samsi
NSF·Duke·NCSU·UNC

Interdisciplinary Workshop for Undergraduate Students
May 14-19, 2017

SPEAKER and PROJECT TITLES/ABSTRACTS

SPEAKERS

Peter J. Mucha
University of North Carolina, Chapel Hill

Bio:
Peter grew up in Minnesota and moved east to attend college at Cornell University where he majored in Engineering Physics. After taking a Churchill Scholarship to study in the Cavendish Laboratory at Cambridge with an M.Phil. in Physics, he returned to the States to continue his studies at Princeton with an M.A. and Ph.D. in Applied and Computational Mathematics. Following a postdoctoral instructorship in applied mathematics at MIT, and a tenure-track assistant professorship in Mathematics at Georgia Tech, he moved to Chapel Hill to join the Department of Mathematics and the Institute for Advanced Materials (now folded into the new Department of Applied Physical Sciences) at UNC.

Title: "Communities in Networks"

Abstract: Networks are all around us, from online social networks to business relationships to family and friends. The patterns in these connections control the way that information, ideas, and diseases spread across a population. In many cases, these processes are strongly influenced by the large-scale organization of nodes into groups. As such, the algorithmic detection of tightly-connected groups of nodes, known as communities, has become a prominent method for studying various networks across different disciplines. Examples discussed in this talk include online social networks, political data, and features of pathogenic E.coli. No previous knowledge about networks will be assumed.

Michael Rappa
Institute for Advanced Analytics – N.C. State University

Bio:
Michael is the founding director of the Institute for Advanced Analytics and a member of the faculty in the Department of Computer Science at North Carolina State University. As head of the Institute, he leads the nation's first Master of Science in Analytics as its originator and principal architect. Before joining NC State as Distinguished University Professor in 1998, for nine years he was a professor at the Massachusetts Institute of Technology. Appointed the inaugural Goodnight Director in 2015, his current endowed position is named in honor of NC State's distinguished alumnus and prominent business leader Dr. James Goodnight.

NOTE: Michael Rappa will speak about opportunities in the analytics field of study

---

PANEL MEMBERS

Rob Chew
RTI International Research Data Scientist
Rob is a versatile data scientist with more than three years of research consulting experience. He uses his training in information retrieval, data management, predictive analytics, text mining, applied econometrics, big data methodologies, and data visualization to help solve complex data problems in public health, social science, and environmental applications.

Prior to joining RTI, Mr. Chew worked as an environmental economist at Abt Associates specializing in water, energy, and climate change. In addition to conducting cost-benefit analysis for several national-scale Environmental Protection Agency (EPA) regulations, he used his data modeling skills to develop a scenario model for forecasting greenhouse gas emissions for the Massachusetts solid waste sector and contributed to economy-wide input-output models for the U.S. Department of Energy (DOE).
B.A. Oberlin College, NCSU M.S. Class of 2014

Melissa Nyeswander

Ph.D., Director, Data Science, Fidelity Investments

Melissa is Director of Data Science at Fidelity Investments, where she's worked across the firm on a number of different subject areas, including customer analytics, fraud, and has recently began to focus on Workforce Analytics. In a previous life, she was a real scientist, with a PhD in Astrophysics from UNC-Chapel Hill, and a career in both astronomy and environmental science.

Diego Malaver

M.S.A. Student, N.C. State Institute for Advanced Analytics

Diego was committed to finding the answers to complex problems at an early age. Diego developed his curiosity for understanding exceptional individuals because of experiences with his younger brother, Alvaro, who has autism and Waardenburg syndrome. As a student at UNC-Chapel Hill, Diego joined the Autism Society, which gave him the ability to assess the progress of children with autism. He also used that opportunity to study the genetics of the disease at UNC's dynamic laboratories. Throughout his academic career and professional experiences, he has been known for the gift of being able to interpret and communicate findings in ways that change minds and impact behaviors. Diego graduated from UNC-Chapel Hill with a B.S. in Biology and Economics with a minor in Chemistry in 2015. He is also pursuing an M.A. at the Institute for Advanced Analytics.

*****NCSU M.S.A. Class of 2017, joining Wake Forest Baptist Medical Center as Data Scientist***

Kaili Stevens

M.S.A. Student, N.C. State Institute for Advanced Analytics

Kaili has always had a passion for learning. She knew from a very young age that she would one day pursue a degree working in mathematics. While studying as an undergraduate student at N.C. State, she realized she enjoyed the challenge of analyzing data and utilizing statistical tools to solve problems. After receiving her B.S., *summa cum laude* in 2016, Kaili received an internship at the US Environmental Protection Agency (EPA), where she developed a model to predict vulnerabilities to dust storms in Arizona under a changing climate scenario. It was during this internship that she realized she was quick to learn new coding languages and adapt to new situations. She enjoys working with large data sets and using analytical techniques to solve complex problems. Kaili is currently pursuing her M.A. at the Institute for Advanced Analytics.

*****NCSU M.S.A. Class of 2017, joining Bain & Company as an Advanced Analytics Specialist***

PROJECTS

Ahmed Attia
SAMSI Postdoctoral Fellow

Bio:
Ahmed received his B.S. degree in Mathematics, Statistics and Computer Science in 2004, and his M.S degree in Statistics and Computer Science in 2008, from Mansoura University, Egypt. In 2016, he obtained his Ph.D. degree, in Computer Science and Applications from Virginia Polytechnic Institute and State University (Virginia Tech), USA. He worked as an intern, in 2014 and in 2015, at Argonne National Laboratory in the Mathematics and Computer Science department. In 2016, he started working as a postdoctoral research scholar at SAMSI, the Statistical and Applied Mathematical Science Institute (RTP, NC). He is affiliated with the Mathematics department at North Carolina State University. Ahmed's research interests are in the area of computational science and engineering.

Project Title: "Data Assimilation for Numerical Weather Prediction"

Abstract: Modern fields of science and engineering rely on sophisticated computer models to describe, predict, and understand the behavioral patterns of complex physical phenomena such as the weather, ocean dynamics, oil reservoirs, earthquakes, and volcanoes. Predictions made by a computer model form an essential source of information about the unknown true state of the physical system of concern. These predictions encapsulate the knowledge about the underlying physical phenomena "prior" to incorporating additional sources of information. In addition to the prior information characterized by the model prediction, complex sensor networks are also employed to collect measurements of the physical phenomena, producing sparse snapshots of reality. Data assimilation (DA) is the inverse problem that fuses information from priors, computer model results, and measurements of reality, in order to provide a consistent description of the true state of the system of concern. In this project, we will learn the basics of numerical weather prediction, and will (learn how to) use a Python-based highly extensible data assimilation testing suite, named DATeS, to develop a practical algorithm for weather prediction. A simple quasi-geostrophic model (already implemented in DATeS to simulate sea surface height of the ocean) will be used as a test case for our project.

Peter Diao
SAMSI Postdoctoral Fellow

Bio:
Peter is currently a postdoctoral fellow at SAMSI in the optimization program. He received his BA in mathematics from Princeton University in 2010. He received his Ph.D. in mathematics from Stanford University in 2016. His thesis was titled "Differential Calculus on Graphon Space and Statistical Applications of Dense Graph Limits." He is interested in mathematical foundations for the analysis of complex systems, representation learning, and optimization.

Project Title: "Distributionally Robust Stochastic Programming for Financial Applications"

Abstract: If we knew the future prices of stocks, then it would be easier to pick a portfolio of investments optimally. Since we do not know, we have to optimize our portfolio according to a probability distribution characterizing our uncertainty about the future. The goal of this project is to introduce these ideas and apply them to real stock data. We will also discuss about how to make our portfolios robust to misspecification of our probability distribution.

David Jones
SAMSI Postdoctoral Fellow

Bio:
David Jones is currently a postdoctoral fellow at SAMSI and is involved in the ASTRO program. He received his Master's Degree in Math in Mathematics from the University of Nottingham (UK) in 2009 and earned his Ph.D. in Statistics from Harvard University in 2016. He is interested in astrostatistics, Bayesian methods, and statistical information.

Project Title: "Lightcurve classification for periodically varying stars"

Abstract: Periodically varying stars constitute many of the most important and intriguing phenomena in astronomy e.g. pulsating stars such as RR Lyrae and systems with two stars orbiting each other (eclipsing binaries). The goal of this project is to classify periodic stars based on features computed from stellar lightcurves (stellar brightnesses observed at many time points). A real dataset with basic features such as estimated periods for the stellar brightness oscillations will be provided and the group will work together to construct a classification algorithm and design any additional features needed in order to get the most accurate classification possible. At the end of the week the true classes of a test dataset will be revealed and the performance of the method will be assessed.

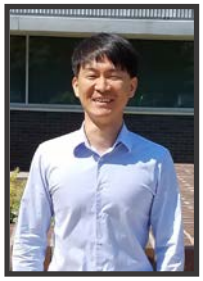
David Stenning
SAMSI Postdoctoral Fellow

Bio:
David Stenning is currently a postdoctoral fellow at SAMSI in the ASTRO program. He received a BSc in Applied Mathematics from Columbia University in 2007 and a PhD in Statistics from the University of California, Irvine in 2015. Between undergraduate studies and graduate school David spent two years

Project Title: "Finding Exoplanets Using Radial Velocity Data"

Abstract: The detection and characterization of exoplanets—planets that orbit stars other than the Sun—is one of the most active areas of research in modern astronomy. Many exoplanets are discovered using the radial velocity method, which involves detecting the Doppler shift in a star's spectral lines resulting from the gravitational effects of an orbiting planet. The goal of this project is to "discover" exoplanets in historical datasets by correctly determining the orbital periods of planets using noisy radial velocity measurements. The radial velocity data is sparse, unevenly sampled, and subject to (known) heteroscedastic measurement errors; determining periodicities is thus statistically challenging.

Hyungsuk Tak
SAMSI Postdoctoral Fellow



Bio:
Tak is currently a postdoctoral fellow at SAMSI in the ASTRO program. He received his BA in Statistics from Korea University in 2009. He received his AM and PhD in Statistics from Harvard University in 2012 and 2016, respectively. He is interested in Bayesian-frequentist unification and astronomical data analysis.

Project Title: "Time Delay Estimation for Gravitationally Lensed Light Curves"

Abstract: The gravitational field of a galaxy can act as a lens and deflect the light emitted by a more distant object such as a quasar. Strong gravitational lensing causes multiple images of the same quasar to appear in the sky. Since the light in each gravitationally lensed image traverses a different path length from the quasar to the Earth, fluctuations in the source brightness are observed in the several images at different times. The time delay between these fluctuations can be used to constrain cosmological parameters, e.g., Hubble constant (the current expansion rate of the Universe), and can be inferred from the time series of brightness data or lightlight curves of each image. We will overview and discuss pros and cons of the existent time delay estimation methods and then each student (or team) is supposed to choose one of the methods or develop a new idea to estimate the blinded time delays and their estimation uncertainties from 50 simulated data sets. Support on coding and test data sets with known time delays (for calibration) will be given. We will discuss the performance of each method to understand why some methods work well and why others do not work well. This discussion will be summarized and reported on the final day.

Sercan Yildiz
SAMSI Postdoctoral Fellow

Bio:
Sercan is currently a postdoctoral fellow at SAMSI in the optimization program. He is also affiliated with the Department of Statistics and Operations Research at the University of North Carolina, Chapel Hill. He received his PhD in Algorithms, Combinatorics, and Optimization from Carnegie Mellon University in 2016. Previously, he received MSc degrees in Algorithms, Combinatorics, and Optimization from Carnegie Mellon University and in Industrial Engineering from the University of Pittsburgh, and a BSc degree in Industrial Engineering from Bogazici University, Turkey. He is interested in mathematical optimization and its applications in statistical learning.

Project Title: "Automatic Genre Classification of Music Pieces"

Abstract: Music information retrieval is an interdisciplinary area that uses tools from signal processing and pattern recognition to extract information from music samples. An important task in music information retrieval is genre classification. In this project, the participants will use several machine learning techniques to develop music genre classifiers based on audio and textual features, and evaluate the performance of these classifiers real data.