Stochastic Networks 2022

Cornell University, June 20-24th, 2022

Location: All talks are held in Philips Hall room #101. Breakfast, and lunches are held in the Duffield Atrium. Duffield and Philips are two adjacent (and connected) buildings. See google map here. Also see here for a map of the campus.

Tentative Conference Schedule

Monday June 20

8:30-9:00	Breakfast and Registration
9:00 - 9:30	Welcome and Information
9:30 - 10:30	Kuang Xu
10:30 - 11:00	Coffee Break
11:00 - 12:00	Rouba Ibrahim
12:00 - 1:30	Lunch Break
1:30 - 2:30	Henry Lam
2:30 - 3:30	"Lightning Round", Part 1
3:30 - 4:30	Poster session, Part 1
4:30 - 5:00	"Lightning Round", Part 2
5:15 - 8:00	Poster Session, Part 2 + Reception @ Johnson Art Museum

Session chairs: Bert Zwart (AM), John Hasenbein (PM)

Note*: Todd Colemans's talk (scheduled originally for 3:00) is canceled due to a personal emergency

Tuesday June 21

8:30 - 9:00	Breakfast

9:00 - 10:00	Kavita Ramanan
10:00 - 10:30	Coffee Break
10:30 - 11:30	Nicolas Gast
11:30 - 1:30*	Lunch Break
1:30 - 2:30	Carri Chan
2:30 - 3:30	Varun Gupta
3:30-4:00	Coffee Break
4:00 - 5:30	Interactive Hybrid Forum — Stochastic Networks: Past, Present, and Future
Dinner on your own	
*Group photo will be taken during lunch break (11:30-11:45)	

Session chairs: Amber Puha (AM), Marty Reiman (PM)

Wednesday June 22

8:30-9:00	Breakfast
9:00 - 10:00	Sophie Hautphenne
10:00 - 10:30	Coffee Break
10:30 - 11:30	Daniele Cappelletti
11:30 - 12:30	Nelly Litvak
12:30 - 2:00	Lunch Break
2:00 - 3:00	Neil Walton
3:00 - 3:30	Coffee Break
3:30 - 4:30	Alexandre Proutiere
4:30 - 5:30	Jon Kleinberg
6:00pm : Banquet @ Statler Terrace (see map)	

Session chairs: Ton Dieker (AM), Rami Atar (PM)

Thursday June 23

8:30-9:00	Breakfast
9:00 - 10:00	Weina Wang
10:00 - 10:30	Coffee Break
10:30 - 11:30	Sergey Foss
11:30 - 12:30	Onno Boxma
12:30	Lunch
1:00	Excursion / Free Afternoon

Session chair: Shane Henderson

Friday June 24

8:30-9:00	Breakfast
9:00 - 10:00	Mengdi Wang
10:00 - 10:30	Coffee Break
10:30 - 11:30	Christina Yu
11:30 - 12:30	R. Srikant
Conference ends	

Session chair: Vikram Krishnamurthy

Onno Boxma

Title: From the single server queue to reflected autoregressive processes

Abstract: We discuss a reflected AR(1) process $(W_n)_n$ obeying a recursion that generalizes the well-known Lindley recursion. Using a Wiener-Hopf decomposition, we obtain explicit results for transforms of the distribution of W_n and of its limiting distribution. Note: This concerns joint studies with Andreas L\"opker, Michel Mandjes, Zbigniew Palmowski and Josh Reed.

Bio: Onno Boxma is emeritus professor of Stochastic Operations Research in the Department of Mathematics and Computer Science of Eindhoven University of Technology. His main research interests are in queueing and insurance risk. He has been editor-in-chief of Queueing Systems (2004-2009) and scientific director of Eurandom (2005-2011). Onno Boxma has received honorary doctorates from the University of Haifa and Heriot-Watt University (Edinburgh), and was recipient of the 2011 ACM SIGMETRICS Achievement Award and the 2014 Arne Jensen Lifetime Award of ITC.

Daniele Cappalletti

Title: Solving the chemical recurrence conjecture in 2D

Abstract: Stochastic reaction networks are continuous-time Markov chain models typically used in biology, epidemiology, and population dynamics. What makes them special from a mathematical point of view is the fact that their qualitative dynamics is described by the finite directed graph of allowed reactions (typically biomolecular

transformations but not necessarily so), referred to as "reaction graph". A long standing conjecture is that models with a reaction graph composed by a union of strongly connected components are necessarily positive recurrent. In my talk I will discuss why the conjecture makes intuitive sense and why it is difficult to prove it. I will then show how my collaborators and I adapted Forster-Lyapunov techniques to prove the conjecture in two dimensions, in the hope that the method we developed can be used in other contexts as well.

Bio: I started studying stochastic reaction networks during my PhD in Copenhagen, under the guide of Carsten Wiuf. I continued my mathematical investigations of these models at the University of Wisconsin-Madison, working with David Anderson and Thomas Kurtz. I could then observe the concrete applications of the theoretical results in the field at the Control Theory and Systems Biology Laboratory of the ETH Zurich, led by Mustafa Khammash. Here I worked in close contact with biologists, engineers, and computational scientists. I now work at the Department of Mathematical Sciences of the Politecnico di Torino, in Italy.

Carri Chan

Title: Managing Queues with Different Resource Requirements

Abstract: Queueing models that are used to capture various service settings typically assume that customers require a single unit of resources (servers) to be processed. However, there are many service settings where such an assumption may fail to capture the heterogeneity in resource requirements of different customers. We propose a multi-server queueing model with multiple customer classes in which customers from different classes may require different amounts of resources to be served. We study the optimal scheduling policy for such systems.

To balance the holding cost, the service rate, the resource requirement, and the priority-induced idleness, we develop a class of index-based policies which we refer to as the idle-aware c-mu/m rule. We establish the asymptotic optimality of this class of policies in the many-server heavy-traffic regime. For a two-class two-server model, where policy-induced idleness can have a big impact on system performance, we establish a uniform performance bound on the amount of sub-optimality incurred by the idle-avoid c-mu/m rule (a special case of the idle-aware c-mu/m rule). This theoretical bound, along with numerical experiments, provides support for the robustness of our proposed class of policies.

Based on joint work with Jing Dong and Noa Zychinski

Bio: Carri W. Chan is a Professor of Business in the Decision, Risk and Operations Division and the Faculty Director of the Healthcare and Pharmaceutical Management Program at Columbia Business School. Her research is in the area of healthcare operations management. Her primary focus is in data-driven modeling of complex stochastic systems, efficient algorithmic design for queuing systems, dynamic control of stochastic processing systems, and econometric analysis of healthcare systems. Her research combines empirical and stochastic modeling to develop evidence-based approaches to improve patient flow through hospitals. She has worked with clinicians and administrators in numerous hospital systems including Northern California Kaiser Permanente, New York Presbyterian, and Montefiore Medical Center. She is the recipient of a 2014 National Science Foundation (NSF) Faculty Early Career Development Program (CAREER) award, the 2016 Production and Operations Management Society (POMS) Wickham Skinner Early Career Award, and the 2019 MSOM Young Scholar Prize. She currently serves as a co-Department Editor for the Healthcare Management Department at Management Science. She received her BS in Electrical Engineering from MIT and MS and PhD in Electrical Engineering from Stanford University.

Todd Coleman

Title: Kuramoto von Mises Autoregressive Statistical Models

Abstract: The Kuramoto model is a widely used models of interacting elements with nonlinear dynamics, specifically within the context of coupled oscillators. It has been utilized in numerous physical, biological, chemical, and societal systems, where there does not exist a central coordinating mechanism and each interacting element can be represented. One example in the biological domain of coupled oscillators is the network of Cajal cells in the digestive tract: each cell individually has a membrane potential that oscillates at 0.05Hz. Cajal cells near each other are coupled and this coupling gives rise to slow waves that mediate sequencing of smooth muscle contractions that propel food down the digestive tract. Here we consider a stochastic differential formulation of the problem and consider estimating the coefficients of coupling from measurements of oscillatory data. We show that at sufficiently high sampling rates, the estimation problem is quadratic and can be solved in closed form. We then show how using group sparsity estimation techniques can allow for estimating time varying sub-network dynamics, and how using high dimensional sparse models can allow for estimating generalizations of the Kuramoto formulation. We provide an example of this inference framework within the context of Cajal cells in the digestive system.

Bio: Todd P. Coleman received B.S. degrees in electrical engineering (summa cum laude), as well as computer engineering (summa cum laude) from the University of Michigan. He received M.S. and Ph.D. degrees from MIT in electrical engineering and did postdoctoral studies at MIT in neuroscience.
Currently on the faculty in the Department of Bioengineering at Stanford University, Dr. Coleman previously was a faculty member in the department of ECE at University of Illinois, Urbana-Champaign, and the department of Bioengineering at UC San Diego. Dr.Coleman's research is very multi-disciplinary, using tools from applied probability, physiology, and bioelectronics. His research spans from developing fundamental information theory and machine learning techniques to developing technologies to

monitor and modulate physiology of the nervous systems in the brain and visceral organs. He has been selected as a National Academy of Engineering Gilbreth Lecturer and is an IEEE Fellow as well as a Fellow of the American Institute for Medical andBiologicalEngineering. He is currently the Chair of the National Academies Standing Committee on Biotechnology Capabilities and National Security Needs.

Sergey Foss

Title: Migration-Contagion Processes

Abstract: S. Foss (jointly with F. Baccelli and S. Shneer)

We consider a migration process based on a closed network of N queues with K = K(N) customers. Each station is an infinite-server queue with i.i.d. exponential service times. The routing is independent and uniform at random to all stations. These customers follow the SIS (Susceptible-Infection-Susceptible) dynamics. That is, customers are in one of two states: I for infected, or S for susceptible. Customers can swap their state only in stations, either from I to S or from S to I. More precisely, at any station, each susceptible customer becomes infected with the instantaneous rate aY if there are Y infected customers in the station, whereas each infected customer recovers and becomes susceptible with rate b (here a and b are positive parameters).

We let N tend to infinity and assume that K(N)/N tends to a finite positive constant that represents the customer density. The main problem of interest is about the set of parameters of such a system for which there exists a stationary regime where the epidemic survives in the limiting system. The latter limit will be referred to as the thermodynamic limit.

We establish several structural properties (monotonicity and convexity) of the system which allow us to give the structure of the phase diagram of this thermodynamic limit. The analysis of this SIS model reduces to that of a wave type PDE which can in turn be reduced to an infinite-dimensional ODE. This plain SIS model is one among a large class of companion stochastic processes that exhibit both migration and contagion, and we present two auxiliary models from the class that provide bounds to the plain SIS model.

We start the talk with discussing an open system consisting of a single station with Poisson input that is instrumental in our analysis of the thermodynamic limits and is also of independent interest. Then we turn to closed systems and their limits.

Bio: Sergey Foss, Professor in Applied Probability at the Maxwell Institute and Heriot-Watt University, Edinburgh; PhD and DSc from S.L.Sobolev Institute of Mathematics; Principal Organiser of a 6-month research semester at the Isaac Newton Institute (2010), EiC of QUESTA (2009--2021), Area Editor for SPA (2018-pres), Fellow of the Royal Society of Edinburgh (2007). Main research interests: stability, limit theorems and rare events in (applied) stochastic processes, coupling methods, directed random graphs.

Nicolas Gast

Title: Computing the bias of mean field approximation

Abstract: Mean field approximation is a widely used technique to study stochastic systems composed of many interacting objects with applications from theoretical physics to biological models and artificial intelligence. In computer science, mean field approximation has been successfully used to analyze the performance of many distributed algorithms, including allocation strategies in server farms, caching algorithms and wireless protocols. In a first part, I will introduce the key concepts behind the classical mean field approximation. By using some of the classical models, I will show how the method can be applied and give ideas of where it cannot be applied. In the second part, I will talk about more recent results: first, how accurate is the approximation for finite systems, and second, how to use this result to define a refined approximation to make it applicable for a few tens of objects.

Bio: Nicolas Gast is a tenured research scientist at Inria (Grenoble, France) since 2014. He graduated from Ecole Normale Superieure (Paris, France) in 2007 and received a Ph.D. from the University of Grenoble in 2010. He was a research fellow at EPFL from 2010 to 2014. His research focuses on the development and the use of stochastic models and optimization methods for the design of distributed control algorithms in large-scale systems, with applications to communication networks, computing infrastructures and more recently energy networks (smart grids).

Varun Gupta

Title: Menu Design of a Bipartite Matching Queueing System with Strategic Users

Abstract: We consider a menu design problem in a multi-class multi-server queueing system. The central planner must decide on a collection of service classes to offer, where each service class is defined by a subset or bundle of servers that can serve it. Customers are heterogeneous in their preferences over servers and self-select the service class to join. The servers process customers in compatible service classes using the FCFS-ALIS service discipline. We study the notion of equilibrium for such a system in heavy-traffic, which turns out to be somewhat subtle, and identify the trade-off of delay vs. matching value obtainable for many special scenarios.

Bio: Varun Gupta is an Associate Professor of Operations Management at the Booth School of Business at the University of Chicago. He received his PhD in Computer Science from Carnegie Mellon University in Pittsburgh, PA and holds a Bachelor of

Technology in Computer Science and Engineering from the Indian Institute of Technology in New Delhi, India. His research interests include stochastic processes and applied probability, algorithm design and analysis, learning and control in non-stationary environments, queueing and scheduling theory, and mechanism design. He is particularly interested in applications of these tools to cloud computing infrastructure, healthcare operations, two-sided markets for freight and other 3PL, and influence propagation in social networks

Sophie Hautphenne

Title: The probabilities of extinction in a branching random walk

Abstract: We consider multitype Galton-Watson branching processes with countably infinite type sets. In particular, we study the probability of extinction in (finite or infinite) subsets of types A, that is, the probability that there exists a generation after which we never see an individual in the set A. We derive conditions under which two distinct subsets of types lead to different extinction probabilities; we discuss how many distinct extinction probabilities may exist; and finally, we make progress towards locating the extinction probabilities in the set of solutions to the fixed-point equation that characterises the branching process.

Bio: Sophie is a Senior Lecturer in the School of Mathematics and Statistics at the University of Melbourne since 2017. Between 2015 and 2019 she was a research fellow on an ARC Discovery Early Career Researcher Award (DECRA), and during that period she was also holding a part-time position as a Scientist in the Chair of Statistics, Swiss Federal Institute of Technology in Lausanne (EPFL, until October 2018). Between 2015 and 2021 she was an Associate Investigator in the ARC Centre of Excellence for Mathematical & Statistical Frontiers (ACEMS). She obtained a PhD in Mathematics from the Université libre de Bruxelles in October 2009. Her fields of research are applied probability and stochastic modelling with a particular focus on branching processes, matrix analytic methods, epidemic models and queueing models. She is particularly interested in applications in population biology and ecology.

Rouba Ibrahim

Title: Size-based scheduling in service systems

Abstract: Size-based scheduling policies, such as shortest-remaining-processing-time (SRPT) and shortest-job-first (SJF), have been extensively studied, for decades, yet almost exclusively in single-server queues with infinitely patient jobs and under exact job-size information. Motivated by applications to service systems, such as call centers or healthcare facilities, we analyze the performance of size-based scheduling in multiserver queues with abandonment and inexact job-size information. In particular, we

demonstrate that system performance under size-based scheduling with noisy service times is asymptotically equivalent, in the many-server heavy-traffic limit, to performance under a two-class priority queue where customers with short service times (below a threshold) are served without wait, and customers with long service times (above a threshold) eventually abandon without service. This is joint work with Jing Dong, from Columbia Business School.

Bio: Rouba is a Professor at the School of Management of University College London. She holds a PhD degree in Operations Research from Columbia University. Her research and teaching interests focus on service operations using both queueing theoretic and data-analytic techniques. She serves on the editorial boards of Management Science, MSOM, IISE Transactions, and Operations Research.

Jon Kleinberg

Title: The Challenge of Understanding What Users Want: Inconsistent Preferences and Engagement Optimization

Abstract: Online platforms have a wealth of data, run countless experiments and use industrial-scale algorithms to optimize user experience. Despite this, many users seem to regret the time they spend on these platforms. One possible explanation is that incentives are misaligned: platforms are not optimizing for user happiness. We suggest the problem runs deeper, transcending the specific incentives of any particular platform, and instead stems from a mistaken foundational assumption. To understand what users want, platforms look at what users do. This is a kind of revealed-preference assumption that is ubiquitous in user models. Yet research has demonstrated, and personal experience affirms, that we often make choices in the moment that are inconsistent with what we actually want: we can choose mindlessly or myopically, behaviors that feel entirely familiar on online platforms.

In this work, we develop a model of media consumption where users have inconsistent preferences. We consider what happens when a platform that simply wants to maximize user utility is only able to observe behavioral data in the form of user engagement. Our framework is based on a stochastic model of user behavior, in which users are guided by two conflicting sets of preferences -- one that operates impulsively in the moment, and the other of which makes plans over longer time-scales. By linking the behavior of this model to abstractions of platform design choices, we can develop a theoretical framework and vocabulary in which to explore interactions between design, behavioral science, and social media.

The talk is based on joint work with Sendhil Mullainathan and Manish Raghavan.

Bio: Jon Kleinberg is the Tisch University Professor in the Departments of Computer Science and Information Science at Cornell University. His research focuses on the interaction of algorithms and networks, the roles they play in large-scale social and information systems, and their broader societal implications. He is a member of the National Academy of Sciences and the National Academy of Engineering, and the recipient of MacArthur, Packard, Simons, Sloan, and Vannevar Bush research fellowships, as well awards including the Harvey Prize, the Nevanlinna Prize, and the ACM Prize in Computing.

Henry Lam

Title: Towards Model-Agnostic Rare-Event Simulation with Guarantees

Abstract: In estimating rare-event probabilities, naive Monte Carlo is inefficient and techniques such as importance sampling have shown to be powerful in speeding up computation. However, to attain statistical guarantees, these techniques typically rely on model structural knowledge that is not always available or amenable to analysis, a challenge that is especially prominent for complex black-box systems arising in recent AI safety applications. We describe some approaches towards obtaining guarantees for importance sampling with lighter requirements on model knowledge. Specifically, we propose and justify relaxations of classical efficiency notions on several fronts, to account for: 1) over-conservativeness of variance-based criteria in enforcing algorithmic properties; 2) lack of precise geometric knowledge on rare-event sets; and 3) lack of a priori knowledge on the solution when solving optimization problems or training models that are rare-event-aware. We present simple procedures to attain these relaxed efficiency notions and discuss their performances.

Bio: Henry Lam is an Associate Professor in the Department of Industrial Engineering and Operations Research at Columbia University. His research interests include Monte Carlo methods, uncertainty quantification, data-driven optimization and rare-event analysis. His works have been recognized by several venues such as the NSF CAREER Award, NSA Young Investigator Award, JP Morgan Chase Faculty Research Award and Adobe Faculty Research Award. Henry serves on the editorial boards of Operations Research, INFORMS Journal on Computing, Applied Probability Journals, Stochastic Models, Manufacturing and Service Operations Management, and Queueing Systems, and as the Area Editor in Stochastic Models and Data Science in Operations Research Letters.

Nelly Litvak

Title: Red Light - Green Light Solution for Large Markov Chains

Abstract: Markov chains are versatile models that provide a mathematical representation for a wide range of real-life stochastic processes, including random walks and other random processes on networks. One of most common tasks in studies

of Markov chains is computation of the stationary distribution. This task is especially difficult when the transition matrix of a Markov chain has a prominent clustered structure, typical for applications in large complex networks. We propose a completely new controlled distributed algorithm for computing stationary distribution of large Markov chains, prove its exponential convergence, and demonstrate its high efficiency, both mathematically and experimentally. In particular, we derive control strategies that achieve convergence rate faster than the standard power iterations, and easily handle clustered structure.

Bio: Nelly Litvak obtained her PhD in Stochastic Operations Research in 2002 from Eindhoven university of Technology. She is currently a professor at the University of Twente and Eindhoven University of Technology in the Netherlands. She received the Stieltjes prize for the best PhD thesis and the Google faculty research award. She is an associate editor of Stochastic Processes and their Applications and Queuing Systems. Her research interests are in the study of large networks such as on-line social networks and the World Wide Web, randomized algorithms, and random graphs.

Alexandre Proutiere

Title: Regret in Online Recommendation Systems

Abstract: We propose a theoretical analysis of recommendation systems in an online setting, where items are sequentially recommended to users over time. In each round, a user, randomly picked from a population of m users, requests a recommen- dation. The decision-maker observes the user and selects an item from a catalogue of n items. Importantly, an item cannot be recommended twice to the same user. The probabilities that a user likes each item are unknown. The performance of the recommendation algorithm is captured through its regret, considering as a reference an Oracle algorithm aware of these probabilities. We investigate various structural assumptions on these probabilities: we derive for each structure regret lower bounds, and devise algorithms achieving these limits. Interestingly, our analysis reveals the relative weights of the different components of regret: the component due to the constraint of not presenting the same item twice to the same user, that due to learning the chances users like items, and finally that arising when learning the underlying structure.

Bio: Alexandre Proutiere is professor in the EECS school at KTH, Royal Institute of Technology, Stockholm, Sweden. He was researcher at Microsoft Research (Cambridge) from 2007 to 2011, and a research Engineer at France Telecom R&D from 2000 to 2006. He also was an invited lecturer and researcher at the computer science department ENS Paris from 2004 to 2006. Alexandre received a PhD in Applied Mathematics from Ecole Polytechnique, he graduated in Mathematics from Ecole Normale Superieure (Paris), and has an engineering degree from Telecom Paris. He is an engineer from Corps des Mines. He received the ACM Sigmetrics rising star award in 2009, the ACM best papers awards at Sigmetrics 2004 and 2010, and Mobihoc 2009. He held an ERC consolidator grant from 2012 to 2017.

Kavita Ramanan

Title: Beyond mean-field limits in the study of large-scale networks

Abstract: Large-scale networks capture various models in engineering and operations research, ranging from load balancing in computer networks to routing in call centers, and their dynamics can often be modelled in terms of coupled stochastic processes whose interactions are local with respect to an underlying graph. The dynamics of these networks are complex and not very tractable, and when the interaction graph is the complete graph, so-called mean-field approximations are well known to serve as tractable approximations for these processes. We will describe novel approximations that are valid in the case when the interaction graph is truly sparse, and which can be rigorously shown to be accurate in a suitable asymptotic. We will also demonstrate their efficacy of these approximations and describe potential applications of these results. This is based on joint work with Ankan Ganguly.

Bio: Kavita Ramanan is the Roland George Dwight Richardson University Professor of Applied Mathematics at Brown University, and an Associate Director with the Institute for Computational and Experimental Research in Mathematics (ICERM), having also previously served as Deputy Director. She is currently the Associate Chair of the Division of Applied Mathematics from 2017, and served as Director of Graduate Studies from 2011--2014.

She works in the area of probability theory and stochastic processes, and has made fundamental contributions to several fields, including the theory of large deviations, stochastic analysis, interacting particle systems, high-dimensional probability and the analysis of stochastic networks. She was awarded the Erlang Prize from the Applied Probability Society in 2006 and a Medallion from the Institute of Mathematical Statistics in 2015. She is an elected Fellow of several societies, including the Institute of Mathematical Statistics (IMS), the American Mathematical Society (AMS), the Institute for Operations Research and the Management Sciences (INFORMS), the American Association for the Advancement of Sciences (AAAS) and the Society of Industrial and Applied Mathematics (SIAM). She was awarded the Simons Fellowship in 2018 and a Guggenheim Fellowship in the Natural Sciences in 2020. She was also a recipient of the Newton Award from the Department of Defense in 2020.

R. Srikant

Title: A Policy Gradient Algorithm for the Risk-Sensitive Exponential Cost MDP

Abstract: We consider the risk-sensitive exponential cost MDP formulation and develop a trajectory-based gradient algorithm to find the stationary point of the cost associated with a set of parameterized policies. We present a formula that can be used to compute the policy gradient from (state, action, cost) information collected from sample paths of

the MDP for each fixed parameterized policy. Unlike the traditional average-cost problem, standard stochastic approximation theory cannot be used to exploit this formula. To address the issue, we introduce a truncated and smooth version of the risk-sensitive cost and show that this new cost criterion can be used to approximate the risk-sensitive cost and its gradient uniformly under some mild assumptions. We then develop a trajectory-based gradient algorithm to minimize the smooth truncated estimation of the risk-sensitive cost and derive conditions under which a sequence of truncations can be used to solve the original, untruncated cost problem. Joint work with Mehrdad Moharrami, Yashaswini Murthy and Arghyadip Roy.

Bio: R. Srikant is the Fredric G. and Elizabeth H. Nearing Endowed Professor of Electrical and Computer Engineering and the Coordinated Science Lab and a Co-Director of the C3.ai Digital Transformation Institute at the University of Illinois at Urbana-Champaign. His research interests are in the areas of applied probability, stochastic networks, and control theory, with applications to machine learning, cloud computing, and communication networks. He is the recipient of the 2021 ACM SIGMETRICS Achievement Award, the 2019 IEEE Koji Kobayashi Computers and Communications Award and the 2015 IEEE INFOCOM Achievement Award. He has also received several Best Paper awards, including the 2017 Applied Probability Society Best Publication Award, the 2015 IEEE INFOCOM Best Paper Award and the 2015 WiOpt Best Paper Award. He was the Editor-in-Chief of the IEEE/ACM Transactions on Networking from 2013-2017.

Neil Walton

Title: Regret Analysis of a Markov Policy Gradient Algorithm for Multi-arm Bandits.

Abstract: We consider a policy gradient algorithm applied to a finite-arm bandit problem with Bernoulli rewards. We allow learning rates to depend on the current state of the algorithm, rather than use a deterministic time-decreasing learning rate. The state of the algorithm forms a Markov chain on the probability simplex. We apply Foster-Lyapunov techniques to analyse the stability of this Markov chain. We prove that if learning rates are well chosen then the policy gradient algorithm is a transient Markov chain and the state of the chain converges on the optimal arm with logarithmic or poly-logarithmic regret.

Bio: Neil Walton is a reader at the University of Manchester. His research is in applied probability and principally concerns the decentralized minimization of congestion in networks. He received his undergraduate ('05), Masters ('06) and PhD ('10) in Mathematics at the University of Cambridge. He was a lecturer at University of Amsterdam where he held an NWO Veni Fellowship. He then moved to the University of Manchester where he is a Reader in Mathematics. Neil has conducted research visits at Microsoft Research Cambridge, the Basque Centre for Mathematics and the Automatic Control Laboratory ETH Zurich. From 2017 to 2019 Neil was the head of probability and statistics group at the University of Manchester. Neil is currently a Fellow of the Alan Turing Institute, a guest lecturer at London Business School, and has an honorary position at the Manchester University NHS Foundation Trust. Here he is using queueing theory to analyze the waiting list backlog from the covid pandemic. He is an associate editor at the journal Stochastic Systems. He is an area editor for stochastic models at Operations Research. He has won best papers awards at the ACM Sigmetrics conference and he was awarded the 2018 Erlang Prize by the Informs Applied Probability Society.

Mengdi Wang

Title: Thompson Sampling-Guided Directed Evolution for Sequence Optimization

Abstract: Directed Evolution (DE), a landmark wet-lab method originated in 1960s, enables discovery of novel protein designs via evolving a population of candidate sequences. Recent advances in biotechnology has made it possible to collect high-throughput data, allowing the use of machine learning to map out a protein's sequence-to-function relation. There is a growing interest in machine learning-assisted DE for accelerating protein optimization. Yet the theoretical understanding of DE, as well as the use of machine learning in DE, remains limited.
In this paper, we connect DE with the bandit learning theory and make a first attempt to study regret minimization in DE. We propose a Thompson Sampling-guided Directed Evolution (TS-DE) framework for sequence optimization, where the sequence-to-function mapping is unknown and querying a single value is subject to costly and noisy measurements. TS-DE updates a posterior of the function based on collected measurements. It uses a posterior-sampled function estimate to guide the crossover recombination and mutation steps in DE. In the case of a linear model, we show that TS-DE enjoys a Bayesian regret of order $\tilde O(d^{2}\sqrt{MT})$\footnote{$\tilde O(\cdot)$ ignores the logarithmic terms.}, where d is feature dimension, M is population size and T is number of rounds. This regret bound is nearly optimal, confirming that bandit learning can provably accelerate DE. It may have implications for more general sequence optimization and evolutionary algorithms.

Bio: Mengdi Wang received her Ph.D. in Electrical Engineering and Computer Science from Massachusetts Institute of Technology in 2013. At MIT, Mengdi was affiliated with the Laboratory for Information and Decision Systems and was advised by Dimitri P. Bertsekas. Mengdi joined Princeton University in 2014. She received the Young Researcher Prize in Continuous Optimization of the Mathematical Optimization Society in 2016 (awarded once every three years), the Princeton SEAS Innovation Award in 2016, the NSF Career Award in 2017, the Google Faculty Award in 2017, and the MIT Tech Review 35-Under-35 Innovation Award (China region) in 2018. She is currently serving as an associate editor for Operations Research.

Her research interests on data-driven optimization, reinforcement learning, statistical dimension reduction and system identification. Mengdi's research group studies the statistical and algorithmic foundation of reinforcement learning and sequential decision-making, as well as their applications in finance, operations research and biomedical research.

Weina Wang

Title: Sharp Waiting-Time Bounds for Multiserver Jobs

Abstract: Multiserver jobs, which are jobs that occupy multiple servers simultaneously during service, are prevalent in today's computing clusters. However, little is known about the delay performance of systems with multiserver jobs. In this talk, to capture the large scale of modern computing clusters, we consider the scheduling problem of multiserver jobs in systems where the total number of servers becomes large. The multi-server job model opens up new scaling regimes where both the number of servers that a job needs and the system load scale with the total number of servers. Within these scaling regimes, we first characterize when the queueing probability diminishes as the system becomes large, and then turn our focus to the mean waiting time, i.e., the time a job spends in queueing rather than in service. In particular, we derive order-wise sharp bounds on the mean waiting time under various policies. The sharpness of the bounds allows us to establish the asymptotic optimality of a priority policy that we call P-Priority, and the strict suboptimality of the commonly used First-Come-First-Serve (FCFS) policy. This talk is based on joint works with Yige Hong, Qiaomin Xie, and Mor Harchol-Balter.

Bio: Weina Wang is an Assistant Professor in the Computer Science Department at Carnegie Mellon University. Her research lies in the broad area of applied probability and stochastic systems, with applications in cloud computing, data centers, and privacy-preserving data analytics. She was a joint postdoctoral research associate in the Coordinated Science Lab at the University of Illinois at Urbana-Champaign, and in the School of ECEE at Arizona State University, from 2016 to 2018. She received her Ph.D. degree in Electrical Engineering from Arizona State University in 2016, and her Bachelor's degree from the Department of Electronic Engineering at Tsinghua University in 2009. Her dissertation received the Dean's Dissertation Award in the Ira A. Fulton Schools of Engineering at Arizona State University in 2016. She received the Kenneth C. Sevcik Outstanding Student Paper Award at ACM SIGMETRICS 2016.

Kuang Xu

Title: Information design for load balancing

Abstract: Information has long played a crucial, if somewhat understated, role in dynamic resource allocation and control for queueing networks. I will discuss some recent results, as well as open questions, in understanding how to perform efficient inference and information exchange in load balancing systems. In the first part of the talk, we will examine the problem of using empirical operational data to perform statistical inference. We ask what types of observables provide the most information and how dynamic control polices may impact the inferential power of the data produced. In the second part, we will look at the problem of load balancing under a limited communication budget by using approximations of the system state. Our goal is to obtain conditions under which quality approximation leads to strong performance guarantees, and to generate desirable state approximations using the least amount of communication possible. Based on joint work with Gal Mendelson (Stanford).

Bio: Born in Suzhou, China, Kuang Xu is an Associate Professor of Operations, Information and Technology at the Stanford Graduate School of Business, Stanford University, and Associate Professor by courtesy at Stanford Electrical Engineering. He received the B.S. degree in Electrical Engineering (2009) from the University of Illinois at Urbana-Champaign, Urbana, Illinois, USA, and the Ph.D. degree in Electrical Engineering and Computer Science (2014) from the Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. His research interests lie in the fields of applied probability theory, optimization, and operations research, seeking to understand fundamental properties and design principles of large-scale stochastic systems, with applications in queueing networks, experimentation design and machine learning. He has received a First Place in the INFORMS George E. Nicholson Student Paper Competition, a Best Paper Award as well as Outstanding Student Paper Award at ACM SIGMETRICS, and an ACM SIGMETRICS Rising Star Research Award. He currently serves as an Associate Editor for Operations Research and Management Science. Outside of academia, he serves as the Principal Data Science Advisor for Shipt Inc.

Christina Lee Yu

Title: Graph Agnostic Randomized Experimental Design under Heterogeneous Linear Network Interference and Beyond

Abstract: Randomized experiments are widely used to estimate causal effects of proposed ``treatments'' in domains spanning across physical sciences, social sciences, medicine, and technology industries. However, classical approaches to experimental design rely on critical independence assumptions that are violated when the outcome of an individual a may be affected by the treatment of another individual b, referred to as network interference. Under such network interference, naively using popular estimators and randomized experimental designs can result in significant bias and loss of efficiency. We consider heterogeneous linear and polynomial potential outcomes models for network interference, under which we propose simple estimators for the total treatment effect that output unbiased estimates with low variance under simple

randomized designs. Our solution and statistical guarantees do not rely on restrictive network properties, allowing for highly connected graph structures. When the network is completely unknown, we provide a simple unbiased and efficient estimator under a staggered rollout randomized design, showing that the flexibility from additional measurements over time can relax requirements of network knowledge.

Bio: Christina Lee Yu is an Assistant Professor at Cornell University in the School of Operations Research and Information Engineering. Prior to Cornell, she was a postdoc at Microsoft Research New England. She received her Ph.D. and MS in electrical engineering and computer science from Massachusetts Institute of Technology, and she received her BS in computer science from California Institute of Technology. She received honorable mention for the 2018 INFORMS Dantzig Dissertation Award, and she is a recipient of the 2021 Intel Rising Stars Award and 2021 JPMorgan Faculty Research Award. Her research interests include algorithm design and analysis, high dimensional statistics, inference over networks, sequential decision making under uncertainty, online learning, and network causal inference.