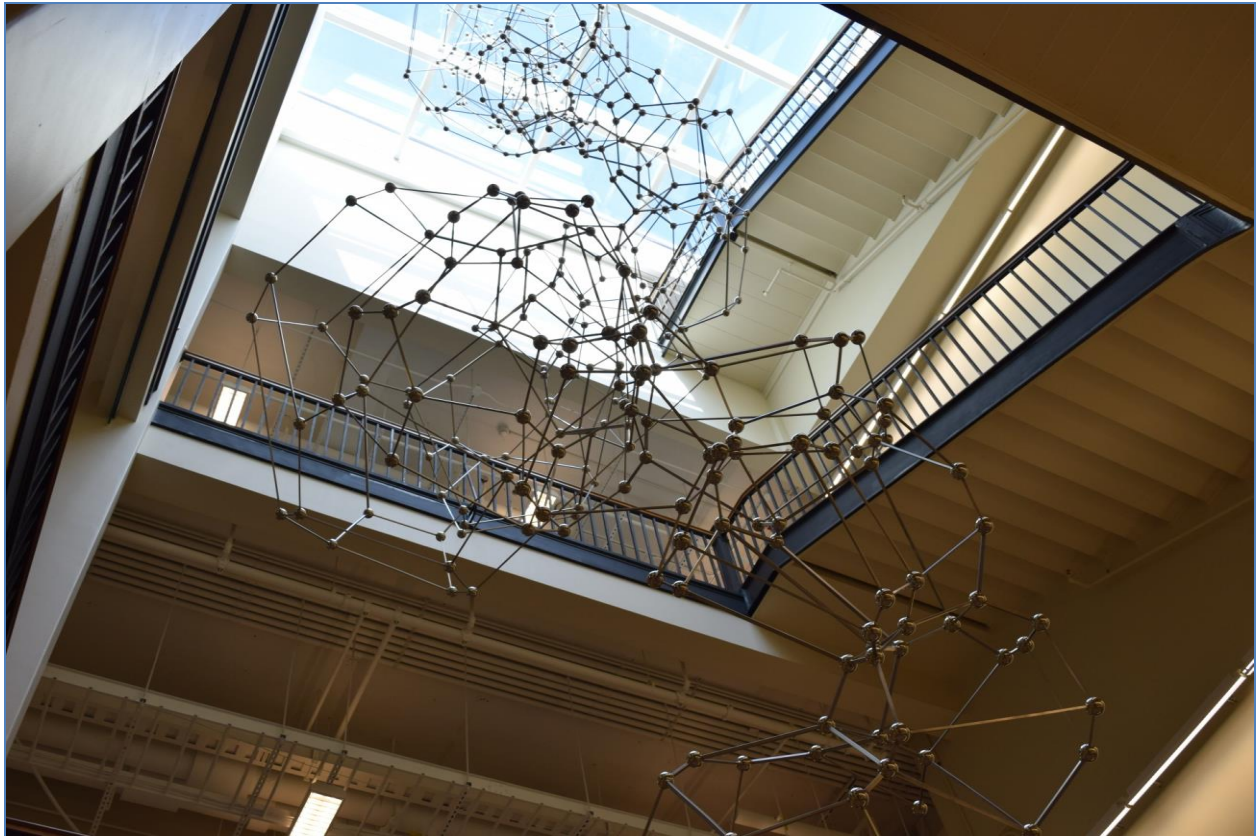

Installation *Chord* by Antony Gormley (2015) at the entrance to MIT's Math Department. Photo by Slava Gerovitch.

SIXTH ANNUAL CONFERENCE

MAY 2016

Sixth annual PRIMES conference

Room 4-270, MIT

Saturday, May 21
Section I. Mathematics

8:20 Welcoming Remarks
Prof. Tom Mrowka, Chair of the MIT Mathematics Department
Prof. Pavel Etingof, PRIMES Chief Research Advisor
Dr. Slava Gerovitch, PRIMES Program Director

8:50 Session 1
Nicholas Guo, "Rational hyperplane arrangements and counting independent sets of symmetric graphs" (mentor Guangyi Yue)
Valerie Zhang, "Computer-based representations and manipulations of paths in the plane" (mentor Umut Varolgunes)
Aaron Yeiser, "A robust spectral PDE solver on skinny triangles" (mentor Dr. Alex Townsend)

9:50-10:00 break

10:00-11:00 Session 2
Kai-Siang Ang, "On the geometry of icosahedral viruses" (mentor Prof. Laura Schaposnik, University of Illinois at Chicago)
Nikhil Marda, "On point separation by arrangements of lines" (mentor Borys Kadets)
Zachary Chroman, "Rational embeddings of convex polyhedra" (mentor Sheela Devadas, Stanford University)

11:00-11:10 break

11:10-12:10 Session 3
Ria Das, "Investigations of mixed reinforcement-memory models for random walks" (mentor Andrew Rzeznik)
PRIMES STEP students, "Who is guilty?" (mentor Dr. Tanya Khovanova)
PRIMES STEP students, "Alternator coins" (mentor Dr. Tanya Khovanova)

12:10-1:10 lunch

1:10-2:20 Session 4
Felix Wang, "Ramification of solutions of functional equations" (mentor Prof. Michael Zieve, University of Michigan)
Nathan Smith, "Square-primitive gaps" (mentor Xiaoyu He, Harvard University)
Meena Jagadeesan and Karthik Karnik, "The Outer Automorphism of S_6" (mentor Akhil Mathew)

2:20-2:25 break

2:25- 3:25 Session 5
Nina Anikeeva, "Applications of ergodic theory to continued fractions on the Heisenberg group" (mentor Prof. Jayadev Athreya, University of Washington)
Rafael Saavedra, "Discreet coin weighings and the Frobenius problem" (mentor Dr. Tanya Khovanova)
Harish Vemuri, "Tiling-harmonic conjugate functions" (mentor Prof. Sergiy Merenkov, CCNY – CUNY)

3:25-3:30 break

3:30-4:35 Session 6
Nelson (Shuheng) Niu, "Extensions of classic combinatorial games" (mentor Dr. Tanya Khovanova)
Kevin Chang, "Ordered Ramsey numbers of small graphs" (mentor William Kuszmaul, Stanford University)
Louis Golowich and Richard Zhou, "Maximum number of pairwise G-different permutations" (mentor Chiheon Kim)

4:35-4:40 break

4:40-5:45 Session 7
Eric Nie and Alok Puranik, "Invariants of knots" (mentor Zhenkun Li)
Albert Yue, "Knot diagram invariants and bounds for the number of Reidemeister moves needed for unknotting" (mentor Piotr Suwara)
Alec Leng, "Independence of the Miller-Rabin and Lucas probable prime tests" (mentor David Corwin)

5:45-6:15 pizza break

6:15-7:15 Session 8
Maya Sankar, "The dimensions of partially directed nil-Temperley-Lieb algebras" (mentor Dr. Tanya Khovanova)
Laura Pierson, "Signatures of stable multiplicity spaces in symmetric group restrictions" (mentor Siddharth Venkatesh)
Dhruv Rohatgi, "A connection between vector bundles over smooth projective curves and representations of quivers" (mentor Vishal Arul)

7:15-7:20 break

7:20-8:20 Session 9
Alec Sun, "Wall crossing bijections and representations of rational Cherednik algebras" (mentor Seth Shelley-Abrahamson)
Matt Lipman, "Representations of Cherednik algebras" (mentor Gus Lonergan)
Matthew Hase-Liu, "Counting points on curves of the form $y^{m_1}=c_1 x^{n_1}+c_2 x^{n_2} y^{m_2}$" (mentor Nicholas Triantafillou)

Sunday, May 22

Section II. Computer Science and Computational Biology

8:45 am: Welcoming Remarks
Prof. Srini Devadas, Department of Electrical Engineering and Computer Science
Dr. Slava Gerovitch, PRIMES Program Director

9:00-10:20 Session 10: Computer Science
Harshal Sheth, Aashish Welling, and Nihar Sheth, "Read-copy update in a garbage collected environment" (mentor Cody Cutler)
Vivek Bhupatiraju, John Kuszmaul, and Vinjai Vale, "Exploring proof of space with hard-to-pebble graphs" (mentors Ling Ren and Albert Kwon)
Leo Alcock "Private publishing using Bitcoin" (mentor Ling Ren)

10:20-10:30 break

10:30-11:45 Session 11: Computer Science
Cristian Gutu, "SeifPass: A secure password manager" (mentor Albert Kwon)
Henry Liu, Justin Kaashoek and Siye Zhu, "Scalable logging algorithm for in-memory database systems" (mentor Xiangyao Yu)
Yatharth Agarwal and Vishnu Murale, "Moving in next door: network flooding as a side channel in cloud environments" (mentors Dr. Jason Hennessey, Kyle Hogan, and Dr. Mayank Varia)

11:45-11:55 break

11:55-1:00 Session 12: Computational and Physical Biology
Prof. Leonid Mirny, Introductory remarks
Laura Braverman, "Protein determinants of chromosome domains" (mentor Nezar Abdennur)
Betsy Pu, "Chromatin states at boundary elements" (mentor Nezar Abdennur)
Krishna Suraj, "Emergent chromosome organization in interphase from loop extrusion" (mentor Dr. Geoffrey Fudenberg)

1:00-2:00 lunch break

2:00-2:50 Session 13: Computational Neuroscience
Prof. Ed Boyden, Introductory remarks
Albert Gerovitch, "Metrics for comparing 3D neuron segmentations in expansion microscopy connectomics" (mentor Dr. Adam Marblestone)
Zachary Steinberg, "Automatic segmentation of punctate 3D super-resolution microscopy data" (mentor Daniel Goodwin)

2:50-3:00 break

3:00-4:10 Session 14: Medical Informatics
Prof. Gil Alterovitz, Introductory remarks

Daniel Lu, "Study of various synergistic drug mechanisms in disordered protein-related diseases" (mentor Prof. Gil Alterovitz)

Kara Luo, "Computational modeling identifies biosynthetic modifications to improve drug inhibition against *Klebsiella pneumoniae*" (mentor Prof. Gil Alterovitz)

Arul Prasad, "Targeting viral envelope proteins: an application to the Zika virus" (mentor Prof. Gil Alterovitz)

4:10-4:15 break

4:15-5:20 Session 15: Medical Informatics
Andrew Gritsevskiy and Adithya Vellal, "Compression and integration of human genomic variants into smart EHR systems" (mentor Prof. Gil Alterovitz)

John Flahive, "Providing clinical decision support to medical providers through interpretation of gene-drug interactions" (mentor Prof. Gil Alterovitz)

James Jusuf, "Exploring the effects of CTCF binding site mutations on transcriptional regulation" (mentor Dr. Mahmoud Ghandi, Broad Institute)

SESSION 1

Nicholas Guo

Rational Hyperlane Arrangements and Counting Independent Sets of Symmetric Graphs

Mentor: Guangyi Yue

Project suggested by Prof. Richard Stanley

A hyperplane arrangement is a finite set of hyperplanes, equations of the form $a_1 x_1 + a_2 x_2 + \cdots + a_n x_n = b$, in a real vector space $\mathbb{R}^n$. Each arrangement has a characteristic polynomial defined by the Möbius function that holds important properties such as the number of separated regions. We investigate several hyperplane arrangements that can be viewed as deformations of Braid arrangements, which are the sets of all hyperplanes: $x_i = 0, \forall\ i;\ x_i = x_j, \forall\ i < j;\ x_i = a_1 x_j, x_i = a_2 x_j, \dots x_i = a_m x_j, \forall\ i \neq j,\ i, j \in [n],\ a_i \in \mathbb{N}$, in $\mathbb{R}^n$, and their corresponding graphs with vertex set $\mathbb{Z}/q\mathbb{Z}$ and edges ij if $i = a_r j$ for some r. In particular, we prove a conjecture of Stanley that for the disjoint union $G = \mathbb{Z}/n_1\mathbb{Z} \cup \mathbb{Z}/n_2\mathbb{Z} \cup \cdots \cup \mathbb{Z}/n_s\mathbb{Z}\ (n_1, n_2, \dots, n_s \gg 1)$, the number of n-element independent sets of G depends only on n, m, and $\sum_i n_i$ if and only if the a_i's are all coprime.

Valerie Zhang

Computer-based representations and manipulations of paths in the plane

Mentor: Umut Varolgunes

Project suggested by Prof. Paul Seidel

Matching paths come up naturally in the study of complex algebraic varieties along with an important operation called the Hurwitz move.

Given n points in the 2-D plane, a matching path is a path that starts at one of these n points and ends at a different one without going through any of the other $n - 2$ points. At the heart of the Hurwitz move is the twist operation, which twists one matching path along another to produce a new (third) matching path. Because the twist is hard to do by hand, we develop computer-based methods to represent matching paths and perform the twist.

Aaron Yeiser

A robust spectral PDE solver on skinny triangles

Mentor: Dr. Alex Townsend

Project suggested by Dr. Alex Townsend

When solving differential equations in two dimensions, it is necessary to consider the shape of the domain. Most domains are not easy to represent exactly, like rectangles or circles. In order to compensate, the domain is subdivided into thousands or even millions of polygons, which are typically triangles. Unfortunately, meshes have one major flaw — they can have skinny triangles. Skinny triangles cause conventional differential equation solvers in two dimensions, like finite element methods, to become numerically unstable. In fact, most of the computation time used by finite element methods is spent generating a mesh with no skinny triangles. When meshes change throughout the simulation, such as in simulations of fluid flow, the simulation has to be periodically remeshed to eliminate skinny triangles. However, the remeshing process is computationally expensive and it also sacrifices some accuracy.

We are developing a method that will be able to accurately solve differential equations on meshes with skinny triangles. Currently, our method is numerically stable on skinny triangles, and untested on multiple triangles.

SESSION 2

Kai-Siang Ang

On the geometry of icosahedral viruses

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

Project suggested by Prof. Laura Schaposnik

Mathematicians often desire methods of classification; we study the methods of classifying biological viruses. We first summarize known classifications of viruses by mathematical and biological properties and give an overview of mathematical ways of describing icosahedral viruses. Then, we use the basic ideas of symmetry of overlapping arguments, rotations in the h, k coordinate system, and the concept of bordering to expand on previous work, deriving a complete classification (with proof) of all possible ways in which regular penta-, tri-, and disymmetrons may be used to construct icosahedral virus capsids. For every class of solutions, we further provide formulas for symmetron sizes and parity restrictions on h, k, and T numbers.

Nikhil Marda

On point separation by arrangements of lines

Mentor: Borys Kadets

Project suggested by Prof. Larry Guth

We investigate the problem of separating a set of points in general position equally among the cells formed by some arrangement of lines. With a set X of N points and some arrangement of D lines in the plane, each cell should contain $O(D^{-2}N)$ points. We show that a large subset of X lying on a convex curve is one, but not the only, obstacle preventing the equal separation of points. The stabbing number of a geometric object is

defined to be the maximum number of intersections any line in the plane can share with the object. We consider $d(X)$, the minimum stabbing number of a curve containing X. Given a large subset Y of X, we show that a small $d(Y)$ prevents equal separation. We further consider properties of $d(X)$ and related curves categorized by stabbing number. We show that $d(X) = O(\sqrt{N})$ and generalize bounds on the Erdos-Szekeres problem on points in convex position by showing the existence of a curve with stabbing number $\leq d$ containing n points in a set of $O(d \cdot 4^{n/d+o(n/d)})$ points.

Zachary Chroman

Rational embeddings of convex polyhedra

Mentor: Sheela Devadas, Stanford University

Project suggested by Prof. Ravi Vakil, Stanford University

In his seminal theorem, Steinitz completely characterized the graphs corresponding to convex polyhedra as identically the 3-connected planar graphs. Following this proof it can be shown that every convex polyhedron has an embedding with all vertices having rational coordinates. Additionally, Hart showed that every convex polyhedron has a unique "canonical" form with all edges adjacent to the unit sphere, and with the origin as the center of mass of the tangency points. However, little more is known about the embeddings of polyhedra. We detail the existing results in this area and outline an approach for studying embeddings of polyhedra with entirely triangular faces.

SESSION 3

Ria Das

Investigations of mixed reinforcement-memory models for random walks

Mentor: Andrew Rzeznik

Project suggested by Dr. Pierre-Thomas Brun

Under certain circumstances, a swarm of a species of trail-laying ants known as army ants can become caught in a doomed revolving motion known as the death spiral, in which each ant follows the one in front of it in a never-ending loop until they all drop dead from exhaustion. This phenomenon, as well as the ordinary motions of many ant species and certain slime molds, can be modeled using reinforced random walks and random walks with memory. In a reinforced random walk, the path taken by a moving particle is influenced by the previous paths taken by other particles. In a random walk with memory, a particle is more likely to continue along its line of motion than change its direction. Both memory and reinforcement have been studied independently in random walks with interesting results. However, real biological motion is a result of a combination of both memory and reinforcement. In this paper, we develop a random walk model that mixes both memory and reinforcement, and write code that simulates these walks on both simple lattice structures and fully generalized graphs. We discover the novel result that there exists an optimum mix of memory and reinforcement in a random walk that produces a minimum travel time between two locations. We also lay the foundation for future work in specifying mixed random walk models that precisely mimic experimental data as well as reproducing and explaining such enigmatic observed phenomena as the ant death spiral.

Benjamin Chen, Ezra Erives, Leon Fan, Michael Gerovitch, Jonathan Hsu, Neil Malur, Ashwin Padaki, Nastia Polina, Will Sun, Jacob Tan, and Andrew The (PRIMES STEP students)

Who is guilty?

Mentor: Dr. Tanya Khovanova

Project suggested by Dr. Tanya Khovanova

We discuss a generalization of logic puzzles in which truth-tellers and liars are allowed to deviate from their pattern in case of one particular question: "Are you guilty?"

Benjamin Chen, Ezra Erives, Leon Fan, Michael Gerovitch, Jonathan Hsu, Neil Malur, Ashwin Padaki, Nastia Polina, Will Sun, Jacob Tan, Andrew The (PRIMES STEP students)

Alternator coins

Mentor: Dr. Tanya Khovanova

Project suggested by Dr. Tanya Khovanova

We introduce a new type of coin: *the alternator*. The alternator can pretend to be either a real or a fake coin (which is lighter than a real one). Each time it is put on a balance scale it switches between pretending to be either a real coin or a fake one.

We solve the following problem: You are given N coins that look identical, but one of them is the alternator. All real coins weigh the same. You have a balance scale which you can use to find the alternator. What is the smallest number of weighings that guarantees that you will find the alternator?

SESSION 4

Felix Wang

Ramification of solutions of functional equations

Mentor: Prof. Michael Zieve, University of Michigan

Project suggested by Prof. Michael Zieve

We study the following question:

(1) What are all solutions to the functional equation $f \circ \hat{f} = g \circ \hat{g}$ in complex rational functions $f, \hat{f}, g, \hat{g} \in \mathbb{C}(X)$?

We utilize the concept of ramification, along with various algebraic, geometric, and analytic results to resolve this question in the case that the numerator of $f(X) - g(Y)$ is an irreducible polynomial in $\mathbb{C}[X, Y]$ of sufficiently large degree. Our work comprises a vast generalization of previous works by Ritt, Avanzi-Zannier, Bilu-Tichy, and others. Our result answers a 1973 question of Fried in all but finitely many cases.

Nathan Smith

Square-primitive gaps

Mentor: Xiaoyu He, Harvard University

Project suggested by Xiaoyu He

In this paper, we investigate primarily the sizes of the gaps of square-primitive sets, which are defined as sets of positive integers such that the quotient of two distinct terms is never the square of an integer. We will show that every infinite square-primitive set has infinitely many gaps of size at least 2. In addition, we will use a probabilistic construction to show the existence of square-primitive sets where the gaps between terms of size n is $O(\log^2 n \log \log n^\epsilon)$ for any given $\epsilon > 0$. We will also outline several other techniques we considered which have yet to yield significant results.

Meena Jagadeesan and Karthik Karnik

The Outer Automorphism of S_6

Reading group mentor: Akhil Mathew, Harvard University

We provide an exposition of automorphisms of the symmetric group, S_n, including the notion of a *complete group*, a centerless group with no outer automorphisms. We then show that S_n is complete for $n \neq 2, 6$. For S_6, we prove that there exists an exceptional outer automorphism of S_6, a result attributed to Hölder. We present two methods of constructing a 120-element transitive subgroup of S_6, using the action of $PGL_2(\mathbb{F}_5)$ on $\boldsymbol{P^1}(\mathbb{F}_5)$ and the action of S_5 on its 5-Sylow subgroups. Using this transitive subgroup, we determine the existence of an outer automorphism of S_6, whose preimage of S_5 is transitive.

SESSION 5

Nina Anikeeva

Applications of ergodic theory to continued fractions on the Heisenberg group

Mentor: Prof. Jayadev Athreya, University of Washington

Project suggested by Prof. Jayadev Athreya

In this talk we give an overview of classical continued fractions and the motivation behind their study, discuss the Heisenberg group $\mathbb{H}$, and how how to construct continued fractions on it using the planar Siegel model of the group. Then we examine and work with major theorems in ergodic theory to explore results concerning growth of denominators and digit frequency.

Rafael Saavedra

Discreet coin weighings and the Frobenius problem

Mentor: Dr. Tanya Khovanova

Project suggested by Dr. Tanya Khovanova

In 2007, Alexander Shapovalov posed an old twist on the classical coin weighing problem by asking for strategies that manage to conceal the identities of specific coins while providing general information on the number of fake coins. In 2015, Diaco and Khovanova studied various cases of these "discreet strategies".

In this talk we discuss a natural coin weighing strategy which we call the sorting strategy: divide the coins into equal piles and sort the remaining piles by weight. We study the instances when the strategy is discreet, and given an outcome of the sorting strategy, the possible number of fake coins. We describe how the values that occur are explained by the relationship to the Frobenius problem.

Harish Vemuri

Tiling-harmonic conjugate functions

Mentor: Prof. Sergiy Merenkov, CCNY CUNY

Project suggested by Prof. Sergiy Merenkov

Tiling-harmonic functions are a class of functions defined on square tilings that minimize a particular energy function for a given set of boundary values. Certain properties of these functions will be helpful in simplifying many results in the study of Sierpinski carpets. Tiling-harmonic functions are closely related to a much more well-known class of functions called graph-harmonic functions, and since there is very useful notion of a conjugate function in the graph-harmonic case, we wish to find an analogue for this in the tiling-harmonic case. We begin by defining tiling-harmonic and graph-harmonic functions and proceed to describe the conjugation procedure for graph-harmonic functions. Then we discuss several directions to take for finding a suitable tiling-harmonic conjugate, including modifying the method for graph-harmonic functions, finding a pseudo-conjugate, and generalizing the energy function, and present our conjectures in each of the directions. Finding a suitable conjugate function for tiling-harmonic functions will allow us to relate the properties of tiling-harmonic functions to those of more well-known functions and may help prove the main conjectures in this field.

Session 6

Nelson (Shuheng) Niu

Extensions of classic combinatorial games

Mentor: Dr. Tanya Khovanova

Project suggested by Dr. Tanya Khovanova

We discuss a variant of Wythoff's Game, m-modular Wythoff's Game, and identify the winning and losing positions of the game in two piles. We also discuss generalizations of Wythoff's Game and, in doing so, describe the properties of the winning and losing positions of a related game, the Matchbox Game. We show that a 3-pile generalization of m-modular Wythoff's game has a finite number of P-positions.

Kevin Chang

Ordered Ramsey numbers of small graphs

Mentor: William Kuszmaul, Stanford University

Project suggested by Prof. Jacob Fox, Stanford University

Let L be a set of labellings for a graph H, where each labeling has vertex set $\{1, 2, \ldots, k\}$. The ordered Ramsey number $R_<(L)$ is the smallest n such that every two-coloring of the

edges of the complete graph on $\{1, 2, \ldots, n\}$ contains a copy of H with vertices appearing in the same order as in one of the labellings in L. Our work concerns the ordered Ramsey numbers of labellings of small graphs. In particular, we study the ordered Ramsey number for labellings of the diamond graph. When L contains only the standard ordering, we prove that $R_<(L) \geq 12$. When L contains all orderings whose smallest vertex is on the diamond's diagonal, we prove that $R_<(L) \leq 13$. When L contains all orderings whose smallest vertex is on the diamond's tip, we prove that $R_<(L) \leq 14$.

Louis Golowich and Richard Zhou

Maximum number of pairwise G-different permutations

Mentor: Chiheon Kim

Project suggested by Chiheon Kim

Two permutations of the vertices of a graph G are called G-different if there exists an index i such that i-th entry of two permutations forms an edge in G. Let $F(G)$ be the maximum number of pairwise G-different permutations. We show that for all balanced bipartite graphs G whose complement has maximum degree $o(n)$, $F(G) = 2^{(1-o(1))n}$ where n is the number of vertices of G. We also present examples of bipartite graphs G with maximum degree $O(\log_2 n)$ for which this equation holds.

SESSION 7

Eric Nie and Alok Puranik

Invariants of knots

Reading group mentor: Zhenkun Li

It is a key problem in knot theory how to distinguish different types knots from each other and knot invariant is an important tool. In the talk we introduce different kinds of knot invariants including some combinatorial invariants coming from the knot diagram such as crossing number, some topological invariants such as knot genus and also some polynomial invariants like Alexander or Jones polynomial.

Albert Yue

Knot diagram invariants and bounds for the number of Reidemeister moves needed for unknotting

Mentor: Piotr Suwara

Project suggested by Piotr Suwara

Knot diagram invariants are quantities assigned to knot projections and are constant under plane deformations of diagrams. They may change in very controlled ways under Reidemeister moves, and thus can be helpful for establishing polynomial lower or upper bounds with respect to the number of crossings for the number of moves for unknotting. We introduce a new knot diagram invariant called self-crossing index (SCI) and consider the effects of framed Reidemeister moves on the invariant. We prove that SCI only changes under framed Reidemeister moves of type III, by ± 1, meaning SCI provides lower bounds for unknotting. We also investigate a family of knots that has a quadratic lower bound for unknotting established by SCI.

Alec Leng

Independence of the Miller-Rabin and Lucas probable prime tests

Mentor: David Corwin

Project suggested by Dr. Stefan Wehmeier, Mathworks

Fast primality testing is necessary in modern public-key cryptography. In particular, probabilistic tests are used for their speed, despite the potential for pseudoprimes. We look at numbers with many nonwitnesses for the Miller-Rabin and Lucas tests and demonstrate that they are usually either Carmichael or Lucas-Carmichael numbers. We then use these categorizations to prove that there are no numbers with many nonwitnesses for both tests.

SESSION 8

Maya Sankar

The dimensions of partially directed nil-Temperley-Lieb algebras

Mentor: Dr. Tanya Khovanova

Project suggested by Prof. Alexander Postnikov

We extend the definition of nil-Temperley-Lieb algebras to graphs with some directed and some undirected edges by removing some relations. We fully describe the cases where the partially directed nil-Temperley-Lieb algebra is finite. We further completely describe the structure of the basis monomials of these algebras and calculate the dimensions of these algebras in special cases.

Laura Pierson

Signatures of stable multiplicity spaces in symmetric group restrictions

Mentor: Siddharth Venkatesh

Project suggested by Prof. Pavel Etingof

Important properties of the symmetric group S_n, the group of permutations of n objects, can be studied by examining its representations. Irreducible representations of S_n correspond to partitions of n, or equivalently, to combinatorial objects called Young diagrams, and the multiplicity space, space of restrictions, from a fixed irreducible representation of S_n to a fixed irreducible representation of S_{n-k} corresponds to removing k boxes from the corresponding Young diagram. A stable sequence for a multiplicity space can be formed by adding boxes to the first row of both Young diagrams. Many important properties, such as the basis vectors, are fixed under this stable sequence, but the norms of the basis vectors under a certain inner product invariant under S_n are rational functions of n. We will be computing these norms, and also considering the signatures, or number of basis vectors with positive norm minus number with negative norm.

Dhruv Rohatgi

A connection between vector bundles over smooth projective curves and representations of quivers

Mentor: Vishal Arul

Project suggested by Dr. Tristan Bozec

Schiffmann conjectured a relation between the problems of counting over a finite field (1) vector bundles over smooth projective curves, and (2) representations of quivers. This conjecture in part reduces to proving that each function in a certain set, indexed by the set of integer partitions, is polynomial. We show this for the subset of integer partitions whose elements are equal, by means of a bijection of partition tuples.

SESSION 9

Alec Sun

FLOTW multipartitions and rational Cherednik algebras

Mentor: Seth Shelley-Abrahamson

Project suggested by Seth Shelley-Abrahamson

Rational Cherednik algebras were introduced by Etingof and Ginzburg in 2002. Recently, Losev provided a concrete combinatorial method for determining the *support* of a simple module $L_c(\lambda)$ of a cyclotomic rational Cherednik algebra $H_c(W, \mathfrak{h})$ in terms of its multipartition label λ and the Cherednik algebra parameter c. The fundamental combinatorial tools introduced are the *wall crossing bijections*, which are certain size-preserving combinatorial bijections on the set of multipartitions that allow one to reduce the general problem of determining supports to the problem of determining supports in the case that the parameter c lies in an *asymptotic chamber*. Even more recently, Jacon and Lecouvey has given a simple combinatorial rule for the computation of these wall crossing bijections by relating them to combinatorial R-matrices. While the significant progress mentioned above can be applied, in principle, to compute the support of $L_c(\lambda)$ for any particular multipartition λ and multipartition c, it does not immediately give closed form answers to determining, for a given multipartition λ, the locus of parameters c for which $L_c(\lambda)$ is finite-dimensional, or determining, for a given parameter c and support variety X, the multipartitions λ for which $L_c(\lambda)$ has support X. In this project we hope to answer questions of this flavor, using the combinatorial approach permitted by the work of Losev and others.

Matt Lipman

The Polynomial Representation of the Rational Cherednik Algebra of type A_{N-1} in characteristic p for $p | N - 1$

Mentor: Gus Lonergan

Project suggested by Prof. Pavel Etingof

We investigate the type A_{N-1} Cherednik algebra in characteristic $p | N - 1$, specifically the polynomial representation. Apart from minor lemmas, we currently do not have any new theorems, but we have two conjectures for the singular vectors, one based on theory from the $p | N$ case and one based on computational data; if both hold, the result obtained by setting their spans equal is strong. Finally, we set forth a conjecture for the result of repeatedly applying of the Dunkl operator to monomials of a single vector.

Matthew Hase-Liu

Counting points on curves of the form $y^{m_1} = c_1 x^{n_1} + c_2 x^{n_2} y^{m_2}$

Mentor: Nicholas Triantafillou

Project suggested by Dr. Andrew Sutherland

Let $p \neq 2, 3$ be a prime such that $p \equiv 1 \pmod{e}$ and $e = 3$, 4, 6, 8, 9, 10, 11, 12, 15, 20, or 24. We describe an algorithm to quickly determine the number of rational points on curves of the form $y^{m_1} = c_1 x^{n_1} + c_2 x^{n_2} y^{m_2}$ over the finite field $\mathbb{F}_p$, where $c_1, c_2 \in \mathbb{F}_p^{\times}$. Our proposed algorithm computes the number of points on a curve modulo p via the Hasse-Witt matrix and determines the exact number using the Hasse-Weil bound. It has an expected probabilistic runtime of $O\left(M\left(\log p\right) \log p\right)$, a deterministic runtime of $O\left(M\left(\log p\right) \log^2 p \log \log p\right)$ assuming the generalized Riemann Hypothesis, and a deterministic runtime of $O\left(M\left(\log^3 p\right) \log^2 p / \log \log p\right)$, where $M\left(n\right) = O\left(n \log n \log \log n\right)$ is the time needed to multiply two n-digit numbers.

SESSION 10: COMPUTER SCIENCE

Harshal Sheth, Aashish Welling, and Nihar Sheth

Read-copy update in a garbage collected environment

Mentor: Cody Cutler

Project suggested by Prof. Frans Kaashoek

Read-copy update (RCU) is a well-known synchronization mechanism used in parallel systems with high reading thread to writing thread ratios. RCU has mainly been used in lower level languages such as C or C++. This paper explores the viability of RCU in a garbage collected language, specifically, Go. The results of various performance tests indicate that RCU in Go provides similar performance improvements as C++ RCU does to read-write mutexes. Furthermore, the ease of use of RCU in Go is far superior to that of RCU in C++.

Vivek Bhupatiraju, John Kuszmaul, and Vinjai Vale

Exploring proof of space with hard-to-pebble graphs

Mentors: Ling Ren and Albert Kwon

Project suggested by Prof. Srini Devadas

Proof of Work (PoW) is a protocol for an interaction between two entities, the prover and the verifier. The prover wants to show that she has done a nontrivial amount of computational work in order to be allowed to complete some transaction. This system has a variety of uses: for example, requiring a PoW attached to each email being sent is a way to prevent email spam, and PoW is also used as the basis for several consensus protocols, most notably Bitcoin. However, PoW comes with its drawbacks. Solving a proof of work takes a lot of compute power on the provers part, and that energy is effectively wasted. Also, well-funded provers can afford expensive computing chips that are specially designed for backsolving hashes, giving them a significant unfair advantage over their peers. In the last few years, Proof of Space (PoS) has emerged as an alternative. At a high level, it operates in a similar fashion to the Proof of Work. A prover is trying to demonstrate to a verifier that he has devoted a significant amount of storage towards the proof of space. In this presentation, we discuss how one can construct a successful PoS scheme. We briefly describe our recent work, which has been focused on implementing and evaluating the performance of various structures and algorithms, and our plans for future work.

Leo Alcock

Private publishing using Bitcoin

Mentor: Ling Ren

Project suggested by Prof. Srini Devadas

We implemented a global commitment scheme by piggybacking Bitcoins consensus protocol. The commitment scheme allows one to assure to the world that one had data x at time t without revealing any features at time t. This protocol is very useful in intellectual

property disputes. Take a case where Alice has made an invention or discovery and wants to secure credit for her work before publishing to develop her idea. To do so Alice would simply apply the protocol.

SESSION 11: COMPUTER SCIENCE

Cristian Gutu

SeifPass: A secure password manager

Mentor: Albert Kwon

Project suggested by Prof. Srini Devadas

Passwords are common. We use them everyday to authenticate to devices and online services. Despite their importance, the way we store passwords is fundamentally broken. Often times, password databases get compromised and when they get compromised they are vulnerable to "offline attacks." Ideally, we want to encrypt passwords in such a way that if the password database gets compromised the passwords remain secure. This is possible by implementing a password encryption system in which trust is distributed amongst multiple machines. This is nice because if one machine gets compromised but not the other the attacker has insufficient information to conduct an offline attack. Pythia is a password encryption protocol which inherits this idea of distributing trust. Pythia also exposes enough information for us to be able to detect online attacks, rotate the encryption key and, as a result, conduct remote cryptographic erasure of compromised databases. While powerful, Pythia still has its limitations. Namely, it may be hard for large companies like Google, Facebook, and Microsoft to implement it in an already sophisticated code base. Ultimately, not all companies are guaranteed to implement Pythia in their systems. This sets the user at a disadvantage. My project, SeifPass, is a secure password manager which implements Pythia on behalf of such companies. SeifPass allows users to be proactive they do not need to depend on anyone to secure their passwords the right way.

Henry Liu, Justin Kaashoek, and Siye Zhu

Scalable logging algorithm for in-memory database systems

Mentor: Xiangyao Yu

Project suggested by Prof. Srini Devadas

Logging is a vital part of ensuring system dependency and mitigating the damage of potential system crashes. However, with an exponentially increasing amount of critical data being processed on a daily basis, such as with ATM transactions or online retail sales, the need for better logging algorithms arises. We establish a baseline with serial logging using the widely-implemented ARIES protocol, and then compare it with two new algorithms that we developed called batch logging and parallel logging. Our results show that Batch logging was faster than series logging, but had the drawback of increased latency. Parallel logging achieved the best in both worlds with faster processing and no increase in latency.

Yatharth Agarwal and Vishnu Murale

Moving in next door: Network flooding as a side channel in cloud environments

Mentor: Dr. Jason Hennessey, Kyle Hogan, and Dr. Mayank Varia, Boston University

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

While side channel attacks have been successfully executed in cloud environments, network transmission remains a relatively unexplored channel for information leakage. This report demonstrates the presence of a signal in terms of bandwidth and packet loss. The relevant experiments were conducted with multiple virtual servers sharing a physical host in laboratory and production clouds. The encompassing project is poised to explore the exploitability of the signal, thus holding implications for the security and privacy of all cloud environments.

SESSION 12: COMPUTATIONAL AND PHYSICAL BIOLOGY

Laura Braverman

Protein determinants of chromosome domains

Mentor: Nezar Abdennur

Project suggested by Prof. Leonid Mirny

An experimental method called Hi-C quantifying the 3D position of various regions of the genome inside the nucleus has revealed that mammalian genomes are composed of domains that have higher contact probability with each other than with adjacent domains [1]. A more recent study using the same technique identified more specific peaks of high contact probability between pairs of loci on the genome. Often occurring near domain boundaries, these peaks have been interpreted as the bases or anchor points of physical loops in the genome [2]. The DNA elements defining these regions, domain boundaries and loop bases, seem to be closely linked with binding of various proteins related both to genome architecture and transcriptional regulation. Our study furthers our last years research into the differences between loops and domains and investigates the possible function of loops based on the nature of proteins that bind DNA in the vicinity of these architectural elements using publicly available ChIP-seq data. While we observed no significant enrichment of measured protein binding sites around domain boundaries or inside of loop and domain segments, we did see strong enrichment of proteins CTCF, SMC3, RAD21, YY1, and ZNF143 at loop bases. Subsequently we found a strong correlation between the localization of architectural proteins, CTCF and cohesin subunits, SMC3 and RAD21, to loop boundaries. We also found that in the absence of CTCF and RAD21, the transcription factors YY1 and ZNF143 are significantly less likely to occur, but that many sites enriched in CTCF and cohesin lacked YY1 and ZNF143. Additionally, RNA Polymerase II, which is responsible for messenger RNA synthesis of most genes and is enriched at transcription start sites is found only in a small portion of loop boundaries.

1. Dixon, Jesse R., et al. "Topological domains in mammalian genomes identified by analysis of chromatin interactions." Nature 485.7398 (2012): 376-380.
2. Rao, Suhas SP, et al. "A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping." Cell 159.7 (2014): 1665-1680.

Betsy Pu

Chromatin states at boundary elements

Mentor: Nezar Abdennur

Project suggested by Prof. Leonid Mirny

Next-generation sequencing technologies now allow genome-wide measurements of large sets of epigenomic profiles, including histone modifications and chromatin-bound factors. In turn, these profiles can be jointly used to segment the genome into a small number of chromatin states of functional significance. Our study analyzed chromatin state annotations in regions around genomic loci relevant to 3D genome organization. These loci include the boundaries of topological domains and the bases of chromatin loops in Hi-C maps. Domains are contiguous regions with enriched internal contact probability, and loops are detected as pairs of loci on the genome with peaks of high contact probability. Previously, we looked at the distribution of functional elements at these border regions. Here, we compared these chromatin state distributions between cell types (GM12878, lymphoblast cell line, and IMR90, a lung fibroblast line). We found that GM12878-specific loops were enriched for GM12878-specific transcription states, suggesting that cell type specific looping determines cell type specific gene expression. Additionally, GM12878 domain boundaries enriched in the architectural protein CTCF tend to possess more transcription-related chromatin states at their borders. We also analyzed whether the chromatin state distribution at one domain boundary or loop base affected the chromatin states at its corresponding domain boundary / loop base, and concluded that they are independent. Our results motivate further inquiry into the relationship between the 3D structure of the genome and gene regulation.

Krishna Suraj

Emergent chromosome organization in interphase from loop extrusion

Mentor: Dr. Geoffrey Fudenberg

Project suggested by Prof. Leonid Mirny

Topologically Associating Domains (TADs) are basic structures of human interphase chromosomes. TADs have been recently hypothesized to be formed by Loop Extruding Factors (LEFs) which interact with each other and with translocation-limiting Boundary Elements (BEs) to create three-dimensional polymer structures. Simulating the behavior of such processes currently requires multi-scale computational models. I will describe components of these models, conclusions drawn from simulations, and propose future investigations.

SESSION 13: COMPUTATIONAL NEUROSCIENCE

Albert Gerovitch

Metrics for comparing 3D neuron segmentations in expansion microscopy connectomics

Mentor: Dr. Adam Marblestone

Project suggested by Prof. Edward Boyden

Visualizing neurons is key to understanding the brains functions. This is particularly important for the new optical approach we are developing in connectomics using expansion microscopy, replacing the traditional approach of electron microscopy. However, there is no reliable method of assessing segmentation algorithms accurateness at capturing the entire neuron in detail. The goal of this project is to develop software tools to evaluate segmentation algorithms and provide feedback for developers to improve their methods. We created an application that uses the pixel error, rand error, and warping error metrics to assess segmentations of single neurons, and returns a result to developers. To further improve our feedback, we visualized the results from the metrics. We are also expanding our application to enable evaluation of multi-cell segmentations. The developed software will provide a resource for neuroscientists to advance their research with high accuracy.

Zachary Steinberg

Automatic segmentation of punctate 3D super-resolution microscopy data

Mentor: Daniel Goodwin

Project suggested by Prof. Edward Boyden

Expansion microscopy is a form of microscopy in which the sample is embedded in a hydrogel and expanded physically in space, enabling standard light-based microscopes to discern details normally too small to view. As a result, studying the connections between neurons to produce a computational model of a full neural circuit is possible and may help further our knowledge of how the brain operates. However, at high levels of expansion, attempts to image using fluorescent proteins separate the fluorescence enough to produce punctate data, complicating attempts to computationally tell apart multiple dendrites in the same space through normal tracing algorithms. We present a new method for segmenting previously-unsegmentable data using a membrane-based mask to apply a clustering algorithm as a proof-of-concept for a full segmentation algorithm on highly expanded data.

SESSION 14: MEDICAL INFORMATICS

Daniel Lu

Study of various synergistic drug mechanisms in disordered protein-related diseases

Mentor: Prof. Gil Alterovitz

Project suggested by Prof. Gil Alterovitz

Recently, there has been an increasing need for new drug development methods. Not only is the number of targetable protein regions small, but the rate of drug target discovery has been going down. What if there was a way to expand the range of targetable proteins and propel drug development? To find more effective treatments addressing more diseases, this study explores drugs that target a new class of proteins and combines these targets in new ways.

Intrinsically disordered proteins (IDPs) are important targets for disease treatment because of their disproportionate involvement in disease. Drug synergy occurs when two or more drugs are administered together and produce an enhanced treatment effect compared to a single drug, making it a desirable goal for disease treatment and another way

of expanding the possibilities of drug development. In this study, an analytical method of identifying drug pairs that act synergistically was developed, targeting IDP-specific features, especially molecular recognition features (MoRFs).

Using a program built on a support vector machine in combination with a spectrophore molecular structure representation, over 7,000 drugs structural similarities to an IDPs MoRF were ranked. Highly ranked drugs were selected as candidates for inhibiting the IDPs interaction through mimicking the MoRF and occupying its binding target. The same analysis was carried out on the MoRFs interactors to identify a second drug that could bind to the MoRF itself. Additional synergy mechanisms were explored following this general method.

The study was successful in identifying synergistic drug pair candidates for eight IDP-related diseases including human papillomavirus, Alzheimers disease, cancer, influenza, human immunodeficiency virus, E.coli-related diseases, type-2 diabetes, and transmissible spongiform encephalopathies (prion diseases). These results yielded statistically significant p-values all smaller than 10^{-14} as well as orthogonal confirmation of data through other studies. This method seems promising as an efficient first step in finding better treatments for many diseases.

Kara Luo

Computational modeling identifies biosynthetic modifications to improve drug inhibition against
Klebsiella pneumoniae

Mentor: Prof. Gil Alterovitz

Project suggested by Prof. Gil Alterovitz

The superbug *K. pneumoniae* is a virulent bacterium that is resistant to available antibiotics. Using the computational technology of virtual screening, here I identified drug molecules that potentially inhibit this bacterium. I specifically targeted PriA, a key enzyme that is critical for bacterial growth. PriA binds to the molecular recognition feature (MoRF) of an intrinsically disordered protein named single-stranded binding (SSB) protein. Initial screening of a large library of 10,000 drugs identified five molecules that mimic the structure of SSB MoRF and likely block the active binding site of PriA. One hit possesses known anti-bacterial activity and treats infectious pneumonia, supporting the validity of my computational approach. I then used the five drugs as prototypes and further discovered biosynthetic modifications showing similar structures. Further docking assays revealed several hits with substantially improved binding affinity, among which includes O589 in the biosynthetic class of antibiotics that has not yet been used to treat infectious pneumonia. These experiments thus reveal promising leads for developing drugs to treat *K. pneumoniae* infection. This study also demonstrates the power of combining multi-layer screening and biosynthetic modifications in finding novel drugs and improving the binding activity of existing drugs. Similar computational approaches may use the expanding structural information of various disease targets to uncover drug molecules that suppress other antimicrobial-resistant micro-organisms.

Arul Prasad

Targeting viral envelope proteins: an application to the Zika virus

Mentor: Prof. Gil Alterovitz

Project suggested by Prof. Gil Alterovitz

The Zika virus has been declared as a global health emergency by the World Health Organization due to its correlation with microcephaly in babies and GuillainBarre Syndrome. As of April 27th, 2016, 55 countries had reported continuing mosquito borne transmission of the virus. Thus, the need for a solution to this issue is critical. This study uses an svm pipeline to sort through a list of FDA approved drug compounds and screen those with the highest binding probabilities with the three dimensional (PDB) structure of the Zika virus, specifically the envelope protein. The proposed drug compound was calculated to have a 0.917 binding probability. A 0.897 binding probability for a drug with tuberculosis worked previously in lab testing. Further binding simulations yielded relatively high docking scores. The results of this study propose an FDA approved drug compound to combat the Zika virus, enabling it to be fast-tracked pending successful lab testing, as well as a general method that can be used to find drugs for any viral disease.

SESSION 15: MEDICAL INFORMATICS

Andrew Gritsevskiy and Adithya Vellal

Compression and integration of human genomic variants into smart EHR systems

Mentor: Prof. Gil Alterovitz

Project suggested by Prof. Gil Alterovitz

With the evolution of Next Generation Sequencing (NGS) technologies, obtaining genomic data has become easier and cheaper than ever before. This data is becoming extremely valuable to medical institutions due to its central importance to precision medicine. Analysis of human genetic variation can allow physicians to determine and recommend personalized drug treatments to each and every one of their patients. However, genomic data is massive, and storage and maintenance of solely the variants in genomic data, a small fraction of the whole genome, is still extremely expensive.

We analyze the variants in genomic data to determine how to effectively exploit biological redundancies in order to losslessly compress the files which store these variants. We determine that patterns found in Single Nucleotide Polymorphisms (SNPs) can be utilized to compress this data, and our basic frequency analysis programs result in compressed files that are 28% of the size of the original files. We also integrate these files into a smart genomics API which will allow addition of genomic data into electronic health records in the future. Finally, we discuss initial progress on an unsupervised deep learning algorithm which will allow for greatly increased compression of these files as well as a better understanding of the biological patterns found in genomic variants.

John Flahive

Providing clinical decision support to medical providers through interpretation of gene-drug interactions

Mentor: Prof. Gil Alterovitz

Project suggested by Prof. Gil Alterovitz

Pharmacogenomics, the study of how genes affect drug response, is a critical, new field in science. Similarly to precision medicine, Pharmacogenomics is important in prescribing patients safe drugs, which will not cause adverse reactions, a significant cause of deaths and hospitalizations in the United States. This presentation introduces a new

pharmacogenomics portal for medical providers that provides patient reports and clinical decision support, based on a DIGITizE Pilot Implementation Guide.

James Jusuf

Exploring the effects of CTCF binding site mutations on transcriptional regulation

Mentor: Dr. Mahmoud Ghandi, Broad Institute

Project suggested by Dr. Mahmoud Ghandi

Analyzing mutations in noncoding regions of DNA is often challenging because these regions do not translate directly into proteins themselves. Instead, noncoding DNA often interacts with proteins and cofactors to regulate the degree to which genes are expressed. CCCTC-binding factor (CTCF) is a protein known to be involved with organizing the 3D structure of DNA. CTCF binds to a specific DNA sequence motif, and CTCF function may be impaired when the binding site is blocked or mutated. Previous studies have shown that methylation of certain binding sites locally impairs CTCF function and causes oncogenesis. We hypothesize that nucleotide alterations in the binding site impair CTCF function in a similar manner. We develop a program that utilizes gkm-SVM, a package developed by Dr. Ghandi, to computationally predict the change in binding affinity given a single nucleotide substitution and the location in the genome at which it occurs. We use the program to identify the mutations that significantly affect CTCF binding activity given a list of mutations present across hundreds of cancer cell lines. Then, we intersect the mutation data with monoallelic expression data to identify sites at which the two phenomena are possibly correlated. In the future, we plan to intersect the mutation data with other datasets, including tumor dependencies published by the Achilles Project.