



Photo: Slava Gerovitch

2020 Tenth Annual
Fall-Term PRIMES Conference
October 17-18, 2020

2020 Fall-Term PRIMES Virtual Conference

Saturday, October 17, 2020

Mathematics

10:20 am: Welcoming Remarks

Prof. Michel Goemans, Head of the MIT Mathematics Department

Prof. Pavel Etingof, PRIMES Chief Research Advisor

Dr. Slava Gerovitch, PRIMES Program Director

10:35 am – 11:55 am: Session 1

- **Yunseo Choi**, “Racial impact on infections and deaths due to COVID-19 in New York City” (mentor Prof. James Unwin, University of Illinois at Chicago)
- **William Li**, “Lebesgue measure preserving Thompson’s monoid” (mentor Prof. Sergiy Merenkov, CCNY – CUNY)
- **Joshua Guo, Karthik Seetharaman, and Ilaria Seidel**, “Upho posets” (mentor Yibo Gao)
- **Quanlin (Andy) Chen, Tianze (Peter) Jiang, and Yuxiao (Tom) Wang**, “On the generational behavior of Gaussian binomial coefficients at root of unity” (mentor Calder Morton-Ferguson)
- **Quanlin (Andy) Chen**, “The center of the q -Weyl algebra over rings with torsion” (mentor Calder Morton-Ferguson)
- **Yuxiao (Tom) Wang**, “Asymptotics for iterating the Lusztig-Vogan bijection for GL_n on dominant weights” (mentor Calder Morton-Ferguson)

12:00 pm – 1:00 pm: Session 2

- **Alexander Zhang**, “On quasisymmetric functions with two bordering variables” (mentor Andrey Khesin)
- **Holden Mui**, “Generating symmetric functions of every arity” (mentor Dr. Zeb Brady)
- **Ankit Bisain**, “The Bernardi formula for non-transitive deformations of the braid arrangement” (mentor Eric Hanson, Brandeis University)
- **William Shi**, “Refinements of product formulas for volumes of flow polytopes” (mentor Prof. Alejandro Morales, UMass Amherst)

1:15 pm – 1:45 pm: Virtual Chat with PRIMES Chief Research Advisor Prof. Pavel Etingof and Head Mentor Dr. Tanya Khovanova (for students and parents)

2:00 pm – 2:50 pm: Session 3

- **Vishaal Ram**, “Applications of a modified age-structured SIR model” (mentor Prof. Laura Schaposnik, University of Illinois at Chicago)
- **Srinath Mahankali**, “Velocity inversion using the quadratic Wasserstein metric” (mentor Dr. Yunan Yang, Courant Institute, New York University)
- **Samuel Florin, Matthew Ho, and Rahul Thomas**, “Group testing via zero-error channel capacity” (mentor Dr. Zilin Jiang)

3:00 pm – 4:00 pm: Session 4

- **Justin Wu**, “Borel cohomology of S^n mapping spaces” (mentor Ishan Levy)
- **Alvin Chen**, “K-semistability of smooth toric Fano varieties” (mentor Kai Huang)
- **Honglin Zhu**, “Irreducible characters for Verma modules for the orthosymplectic Lie superalgebra $osp(3|4)$ ” (mentor Arun Kannan)
- **Kenta Suzuki**, “Value sharing of meromorphic functions” (mentor Prof. Michael Zieve, University of Michigan)

4:00 pm – 5:10 pm: Session 5

- **Brian Liu**, “The geometry of harmonic maps in genus one” (mentor Prof. Dhruv Ranganathan, Cambridge, UK)
- **Fiona Abney-McPeck, Serena An, and Jakin Ng**, “The Stembridge equality for skew dual stable Grothendieck polynomials” (mentor YiYu (Adela) Zhang)
- **Adithya Balachandran, Andrew Huang, and Simon Sun**, “Product expansions of q -character polynomials” (mentor Dr. Nir Gadish)

2020 Fall-Term PRIMES Virtual Conference

Sunday, October 18, 2020

Computer Science

9:25 am: Welcoming Remarks

Dr. Slava Gerovitch, PRIMES Program Director

Prof. Srinivasa Devadas, PRIMES Computer Science Section Coordinator

9:35 am – 10:35 am: Session 6

- **Beining (Cathy) Zhou**, “A High-order cumulant-based sparse ruler for improved lag generation” (mentor Hanshen Xiao)
- **Matthew Ding**, “Relay protocol for approximate Byzantine consensus” (mentor Hanshen Xiao)
- **Linda Chen**, “Reducing round complexity of Byzantine broadcast” (mentor Jun Wan)
- **Jason Yang**, “On updating and querying submatrices” (mentor Jun Wan)

10:50 am – 12:00 pm: Session 7

- **Benjamin Chen**, “Towards practical ambiguity sets” (mentor Kyle Hogan)
- **Yavor Litchev and Abigail Thomas**, “Hybrid privacy scheme” (mentor Yu Xia)
- **Ho Tin (Alex) Fan and Alvin Lu**, “Parallel batch-dynamic subgraph maintenance” (mentors Prof. Julian Shun and Jessica Shi)
- **Kevin Zhao**, “Text is an image: Augmentation via embedding mixing” (mentors Prof. Anna Rumshisky and Vladislav Lialin, UMass Lowell)

12:15 pm – 12:45 pm : Virtual Chat with PRIMES Computer Science Section Coordinator Prof. Srinivasa Devadas and mentors (for students and parents)

Bioinformatics and Computational Biology

1:00 pm – 2:05 pm : Session 8

- **Anish Mudide**, “SARS-CoV-2 drug discovery based on intrinsically disordered regions” (mentor Prof. Gil Alterovitz)
- **Varun Suraj**, “Developing clinical decision support systems for cancer and COVID-19 precision medicine” (mentor Prof. Gil Alterovitz)

- **Daniel Xu**, “Network based digital contact tracing and testing strategies for the COVID-19 pandemic” (mentor Dr. Jesse Geneson, Iowa State University)
- **Tanisha Saxena and Daniel Xu**, “Graph alignment-based protein comparison” (mentor Younhun Kim)

2:20 pm – 3:20 pm: Session 9

- **Sarah Chen**, “In silico prediction of retained intron-derived neoantigens in leukemia” (mentors Dr. Nicoletta Cieri, Dana-Farber Cancer Institute, and Kari Stromhaug)
- **Mikhail Alperovich**, “Data driven quality control for single-cell RNA sequencing analysis” (mentor Dr. Ayshwarya Subramanian, Broad Institute)
- **Neil Chowdhury**, “Modeling the effect of histone methylation on chromosomal organization in colon cancer cells” (mentor Dr. Maxim Imakaev)
- **Vishnu Emami**, “Statistical ranking model for candidate genes in rare genetic disorders” (mentors Dr. Klaus Schmitz-Abe and Dr. Pankaj Agrawal, Boston Children's Hospital)

Mathematics

3:35 pm – 4:45 pm: Session 10

- **Jessica Zhang**, “Tight contact structures on the solid torus” (mentor Prof. Zhenkun Li, Stanford University)
- **William Qin**, “Colored HOMFLY polynomials of genus-2 pretzel knots” (mentor Yakov Kononov, Columbia University)
- **Daniel Hong, Hyunwoo (David) Lee, and Alex Wei**, “Optimal solutions and ranks in the max-cut SDP” (mentor Dr. Diego Cifuentes)
- **Richard Chen, Jason Tang, and Nathan Xiong**, “Few distance sets in l_p spaces and l_p product spaces” (mentor Feng Gui)

5:00 pm – 6:15pm : Session 11

- **Katherine Tung**, “The Sperner property for 132-avoiding intervals in the weak order” (mentor Christian Gaetz)
- **Andrew Cai**, “Ratios of Naruse-Newton coefficients obtained from descent polynomials” (mentor Pakawut Jiradilok)
- **Sean Li**, “The Penney’s Game with group action” (mentor Dr. Tanya Khovanova)
- **Tae Kyu Kim**, “On generalized Carmichael numbers” (mentor Yongyi Chen)
- **William Wang**, “Planar embeddings of periodic time-dependent graphs” (mentor Dr. Jesse Geneson, Iowa State University)

2020 PRIMES CONFERENCE ABSTRACTS

SATURDAY, OCTOBER 17

SESSION 1:

Yunseo Choi

Racial impact on infections and deaths due to COVID-19 in New York City

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

Project suggested by Prof. James Unwin, University of Illinois at Chicago

Redlining is the discriminatory practice whereby institutions avoided investment in certain neighborhoods due to their demographics. Here we explore the lasting impacts of redlining on the spread of COVID-19 in New York City (NYC). Using data available through the Home Mortgage Disclosure Act, we construct a redlining index for each NYC census tract via a multi-level logistical model. We compare this redlining index with the COVID-19 statistics for each NYC Zip Code Tabulation Area. Accurate mappings of the pandemic would aid the identification of the most vulnerable areas and permit the most effective allocation of medical resources, while reducing ethnic health disparities.

William Li

Lebesgue measure preserving Thompson's monoid

Mentor: Prof. Sergiy Merenkov, CCNY – CUNY

Project suggested by Prof. Sergiy Merenkov, CCNY – CUNY

In this talk we define Lebesgue measure preserving Thompson's monoid \mathbb{G} and study its algebraic and dynamical properties. This study is at an intersection of two subjects of research, namely Lebesgue measure preserving interval maps of $[0, 1]$ onto itself and Thompson's group \mathbb{F} , which is the group of piecewise affine maps from $[0, 1]$ onto itself whose derivatives are integer powers of 2 and points at which the derivatives are discontinuous are dyadic numbers. Monoid \mathbb{G} is similar to \mathbb{F} except that the derivatives of piecewise affine maps can be negative. The main results presented in this talk are summarized as follows.

- We show that any continuous measure preserving map can be approximated by a map in \mathbb{G} with any required precision. Moreover, we show that the approximating map in \mathbb{G} can be locally eventually onto (LEO) and achieve any target value of entropy that is at least 2.
- We show that for any element of \mathbb{G} , topological mixing (TM) is equivalent to LEO and every dyadic point is preperiodic. Thus, any map in \mathbb{G} is Markov. We show that for maps in a specific subset of \mathbb{G} there exist periodic points with period of 3, an essential feature of chaotic maps. We characterize periods of periodic points of other maps in \mathbb{G} .

The main results of this talk improve several results of recent papers in the literature. For example, we show that the subset of \mathbb{G} that is both LEO and Markov is dense in the set of continuous measure preserving maps. Because \mathbb{G} is a subset of piecewise affine continuous measure preserving maps, this result is stronger than the state-of-art result, which shows that piecewise affine continuous measure preserving maps that are both LEO and Markov is dense in the set of continuous measure preserving maps.

At an intersection of these two subjects of research, the talk demonstrates an interesting interplay between algebraic and dynamical settings. For example, in general, LEO implies TM and the converse does not hold; however, we show that for any element of \mathbb{G} , TM is equivalent to LEO and any map in \mathbb{G} is Markov. As another example, we show that the algebraic structure of \mathbb{G} leads to a simple characterization of periods of periodic points of maps in \mathbb{G} .

Joshua Guo, Karthik Seetharaman, and Ilaria Seidel

Upho posets

Mentor: Yibo Gao

Project suggested by Yibo Gao

Upper homogeneous finite type (upho) posets are a large class of partially ordered sets with the property that the upper principal order filter at every vertex is isomorphic to the whole poset. Well-known examples include k -array trees, the grid graphs, and the Stern poset. Very little is known about upho posets in general. In this talk, we construct upho posets with Schur-positive Ehrenborg quasisymmetric functions, whose rank-generating functions have rational poles and zeros. We also categorize the rank-generating functions of all planar upho posets. Finally, we prove the existence of an upho poset with uncomputable rank-generating function.

Quanlin (Andy) Chen, Yuxiao (Tom) Wang, Tianze (Peter) Jiang

On the Generational Behavior of Gaussian Binomial Coefficients at Root of Unity

Mentor: Calder Morton-Ferguson

Project suggested by Calder Morton-Ferguson

In this talk, we introduce the generational behavior of Gaussian binomial coefficients at roots of unity, which shadows the relationship between the reductive algebraic group in prime characteristic and the quantum group at roots of unity. We present three ways of obtaining integer values from Gaussian binomial coefficients at roots of unity. We rigorously define the “generations” in this context and illustrate such behavior at odd primes power. Moreover, we present some extended discussions on the vanishing, valuation, and sign behavior under the big picture of generations.

Quanlin (Andy) Chen

The Center of the q -Weyl Algebra over Rings with Torsion

Mentor: Calder Morton-Ferguson

Project suggested by Calder Morton-Ferguson

We will discuss the centers of the Weyl algebra, q -Weyl algebra, and the “first q -Weyl algebra” over the quotient of the ring $\mathbb{Z}/p^N\mathbb{Z}[q]$ by some polynomial $P(q)$. Through this, we generalize and “quantize” part of a result by Stewart and Vologodsky on the center of the ring of differential operators on a smooth variety over $\mathbb{Z}/p^n\mathbb{Z}$. We present that a corresponding Witt vector structure appears for $P(q)$ that are irreducible modulo p and compute the extra term for special $P(q)$ with particular properties, answering a question by Bezrukavnikov of possible interpolation between two known results.

Yuxiao (Tom) Wang

Asymptotics for iterating the Lusztig-Vogan Bijection for GL_n on Dominant Weights

Mentor: Calder Morton-Ferguson

Project suggested by Calder Morton-Ferguson

This talk will introduce Lusztig-Vogan Bijection, which was conjectured by Lusztig and Vogan and proved by Bezrukavnikov. It is a bijection between the set of dominant weights for any reductive group G over an algebraically closed field and the set of irreducible G -equivalent bundles on nilpotent orbits, critical for the representation theory. We iterate the algorithm of computing the Lusztig-Vogan Bijection Type A (GL_n) on dominant weights, which is proposed by Achur and simplified by Rush. We will illustrate the asymptotic behavior between the number of iterations for an input and the length of the input, and a recursive formula to compute the slope of the asymptote. This serves as another contribution to understanding the Lusztig-Vogan Bijection from a combinatorial perspective and a first step in understanding the iterative behavior of the Type A Lusztig-Vogan Bijection.

SESSION 2:

Alexander Zhang

On quasisymmetric functions with two bordering variables

Mentor: Andrey Khesin

Project suggested by Prof. Darij Grinberg, Drexel University

We study a family of formal power series $K_{n,\Lambda}$, parameterized by n and $\Lambda \subseteq [n]$, that largely resemble quasisymmetric functions. This family was conjectured to have the property that the product $K_{n,\Lambda}K_{m,\Omega}$ of any two functions $K_{n,\Lambda}$ and $K_{m,\Omega}$ from the family can be expressed as a linear combination of other functions from the family. We prove that this is indeed the case and that the span of the $K_{n,\Lambda}$'s form an algebra.

Holden Mui

Generating symmetric functions of every arity

Mentor: Dr. Zeb Brady

Project suggested by Dr. Zeb Brady

When can we create symmetric operations for every number of variables, given a small collection of starting operations? After presenting examples of operations, symmetric operations, and operation composition, we dive into two examples.

Ankit Bisain

The Bernardi formula for non-transitive deformations of the braid arrangement

Mentor: Eric Hanson, Brandeis University

Project suggested by Prof. Olivier Bernardi, Brandeis University

We consider a specific type of hyperplane arrangements, known as *deformations of the braid Arrangement*, and *boxed trees*, which are rooted labelled trees paired with a partition of their nodes. Bernardi has given a general formula to compute the number of regions of a deformation of the braid arrangement as a signed sum over boxed trees. We prove that the contribution to this sum of the set of boxed trees sharing an underlying tree is 0 or ± 1 and give an algorithm for computing this value. We then restrict to arrangements which we call *nested Ish arrangements* and construct a sign-reversing involution which reduces Bernardi's signed sum to the enumeration of a set of rooted labeled trees in this case.

William Shi

Refinements of product formulas for volumes of flow polytopes

Mentor: Prof. Alejandro Morales, UMass Amherst

Project suggested by Prof. Alejandro Morales, UMass Amherst

Flow polytopes are an important class of polytopes in combinatorics whose lattice points and volumes have interesting properties and relations. The Chan-Robbins-Yuen (CRY) polytope is a flow polytope with normalized volume equal to the product of consecutive Catalan numbers. Zeilberger proved this by evaluating the Morris constant term identity, but no combinatorial proof is known. There is a refinement of this formula that splits the largest Catalan number into Narayana numbers, which Mészáros gave an interpretation as the volume of a collection of flow polytopes. We introduce a new refinement of the Morris identity, compute an explicit product formula, and give combinatorial interpretations in terms of the volume of a collection of flow polytopes. Our results generalize Mészáros's construction to a recent flow polytope interpretation of the Morris identity by Corteel-Kim-Mészáros.

SESSION 3:

Vishaal Ram

Applications of a modified age-structured SIR model

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

Project suggested by Prof. Laura Schaposnik, University of Illinois at Chicago

We present a modified age-structured SIR model based on known patterns of social contact and distancing measures within Washington, USA. We find that population age-distribution has a significant effect on disease spread and mortality rate, and contribute to the efficacy of age-specific contact and treatment measures. We consider the effect of relaxing restrictions across less vulnerable age-brackets, comparing results across selected groups of varying population parameters. Moreover, we analyze the mitigating effects of vaccinations and examine the effectiveness of age-targeted distributions. Lastly, we explore how our model can be applied to other states to reflect social-distancing policy based on different parameters and metrics.

Srinath Mahankali

Velocity inversion using the quadratic Wasserstein metric

Mentor: Dr. Yunan Yang, Courant Institute, New York University

Project suggested by Dr. Yunan Yang, Courant Institute, New York University

Full-waveform inversion (FWI) is a method used to determine properties of the Earth from information on its surface. We use the squared Wasserstein distance (squared W_2 distance) as an objective function to invert for the velocity as a function of position in the Earth, and we discuss its convexity with respect to the velocity parameter. In one dimension, we consider constant, piecewise increasing, and linearly increasing velocity models as a function of position, and we show the convexity of the squared W_2 distance with respect to the velocity parameter on the interval from zero to the true value of the velocity parameter when the source function is a probability measure. Furthermore, we consider a two-dimensional model where velocity is linearly increasing as a function of depth and prove the convexity of the squared W_2 distance in the velocity parameter on large regions containing the true value. We discuss the convexity of the squared W_2 distance compared with the convexity of the squared L^2 norm, and we discuss the relationship between frequency and convexity of these respective distances. We also discuss multiple approaches to optimal transport for non-probability measures by first converting the wave data into probability measures.

Samuel Florin, Matthew Ho, and Rahul Thomas

Group testing via zero-error channel capacity

Mentor: Dr. Zilin Jiang

Project suggested by Dr. Zilin Jiang

We discuss a problem in group testing, and describe a connection between this problem and the problem of finding the zero-error capacity of a communication channel in information theory. We then give an equivalent optimization problem. Finally, we present our results in bounding the cardinality of this optimization problem.

SESSION 4:

Justin Wu

Borel cohomology of S^n mapping spaces

Mentor: Ishan Levy

Project suggested by Prof. Haynes Miller

The mapping space X^{S^n} of continuous functions from S^n to X is a natural space to study. We produce approximations to the mod 2 cohomology and the $SO(n+1)$ mod 2 Borel cohomology. They are approximations in the sense that it agrees with the actual cohomology when X is an Eilenberg MacLane space $K(\mathbb{Z}/2, m)$. Along the way I will give a brief introduction to cohomology, Steenrod algebras, and spectral sequences.

Alvin Chen

K -semistability of smooth toric Fano varieties

Mentor: Kai Huang

Project suggested by Kai Huang

Jiang conjectured that Tian's α -invariant for n -dimensional K -semistable smooth Fano varieties has a gap between $\frac{1}{n}$ and $\frac{1}{n+1}$. We prove this conjecture in the toric case using a combinatorial interpretation assuming a simplified version of Ewald's conjecture. We also prove a classification and a construction for all possible values of the α -invariant for these varieties.

Honglin Zhu

Irreducible characters for Verma modules for the orthosymplectic Lie superalgebra $\mathfrak{osp}(3|4)$

Mentor: Arun Kannan

Project suggested by Arun Kannan

We introduce the fundamentals of the representation theory of Lie algebras. Then, we present the problem of finding Jordan-Hölder multiplicities of Verma modules of basic Lie superalgebras. We outline the strategy used to solve this problem in the case of the orthosymplectic Lie superalgebra $\mathfrak{osp}(3|4)$.

Kenta Suzuki

Value sharing of meromorphic functions

Mentor: Prof. Michael Zieve, University of Michigan

Project suggested by Prof. Michael Zieve, University of Michigan

Let p and q be nonconstant meromorphic functions on \mathbb{C}^m . We show that if p and q have the same preimages as one another, counting multiplicities, at each of four nonempty pairwise disjoint subsets S_1, \dots, S_4 of \mathbb{C} , then p and q have the same preimages as one another at each of infinitely many subsets of \mathbb{C} , and moreover $g(p) = g(q)$ for some nonconstant rational function $g(x)$ whose degree is bounded in terms of the sizes of the S_i 's. This result is new already when $m = 1$, and it implies many previous results about the extent to which a meromorphic function is determined by its preimages of a few points or a few small sets, in addition to yielding new consequences such as a classification of all possibilities when two of the S_i 's have size 1.

SESSION 5:

Brian Liu

The geometry of harmonic maps in genus one

Mentor: Prof. Dhruv Ranganathan

Project suggested by Prof. Dhruv Ranganathan

Tropicalization is a way to convert algebraic maps into piecewise linear functions. In this presentation, we give necessary conditions for a piecewise linear function to be the result of a tropicalization of an algebraic map. In particular, we claim that the function must be balanced and well-spaced.

Fiona Abney-McPeck, Serena An, and Jakin Ng

The Stembridge equality for skew dual stable Grothendieck polynomials

Mentor: YiYu (Adela) Zhang

Project suggested by Prof. Darij Grinberg, Drexel University

The Schur polynomials s_λ are essential in understanding the representation theory of the general linear groups. They also describe the cohomology ring of the Grassmannians. For $\rho = (n, n-1, \dots, 1)$ a staircase shape and $mu \subseteq \rho$ a sub-tableau, the Stembridge equality states that $s_{\rho/\mu} = s_{\rho/\mu^T}$. This equality provides information about the symmetry of the cohomology ring. The dual stable Grothendieck polynomials g_λ are variants of the Schur polynomials developed by Buch, Lam, and Pylyavskyy, which describe the K-theory of the Grassmannians. The goal of our project is to prove the Stembridge equality for the skew dual stable Grothendieck polynomials. Using the Hopf algebra structure of the ring of symmetric functions and a generalized Littlewood-Richardson rule, we show that $g_{\rho/(k)} = g_{\rho/(1^k)}$, where k is the straight shape with k boxes.

Adithya Balachandran, Andrew Huang, and Simon Sun

Product expansions of q -character polynomials

Mentor: Dr. Dr. Nir Gadish

Project suggested by Dr. Nir Gadish

This project deals with certain class functions defined simultaneously on finite general linear groups of all dimensions, which we interpreted as statistics on matrices. It has been previously shown that these class functions are closed under multiplication, and we work towards computing the structure constants of this ring of functions. We find a closed form for evaluations of the functions and determine their product expansions in the cases associated with identity matrices and single Jordan blocks. These coefficients are of importance as they compute and demonstrate stability phenomena of joint moments as well as the irreducible decomposition of tensor products of representations of finite general linear groups.

SUNDAY, OCTOBER 18

SESSION 6:

Beining (Cathy) Zhou

A High-order cumulant-based sparse ruler for improved lag generation

Mentor: Hanshen Xiao

Project suggested by Hanshen Xiao

The sparse ruler is a classical problem in theoretical computer science that seeks to find a sequence of integers, or integer marks on a ruler, such that the differences of the marks can generate as many consecutive integers lag as possible. Most recently, the sparse ruler is generalized with high-order cumulants to optimize the lags generated. We will present a new construction of the ruler for the 4-th and 6-th order and extend it to the $2q$ -th order by layering. Compared to two other prominent methods, this approach significantly improves the lag generation by exploiting more sign combinations of the permutation invariants. This problem and the construction could also be applied to spatial signal processing to provide more efficient sensor array geometries.

Matthew Ding

Relay protocol for approximate Byzantine consensus

Mentor: Hanshen Xiao

Project suggested by Hanshen Xiao

This paper presents a novel algorithm for Approximate Byzantine Consensus (ABC), called Relay-ABC. The algorithm allows machines to achieve approximate consensus to arbitrary exactness in the presence of byzantine failures. The algorithm relies on the novel usage of a relayed messaging system and signed messages with unforgeable signatures that are unique to each node. The use of signatures and relays allows the strict necessary network conditions of traditional approximate byzantine consensus algorithms to be circumvented.

We also provide theoretical guarantees of validity and convergence for Relay-ABC. To do this, we utilize the idea that the iteration of states in the network can be modelled by a sequence of transition matrices. We extend previous methods, which use transition matrices to prove ABC convergence, by having each state vector model not just one iteration, but a set of D iterations, where D is a diameter property of the graph. This allows us to accurately model the delays of messages inherent within the relay system.

Linda Chen

Reducing round complexity of Byzantine broadcast

Mentor: Jun Wan

Project suggested by Jun Wan

Byzantine Broadcast is a variation of the Byzantine Agreement problem, an important topic in distributed systems. In Byzantine Broadcast, a designated leader must ensure that all honest users in a distributed system reach a consensus, even in the presence of some corrupt users. Improving the round complexity of Byzantine Broadcast has long been a focused challenge. Under dishonest majority, previous work has shown a protocol that achieves expected constant round complexity. In this project, we describe a method to further reduce the round complexity of this dishonest majority protocol. In addition, we introduce similar protocols for the honest majority settings, which achieve lower round complexities in comparison to the state of the art results.

Jason Yang

On updating and querying submatrices

Mentor: Jun Wan

Project suggested by Jun Wan

In this talk, we study the d -dimensional update-query problem. We provide lower bounds on update and query running times, assuming a long-standing conjecture on min-plus matrix multiplication, as well as algorithms that are close to the lower bounds. Given a d -dimensional matrix, an *update* changes each element in a given submatrix from x to $x \nabla v$, where v is a given constant. A *query* returns the Δ of all elements in a given submatrix. We study the cases where ∇ and Δ are both commutative and associative binary operators. When $d = 1$, updates and queries can be performed in $O(\log N)$ worst-case time for many (∇, Δ) by using a segment tree with lazy propagation. However, when $d \geq 2$, similar techniques usually cannot be generalized. We show that if min-plus matrix multiplication cannot be computed in $O(N^{3-\varepsilon})$ time for any $\varepsilon > 0$ (which is widely believed to be the case), then for $(\nabla, \Delta) = (+, \min)$, either updates or queries cannot both run in $O(N^{1-\varepsilon})$ time for any constant $\varepsilon > 0$, or preprocessing cannot run in polynomial time. Finally, we show a special case where lazy propagation can be generalized for $d \geq 2$ and where updates and queries can run in $O(\log^d N)$ worst-case time. We present an algorithm that meets this running time and is simpler than similar algorithms of previous works.

SESSION 7:

Benjamin Chen

Towards practical ambiguity sets

Mentor: Kyle Hogan

Project suggested by Kyle Hogan

Perfectly private communication over the internet has always come at a cost of poor performance, and ambiguity sets have been proposed as a means to improve the performance while retaining some degree of privacy. The practicality of such sets was explored, using real internet forum activity scraped from Reddit. Several different methods of forming anonymity sets and defining an activity budget for users within the set are examined and evaluated. The effects of various variables on the performance and privacy of the resulting system were explored, and it was found that K-means clustering on users with a budget created by adding a predetermined multiple of the standard deviation had both decent performance while still retaining some degree of privacy. Gaussian mixture models, spectral clustering, and self-organizing (Kohonen) maps were also tested. These findings suggest that the usage of ambiguity sets in the real world is practical and demonstrates one way they could be realistically implemented through common algorithms.

Yavor Litchev and Abigail Thomas

Hybrid privacy scheme

Mentor: Yu Xia

Project suggested by Yu Xia

Local Differential Privacy (LDP) is an approach that allows a central server to compute on data from multiple users while maintaining the privacy of each user by adding noise to their inputs. LDP is a very efficient approach; however, as privacy increases, the accuracy of these computations decreases. Multi-Party Computation (MPC) is a process by which multiple parties can work together to compute the output of a function, without revealing their own information. MPC is very secure and accurate for such computations, but it is very slow. These costs and benefits become very apparent when one wishes to train a model with artificial intelligence while keeping source data private, due to the training being computationally intensive. In this project, we seek to harness the benefits of LDP and MPC to form a hybrid model that is accurate, secure, and fast. In our experimentation, we implement such a hybrid model on the example of training a support vector machine (SVM) on the famous MNIST handwritten digits dataset. This is accomplished by first training the SVM using LDP to quickly optimize the model, and then afterwards training the SVM with MPC to fine-tune the model. Our experimentation shows that our hybrid algorithm is much faster to train than a pure MPC algorithm and is also more precise than a pure LDP algorithm.

Ho Tin (Alex) Fan and Alvin Lu

Parallel batch-dynamic subgraph maintenance

Mentors: Prof. Julian Shun and Jessica Shi

Project suggested by Prof. Julian Shun

Counting certain subgraphs is a fundamental problem that is crucial in recognizing patterns in large graphs, such as social networks and biological interactomes. However, many real world graphs are constantly evolving and are subject to changes over time, and previous work on efficient parallel subgraph counting algorithms either do not support dynamic modifications or do not extend to general subgraphs. This talk presents a theoretically-efficient and demonstrably fast algorithm for parallel batch-dynamic 3-vertex subgraph counting, and the underlying data structure can be extended to counting 4-vertex subgraph counts as well. The algorithm maintains the h -index of the graph, or the maximum h such that the graph contains h vertices with degree at least h , and uses this to update subgraph counts through an efficient traversal of two-paths, or wedges. For a batch of size b , the algorithm takes $O(bh)$ expected work and $O(\log(bh))$ span with high probability.

Kevin Zhao

Text is an image: Augmentation via embedding mixing

Mentor: Prof. Anna Rumshisky and Vladislav Lialin, UMass Lowell

Project suggested by Prof. Anna Rumshisky, UMass Lowell

Data augmentation techniques are essential for computer vision, yielding significant accuracy improvements with little engineering costs. However, data augmentation for text has always been tricky. Synonym replacement techniques require a good thesaurus and domain-specific rules for synonym selection from the synset, while backtranslation techniques are computationally expensive and require a good translation model for the language in interest. In this talk, we present simple text augmentation techniques on the embeddings level, inspired by mixing-based image augmentations. These techniques are language-agnostic and require little to no hyperparameter tuning. We evaluate augmentation techniques on GLUE tasks and show that they significantly improve the score of the RoBERTa model.

SESSION 8:

Anish Mudide

SARS-CoV-2 drug discovery based on intrinsically disordered regions

Mentor: Prof. Gil Alterovitz

Project suggested by Prof. Gil Alterovitz

Currently, many drug discovery approaches follow the typical protein structure-function paradigm, designing drugs to bind to fixed three-dimensional structures. However, in recent years such approaches have failed to address drug resistance and limit the set of possible drug targets and candidates. For these reasons we instead focus on targeting protein regions that lack a stable structure, known as intrinsically disordered regions (IDRs). Such regions are essential to numerous biological pathways that contribute to the virulence of various viruses. In our early work, we focus on SARS-CoV-2 therapeutic discovery. Using modern molecular docking procedures and machine learning, we identify eleven new SARS-CoV-2 drug candidates targeting IDR mechanisms. Later, we investigate the use of learned molecular representations for biomimicry-based IDR drug discovery. We find that our model outperforms previous baselines in this area, demonstrating its promise for future work on disordered targets within SARS-CoV-2 or elsewhere.

Varun Suraj

Developing clinical decision support systems for cancer and COVID-19 precision medicine

Mentor: Prof. Gil Alterovitz

Project suggested by Prof. Gil Alterovitz

The extent of patient information available in electronic form has grown remarkably in the past few years. This includes availability of patient information through electronic health records and gene sequencing data. Both of these developments allow for the increased use of patient data in precision medicine. This project expands the use of clinical decision support platforms to take advantage of the increased availability of EHRs and genetic information to support physicians.

The first phase of the project incorporates multiple EHRs into the existing SMART Cancer Navigator platform. By doing so, physicians can identify any discrepancies between data sources, and can be sure that the information provided from multiple EHRs is completely up-to-date. The next phase of the project incorporates patient genetic information into the SMART Cancer Navigator platform. Gene sequencing results are stored in Variant Call Format (VCF) files. The application converts these VCF files into more standardized formats accessible by EHRs. It then reads the genetic data to display information such as diseases that could potentially be caused due to a patient's genetic variants. The final phase of the project takes the SMART Cancer Navigator beyond just the use in cancer clinical support to other types of diseases. In this case, the platform was extended for COVID-19 decision support. Patient medical conditions stored in EHRs were linked to potential risk factors for COVID-19 - along with research on these conditions - allowing users to view data on COVID-19's specific risk for the individual. By doing so, physicians can easily identify patient outcomes for particular risk factors.

Daniel Xu

Network based digital contact tracing and testing strategies for the COVID-19 pandemic

Mentor: Dr. Jesse Geneson, Iowa State University

Project suggested by Daniel Xu

We developed a model of digital contact tracing and testing strategies of COVID-19. We incorporated infectivity variations, test sensitivities, incubation period and asymptomatic cases to enhance the stochastic dynamic network based compartmental SEIR model. Our model is generated using a real world contact data set. The simulation results showed that the testing plus contact tracing can effectively reduce the outbreak size. We also use this model's simulation to find quarantine rules and testing strategies.

Tanisha Saxena and Daniel Xu

Graph alignment-based protein comparison

Mentor: Younhun Kim

Project suggested by Upasana Das Adhikari, Mass General Hospital, and Younhun Kim

We are in an era when DNA and proteins can be measured down to their building blocks. On the other hand, there are many algorithmic challenges due to the sheer size and complexity of data. Inspired by the question of identifying mechanisms of viral infection, we are interested in the problem of comparing pairs of proteins, given by amino acid sequences and traces of their 3-dimensional structure. While it is true that the problem of predicting and comparing protein function is one of the most famous unsolved problems in computational biology, we propose a heuristic which poses it as a simple alignment problem, which — after some linear-algebraic pre-processing — is amenable to a dynamic programming solution.

SESSION 9:

Sarah Chen

In silico prediction of retained intron-derived neoantigens in leukemia

Mentors: Dr. Nicoletta Cieri, Dana-Farber Cancer Institute, and Kari Stromhaug

Project suggested by Dr. Nicoletta Cieri, Dana-Farber Cancer Institute, and Kari Stromhaug

Alternative splicing is critical for the regulation and diversification of gene expression. Conversely, splicing dysregulation, caused by mutations in splicing machinery or splice junctions, is a hallmark of cancer. Tumor-specific isoforms are a potential source of neoantigens, cancer-specific peptides presented by human leukocyte antigen (HLA) class I molecules and potentially recognized by T cells. For cancers such as acute myeloid leukemia (AML) with a low mutation burden but widespread splicing aberrations, splice variants and retained introns (RIs) in particular, may broaden the number of suitable targets for immunotherapy. We developed a computational pipeline to predict RI-derived neoepitopes from tumor RNA-Seq. We first used the B721.221 B cell line as a model system, for which RNA-Seq, Ribo-Seq, and immunoproteome data from > 90 HLA class I monoallelic lines were available. We performed de novo transcriptome assembly with StringTie identifying on average 694 RI isoforms across 4 technical replicates. Using HLATHENA, we identified a median of 43 RI-derived neoepitopes predicted to bind with the most stringent cut-off (0.1% percentile rank) across 4 frequent HLA alleles. We validated 23 predicted neoepitopes using the alignment of ≥ 1 Ribo-Seq reads as a measure of mRNA translation. To further increase prediction accuracy, we are currently analyzing the HLA I immunopeptidome to define the features of predicted RIs more likely to be translated and presented. Finally, we are applying our prediction pipeline to AML cell lines and primary samples. Preliminary results on the Kasumi-1 cell line yielded 785 RI isoforms (containing 724 Kasumi-1-specific RIs and 54 shared with B721.221). Kasumi-1 displayed a 3x higher proportion of isoforms containing full RIs (as opposed to alternative 5' and 3' splice sites) than B721, in line with the biological relevance of RIs in particular in this disease setting. Accurate prediction of RI-derived neoantigens through our pipeline will contribute to the design of novel cancer immunotherapies.

Mikhail Alperovich

Data driven quality control for single-cell RNA sequencing analysis

Mentor: Dr. Ayshwarya Subramanian, Broad Institute

Project suggested by Dr. Ayshwarya Subramanian, Broad Institute

In recent years, single cell technologies have enabled breakthrough insights in biology by allowing us to investigate properties of individual cells. Single cell RNA sequencing (scRNAseq) profiles gene expression of individual cells, and allows us to understand cellular heterogeneity, and functions of diverse cell types. A typical scRNAseq data analysis pipeline includes multiple sequential steps, starting from quality control, feature selection, dimensionality reduction, clustering, cell type annotation, and downstream analysis to answer specific biological questions. Quality Control (QC) is an important first step where higher-quality data is retained by accounting for noise and other factors accompanying the process of data generation. In this project we surveyed commonly used QC methods, and developed alternative data-driven approaches. We benchmarked the methods on large public datasets, and evaluated their performance on ability to maximize retention of information for answering relevant downstream biological questions.

Neil Chowdhury

Modeling the effect of histone methylation on chromosomal organization in colon cancer cells

Mentor: Dr. Maxim Imakaev

Project suggested by Dr. George Spracklin, UMass Medical School, and Dr. Nezar Abdennur

Loop extrusion and compartmentalization are the two most important processes regulating the high-level organization of DNA in the cell nucleus. These processes are largely believed to be independent and competing (Sanborn et al. 2015; Nuebler et al. 2018). Chromatin consists of nucleosomes, which contain coils of DNA wrapped around histone proteins. Besides packing DNA, nucleosomes contain an “epigenetic code” — tails of histone proteins are chemically modified at certain positions to leave certain “histone marks” on the chromatin fiber. This talk explores the effect of the H3K9me3 histone modification, which typically corresponds to inactive and repressed chromatin, on genome structure. Interestingly, in H3K9me3 domains, there are much fewer topologically associating domains (TADs) than in other domains, and there is a unique compartmentalization pattern. A high-resolution polymer model simulating both loop extrusion and compartmentalization is created to explore these differences.

Vishnu Emani

Statistical ranking model for candidate genes in rare genetic disorders

Mentors: Dr. Klaus Schmitz-Abe and Dr. Pankaj Agrawal

Project suggested by Dr. Klaus Schmitz- Abe and Dr. Pankaj Agrawal

Genetic mutations are responsible for a significant number of rare diseases, and so investigating the genetic basis of various rare diseases has been a crucial area of study. Genetic analysis includes an examination of genotype data (including allele frequency, the gene's overall resistance to mutation, and its evolutionary conservation) as well as phenotype data (previously reported disease associations and mouse models) (Figure 1). The purpose of this project was to develop an automated algorithm, using a host of parameters, to rank the mutation candidates based on the computed scores for pathogenicity (one for genotype and another for phenotype). In order to train the model for the genotype score, the mutation data of 35 families were imported, along with the manual classification made by the geneticist. Regression models were trained using 25 predictors against the reported pathogenicity. After the predictors with insignificant weights were removed, a k-fold cross-validation was run on the model using other test sets. Receiver operating characteristic (ROC) analysis was performed, producing a graph of the false-positive rate vs. the true-positive rate, and the area under the curve (AUC) was calculated. The results showed a strong prediction accuracy of the model in the test sets (Average AUC 0.89). In order to produce the phenotype score, the genes were searched in a matrix that combined data from many genetic databases, and the key words of the gene phenotype were compared with the patient's phenotype, producing a score based on the specificity and depth of coverage. In the future, the models can be refined to take into account more genetic factors for more complex cases. Nevertheless, the preliminary results of the model's predictions are promising, and suggest that there is room for computational approaches to genetic mutation analyses.

SESSION 10:

Jessica Zhang

Tight contact structures on the solid torus

Mentor: Prof. Zhenkun Li, Stanford University

Project suggested by Prof. Zhenkun Li, Stanford University

One of the most fundamental open questions in contact geometry is the classification of all tight contact structures on a given 3-manifold (up to isotopy). If the manifold has boundary, this classification is based on the dividing set, which is a special set on the boundary. In this research project, we completely classified the tight contact structures for the solid torus with boundary. In particular, we were able to write down a closed formula for the number of non-isotopic tight contact structures on the solid torus in terms of the dividing set on the boundary. Previously, only a few special cases for the solid torus were known by Honda, and only one other manifold with boundary was completely classified, namely the 3-ball.

William Qin

Colored HOMFLY polynomials of genus-2 pretzel knots

Mentor: Yakov Kononov, Columbia University

Project suggested by Yakov Kononov, Columbia University

HOMFLY polynomials are one of the major knot invariants being actively studied. They are difficult to compute in the general case but can be far more easily expressed in certain specific cases. We examine in particular a type of knot called a pretzel knot, and in particular use a method reminiscent of normal polynomials, an analogue of finite differences, to help with this. We find a way to significantly decrease computation time of HOMFLY polynomials of genus-2 pretzel knots with certain large parameters, and conjecture some properties that could lead to a more general algorithm to quickly compute HOMFLY polynomials.

Daniel Hong, Hyunwoo (David) Lee, and Alex Wei

Optimal solutions and ranks in the max-cut SDP

Mentor: Dr. Diego Cifuentes

Project suggested by Dr. Diego Cifuentes

The max-cut problem is a classical graph theory problem which is NP-complete. The best polynomial-time approximation scheme relies on *semidefinite programming* (SDP). We study the conditions under which graphs of certain classes have rank 1 solutions to the max-cut SDP. We apply these findings to look at how solutions to the max-cut SDP behave under simple combinatorial constructions. Our results determine when solutions to the max-cut SDP for cycle graphs are rank 1. We find the solutions to the max-cut SDP of the vertex sum of two graphs. We then characterize the SDP solutions upon joining two triangle graphs by an edge sum.

Richard Chen, Jason Tang, and Nathan Xiong

Few distance sets in ℓ_p spaces and ℓ_p product spaces

Mentor: Feng Gui

Project suggested by Prof. Larry Guth

Kusner asked if $n + 1$ points is the maximum number of points in \mathbb{R}^n such that the ℓ_p distance between any two points is 1. We present an improvement to the best known upper bound when p is large in terms of n , as well as a generalization of the bound to s -distance sets. We also study equilateral sets in the ℓ_p sums of Euclidean spaces, deriving upper bounds on the size of an equilateral set for when $p = \infty$, p is even, and for any $1 \leq p < \infty$.

Katherine Tung

The Sperner property for 132-avoiding intervals in the weak order

Mentor: Christian Gaetz

Project suggested by Christian Gaetz

A well-known result of Stanley from 1980 implies that the weak order on a maximal parabolic quotient of the symmetric group S_n has the Sperner property; this same property was recently established for the weak order on all of S_n by Gaetz and Gao, resolving a long-open problem. In our research, we interpolate between these results by showing that the weak order on any 132-avoiding interval has the Sperner property.

This result is proven by exhibiting an action of sl_2 respecting the weak order on these intervals. As a corollary we obtain a new formula for principal specializations of Schubert polynomials. Our formula can be seen as a strong Bruhat order analogue of Macdonald's reduced word formula.

Andrew Cai

Ratios of Naruse-Newton coefficients obtained from descent polynomials

Mentor: Pakawut Jiradilok

Project suggested by Pakawut Jiradilok

We study the Naruse-Newton coefficients, which are obtained from expanding descent polynomials in a specific Newton basis introduced by Jiradilok and McConville. These coefficients C_0, C_1, \dots, C_s form an integer sequence associated with each finite set of positive integers. They are derived from the expansion of the descent polynomial through Naruse's Hook Length Formula. We characterize finite sets for which $\frac{C_a}{C_b}$ is minimized, strengthening an inequality from Jiradilok and McConville. We also study certain transformations of descent sets, and their effects on $\frac{C_a}{C_b}$. Finally, for fixed nonnegative integers $a < b$, we examine the set $R_{a,b}$ of all $\frac{C_a}{C_b}$ over finite sets of positive integers.

Sean Li

The Penney's Game with group action

Mentor: Dr. Tanya Khovanova

Project suggested by Dr. Tanya Khovanova

Consider equipping an alphabet A with a group action; this partitions the set of words into equivalence which we call patterns. Much of existing word avoidance theory can be generalized to these patterns. We use these generalizations to answer standard questions for the Penney's game on patterns and show non-transitivity for the game on patterns as the length of the pattern tends to infinity. We also analyze bounds on the pattern-based Conway leading number and expected wait time, and further explore the game under the cyclic and symmetric group actions.

Tae Kyu Kim

On generalized Carmichael numbers

Mentor: Yongyi Chen

Project suggested by Dr. Stefan Wehmeier, Mathworks

Given an integer k , define C_k as the set of integers $n > \max(k, 0)$ such that $an - k + 1 \equiv a \pmod{n}$ holds for all integers a . We establish various multiplicative properties of the elements in C_k and give a sufficient condition for the infinitude of $|C_k|$. Moreover, we prove that there are finitely many elements in C_k with one and two prime factors, and if we put certain restrictions on the prime factors, there are finitely many elements in C_k with three prime factors satisfying those restrictions. We also give conjectures about the growth rate of C_k with numerical evidence. We explore a similar question when both a and k are fixed and prove that for fixed integers $a \geq 2$ and k , if $(k, a) \neq (0, 2)$, there are infinitely many integers n such that $an - k \equiv a \pmod{n}$ by building off the works of P éter Kiss and Bui Minh Phong. Finally, we discuss the multiplicative properties of positive integers n such that Carmichael function $\lambda(n)$ divides $n - k$.

William Wang

Planar embeddings of periodic time-dependent graphs

Mentor: Dr. Jesse Geneson, Iowa State University

Project suggested by Dr. Jesse Geneson, Iowa State University

How do we extend normal planar graphs into structures with similar characteristics that vary with time? We introduce the notion of planarity for time-dependent graphs and consider three famous theorems on time-independent graphs to see if they still hold for our new constructs.