*Presenter

Asking Different Questions: A Justice-Oriented Approach to Research

Sarah McCullough*, Eli Alston-Stepnitz, Tara Pozzi, Colleen Sargent, Kalindi Vora, Maya Cruz, Sara Giordano, Ellie Legg

Asking Different Questions is a research training program that provides researchers with the intellectual tools to integrate values of justice and equity more deeply into their research agendas. Insights from feminist science & technology studies, ethnic studies, gender studies, and indigenous studies are taught, and participants apply these insights into their research. Two years into the program, evaluation reveals high levels of efficacy in achieving learning outcomes, increasing participant's sense of belonging, and improvements to participant's ability to apply insights to their own research.

Fair Data Representation for Machine Learning at the Pareto Frontier

Shizhou Xu*, Thomas Strohmer

As machine learning powered decision-making becomes increasingly important in our daily lives, it is imperative to strive for fairness of the underlying data processing. We propose a pre-processing algorithm for fair data representation via which L2-objective supervised learning algorithms result in estimations of the Pareto frontier between prediction error and statistical disparity. Particularly, the present work applies the optimal affine trans- port to approach the post-processing Wasserstein barycenter characterization of the optimal fair L2-objective supervised learning via a pre-processing data deformation. We call it Wasserstein pseudo-barycenter. Furthermore, we show that the Wasserstein geodesics from learning outcome marginals to their barycenter characterizes the Pareto frontier between L2-loss and total Wasserstein distance among the marginals. Thereby, an application of McCann interpolation generalizes the pseudo-barycenter to a family of data representations via which L2-objective supervised learning algorithms estimate the Pareto frontier. Numerical simulations underscore the advantages: (1) the pre-processing step is compositive with arbitrary L2-objective supervised learning methods and unseen data; (2) the fair representation protects data privacy by preventing access to the sensitive information; (3) the optimal affine maps are computationally efficient on high-dimensional data; (4) experimental results shed light on the fairness of L2-objective unsupervised learning via the proposed fair data representation.

Race discrepancies of arrests made by law enforcement

Sam Brinker*, Moss Le, Anastasia Goold, and Neil McRoberts

The 2020 Black Lives Matter protests have increased the calls for police oversight and reform. Previous studies have shown that black individuals are racially targeted by law enforcement while driving, but the data needed to prove this requires a Freedom of Information Act request to obtain. Using publicly available data, we found that black individuals have been overly represented in the arrest datasets of Bay County, Florida for the past 20 years while white individuals have been noticeably underrepresented. This methodology can be applied to aid community level efforts driving activism and potentially provide a citizen science approach for surveilling inequalities brought on by law enforcement.

Predicting the Formation of R-loops using a Formal Grammar

Jacob Lusk*, Dr. Mariel Vazquez

An R-loop is a three-stranded nucleic acid structure that can form during transcription. R-loop formation is affected by both DNA sequence and topology. Applying tools from computational linguistics we generate a formal grammar to model how DNA sequence affects the formation of R-loops. By generating a treebank from experimental data to determine the probabilities associated with a probabilistic context free grammar (PCFG), we aim to predict the formation of R-loops from a given genomic sequence.

The Effects of Environmental Changes on Cyanobacteria Diversity found in UC Davis Soils

Jazmine Reeves*, Sylvia Rodriguez*, Metzli Montero*

Cyanobacteria are astrobiological organisms of interest because of their ability to thrive in extreme environmental conditions, such as the thermal pools in Yellowstone and the frozen lakes of Antarctica. In our research, we investigated how the diversity and morphologies of cyanobacteria can be affected by heat, water distress, soil disruption, and changes in soil pH. Initial samples were found in the crusts of crushed granite pathways outside the EPS building on the UC Davis campus, where summer temperatures reached 110 °F. Various cyanobacteria have been identified with microscopy through their morphology including Nostoc, Anabaena, Phormidium, and Oscillatoria. This research is relevant to astrobiological studies on cyanobacteria resilience, our understanding of the origins of life on Earth, and the study of the potential for life on other planets with extreme environments. Individual studies revealed that variations in environment play a role in the diversity and phenotype expression in

cyanobacteria. Our next steps are to investigate cyanobacteria diversity through sequencing data of our various samples using MinION nanopore sequencing.

Prediction of Binding Affinity for Coronaviruses and their Receptors Using Machine Learning

Apurva Mishra*, Nathan Solomon

This project attempts to create a model to predict the binding affinity between the Receptor Binding Domain (RBD) of the SARS-CoV-2 virus and the human ACE2 receptor. The model relates biochemical properties of the RBD region to its binding affinity with the ACE2 receptor using machine learning in order to bypass the computationally expensive procedures of protein structure prediction and docking. In addition, it characterizes mutations of the virus at two significant locations for binding ability. The pipeline produced can be applied to other infectious diseases.

Analysis of SEIR Model in H1N1 Infection

Yuan Chen*; Yuying Zhang*; Paul Graggs*

In this poster, we are going to present an SEIR model for the H1N1 outbreak in a community region in Venezuela. We acquired our model and parameters from two published papers, and reproduced the model to compare with the real-world statistics.

Human Induced Pluripotent Stem Cell-Based Treatment in a Valproic Acid Rat Model of Autism

Pablo Juarez, Alexa D'Ambra, Vanessa Cervantes,Paul Knoepfler, Diasynou Fioravante and Verónica Martinez-Cerdeño

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder characterized by deficits in social interaction and stereotypical behavior. Deficits of Parvalbumin (PV+) expressing GABAergic interneurons in the prefrontal cortex of ASD patients is believed to contribute to the imbalance in the ratio between excitatory and inhibitory signaling (E/I). We hypothesize that substituting this lost population of PV+ cells in our ASD Valproic Acid (VPA) rat model will improve behavioral deficits by ameliorating E/I signaling. First, we differentiated Human Induced Pluripotent Stem Cells (hIPSC's) into Medial Ganglionic Eminence (MGE) cells (PV+ GABA precursor cells),and transplanted them into the medial PFC of our VPA rat model. We have four groups: Wild Type rats, VPA rats, VPA+dead MGE rats and VPA+MGE rats (n=10-11/group). Behavioral assays (social apparatus, elevated plus maze, open field and marble burry test) were performed 1-month post-transplant using tests that model core behavioral deficits seen in human ASD. hIPSC derived MGE

cells were able to survive and mature after transplantation. Patch clamp recordings 4 weeks post-transplant showed that MGE cell transplants had immature electrophysiological properties. Moreover, MGE cell transplants rescued social, anxiety and exploratory behavioral deficits in VPA rats.

Travelling Salesman Problem Algorithm
Jingren Xu/Yuan Chen

In Travelling salesman problem, we studied genetic operators such as crossover, mutation, and selection in alleviating the premature convergence.

Automated detection of DNA organization in cryoEM micrographs of bacteriophage P4

Wai Ho CHAK*, Tamara Christiani*, Camila Sanudo Thomas, Javier Arsuaga, Naoki Saito, and Mariel Vazquez

The organization of the genome inside viral capsid is critical to the production of infective bacteriophages. While many models have been proposed to describe the global arrangement of the encapsulated viral DNA, these models lack the necessary information obtained from observing viruses directly. Cryogenic election microscopy (cryoEM) micrographs have become the dominant file type in modern structural biology studies. As the level of detail captured in these cryoEM datasets approaches near atomic resolution, there a vital need for advanced image processing techniques to handle these micrographs. Here we present methods developed to detect and analyze the DNA packaged in bacteriophages captured in cryoEM micrographs. Our methods denoise the micrographs and classify the capsids according to their DNA density. The output of this image processing pipeline is an image of averaged packaged capsids that is then used for a radial profile analysis. The profile analysis on fully packaged capsids revealed a series of ordered DNA layers surrounding a disordered core.