FRONTIERS IN COMPUTING + MATHEMATICAL SCIENCES

JANUARY 30 AND FEBRUARY 27 2017

Caltech

FRONTIERS IN COMPUTING + MATHEMATICAL SCIENCES 2017

ALL TALKS ARE ONE HOUR LONG, AND WILL BE HELD IN ANNENBERG 105.
COFFEE BREAKS WILL BE HELD IN THE ANNENBERG LOBBY.

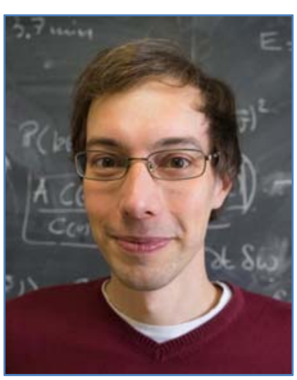
JANUARY 30

9:20am	**Introduction**	**Mathieu Desbrun** Caltech
9:30am	**Data Privacy: How to Survive the Inference Avalanche**	**Reza Shokri** Cornell University
10:30am	Coffee Break	
11:00am	**Data-Driven Probabilistic Modeling and High-Performance Computing: Algorithms and Applications to Physical and Biological Systems**	**Paris Perdikaris** Massachusetts Institute of Technology
1:15pm	**Extreme Events and Metastability in Fluids and Waves**	**Tobias Grafke** New York University
2:15pm	Coffee Break	
2:45pm	**Towards a Theory of Fairness in Machine Learning**	**Jamie Morgenstern** University of Pennsylvania
3:45pm	**Coffee Break**	
4:15pm	**Shared-Memory Parallelism Can Be Simple, Fast, and Scalable**	**Julian Shun** UC Berkeley

Extreme Events and Metastability in Fluids and Waves

Tobias Grafke
New York University

Rare but extreme events often have a dramatic influence on the statistics of stochastic systems, but are notoriously hard to handle both analytically and numerically. In particular in fluid dynamics with its overwhelmingly large number of coupled degrees of freedom, the stochastic forcing interacts with nonlocal nonlinear terms to create coherent structures inducing strong non-Gaussianity. I will present how large deviation theory provides a rigorous framework to quantify these effects, allowing to predict the emergence of extreme events, and similarly transition events in metastable fluid systems (as observed in e.g. atmospheric flows), as well as computing their probability and statistics. The same methods apply to stochastic systems from other fields, such as ocean surface waves, active matter, or reaction/diffusion systems.



Biography:
Tobias Grafke *earned his PhD at the Ruhr-University Bochum, Germany, on the topic of existence and regularity of solutions to the three-dimensional incompressible Euler equations. From 2013 to 2015, he was a was a Dean of Faculty fellow at the Weizmann Institute of Science, Israel, working on turbulence fundamentals. Since 2015, he is an Instructor at the Courant Institute in New York, where he studies the application of large deviation theory to extreme events and metastability in fluid dynamics.*

Towards a Theory of Fairness in Machine Learning

Jamie Morgenstern
University of Pennsylvania

Algorithm design has moved from being a tool used exclusively for designing systems to one used to present people with personalized content, advertisements, and other economic opportunities. Massive amounts of information is recorded about people's online behavior including the websites they visit, the advertisements they click on, their search history, and their IP address. Algorithms then use this information for many purposes: to choose which prices to quote individuals for airline tickets, which advertisements to show them, and even which news stories to promote. These systems create new challenges for algorithm design. When a person's behavior influences the prices they may face in the future, they may have a strong incentive to modify their behavior to improve their long-term utility; therefore, these algorithms' performance should be resilient to strategic manipulation. Furthermore, when an algorithm makes choices that affect people's everyday lives, the effects of these choices raise ethical concerns such as whether the algorithm's behavior violates individuals' privacy or whether the algorithm treats people fairly.

Machine learning algorithms in particular have received much attention for exhibiting bias, or unfairness, in a large number of contexts. In this talk, I will describe my recent work on developing a definition of fairness for machine learning. One definition of fairness, encoding the notion of 'fair equality of opportunity', informally, states that if one person has higher expected quality than another person, the higher quality person should be given at least as much opportunity as the lower quality person. I will present a result characterizing the performance degradation of algorithms which satisfy this condition in the contextual bandits setting. To complement these theoretical results, I then present the results of several empirical evaluations of fair algorithms.

I will also briefly describe my work on designing algorithms whose performance guarantees are resilient to strategic manipulation of their inputs, and machine learning for optimal auction design.

Biography:

Jamie Morgenstern *is a Warren Center postdoctoral fellow in Computer Science and Economics at the University of Pennsylvania. She received her Ph.D. in Computer Science from Carnegie Mellon University in 2015, and her B.S. in Computer Science and B.A. in Mathematics from the University of Chicago in 2010. Her research focuses on machine learning for mechanism design, fairness in machine learning, and algorithmic game theory. She received a Microsoft Women's Research Scholarship, an NSF Graduate Research Fellowship, and a Simons Award for Graduate Students in Theoretical Computer Science.*

Data-Driven Probabilistic Modeling and High-Performance Computing: Algorithms and Applications to Physical and Biological Systems

Paris Perdikaris
Massachusetts Institute of Technology

The analysis of complex physical and biological systems necessitates the accurate resolution of interactions across multiple spatio-temporal scales, the consistent propagation of information between concurrently coupled multi-physics processes, and the effective quantification of model error and parametric uncertainty. Addressing these grand challenges is a multi-faceted problem that poses the need for a highly sophisticated arsenal of tools in stochastic modeling, high-performance scientific computing, and probabilistic machine learning. Through the lens of three realistic large-scale applications, this talk aims to demonstrate how the compositional synthesis of such tools is introducing a new paradigm in scientific discovery. First, we present multi-scale blood flow simulations in the human brain, and show how high-order methods, massively parallel computing, and concurrent coupling of multi-physics solvers can uncover intrinsic physiological mechanisms in health and disease. We will demonstrate how the introduction of probabilistic machine learning techniques, and the key concept of multi-fidelity modeling, provide a scalable platform for information fusion and lead to significant computational expediency gains. The second application involves an environmental study that illustrates how machine learning tools enable the synergistic combination of simulations, noisy measurements and empirical models towards quantifying the anthropogenic effect in the increasing acidification of coastal waters, and developing a cost-effective monitoring and prediction mechanism. Lastly, we consider the shape optimization of super-cavitating hydrofoils of an ultrafast marine vessel for special naval operations. Specifically, we show how the combination of turbulent multi-phase flow simulations and the concept of multi-fidelity Bayesian optimization allows us to tackle complex engineering design problems in which a rigorous assessment of uncertainty and risk becomes critical in policy and decision making.

Biography:

Paris Perdikaris *received his PhD in Applied Mathematics from Brown University in May 2015. His expertise lies in probabilistic machine learning, computational fluid dynamics, multi-fidelity modeling, uncertainty quantification, and parallel scientific computing. While at Brown he developed scalable machine learning algorithms for predictive multi-fidelity modeling of high-dimensional systems. A parallel research thrust involved developing mathematical models for simulating cardiovascular fluid flows and assessing the characteristics of cerebral pathologies such as cerebral aneurysms in-silico. In June 2015, he moved to MIT as a post-doctoral research associate at the department of Mechanical Engineering and the MIT Sea Grant College Program. His research at MIT is focused on designing a scalable data-driven framework for uncertainty quantification, inverse problems, design optimization, and beyond. The developed algorithms are currently used for risk-averse design optimization of super-cavitating hydrofoils, as well as data assimilation of noisy measurements in coastal regions. Moreover, he has co-advised an undergraduate student working on active learning and data acquisition under uncertainty, and a masters student working on deep learning techniques for object recognition, tracking, and autonomous marine navigation. From 2010-present he has been actively involved in several research projects funded by major US agencies including DOE, AFOSR, NIH and DARPA.*

Data Privacy: How to Survive the Inference Avalanche

Reza Shokri
Cornell University

Underestimating the power of inference attacks is the major reason why data privacy mechanisms fail. In this talk, I will describe my general approach to quantifying privacy and illustrate its applications by showing how to rigorously measure privacy risks of location data and machine-learning models. I will then discuss my current research at the junction of privacy and data science in two important practical scenarios: generating privacy-preserving synthetic data and building accurate deep-learning models that respect privacy of the training data.

Biography:
Reza Shokri *is a postdoctoral researcher at Cornell Tech. His research focuses on quantitative analysis of privacy, as well as design and implementation of privacy technologies for practical applications. His work on quantifying location privacy was recognized as a runner-up for the Award for Outstanding Research in Privacy Enhancing Technologies (PET Award). Recently, he has focused on privacy-preserving generative models and privacy in machine learning. He received his PhD from EPFL.*

Shared-Memory Parallelism Can Be Simple, Fast, and Scalable

Julian Shun
UC Berkeley

Parallelism is the key to achieving high performance in computing. However, writing efficient and scalable parallel programs is notoriously difficult, and often requires significant expertise. To address this challenge, it is crucial to provide programmers with high-level tools to enable them to develop solutions more easily, and at the same time emphasize the theoretical and practical aspects of algorithm design to allow the solutions developed to run efficiently under many possible settings. My research addresses this challenge using a three-pronged approach consisting of the design of shared-memory programming techniques, frameworks, and algorithms for important problems in computing. In this talk, I will present tools for deterministic parallel programming, large-scale shared-memory algorithms that are efficient both in theory and in practice, and Ligra, a framework for simplifying the programming of shared-memory graph algorithms.

Biography:
Julian Shun *is currently a Miller Research Fellow (post-doc) at UC Berkeley. He obtained his Ph.D. in Computer Science from Carnegie Mellon University, and his undergraduate degree in Computer Science from UC Berkeley. He is interested in developing large-scale parallel algorithms for graph processing, and parallel text algorithms and data structures. He is also interested in designing methods for writing deterministic parallel programs and benchmarking parallel programs. He has received the ACM Doctoral Dissertation Award, CMU School of Computer Science Doctoral Dissertation Award, Miller Research Fellowship, Facebook Graduate Fellowship, and a best student paper award at the Data Compression Conference.*