Stochastic networks 2018

25 – 29 June 2018

International Centre for Mathematical Sciences, Edinburgh

Abstracts

Blanchet, Jose

Distributionally robust performance analysis for stochastic networks

In most problems involving stochastic modelling, we do not really know the underlying model of randomness, this talk is about quantifying model uncertainty in great generality. We use the theory of optimal transport to explicitly compute bounds on worst-case probabilities within an optimal transport ball around a baseline model. We will show that our theory connects naturally to state-of-the-art machine learning and artificial techniques, so our approach may provide fundamental principles which can be used for data-driven analysis and control of stochastic networks in a principled way which is fully aligned with contemporary machine learning techniques.

Borst, Sem

Scalable load balancing in networked systems

We will discuss scalable load balancing algorithms, which provide favourable delay performance in large-scale systems, and yet only require minimal implementation overhead.

Through the initial stages of the talk we focus on a basic setup - commonly referred to as the supermarket model - with a single dispatcher and N identical parallel servers. A popular class of load balancing algorithms are so-called JSQ(d) policies, where an incoming task is assigned to a server with the shortest queue among d servers selected uniformly at random. As the name reflects, this class includes the celebrated Join-the-Shortest-Queue (JSQ) policy as a special case ($d = N$), which has strong stochastic optimality properties.

In order to explore the fundamental trade-off between delay performance and implementation overhead, we consider an asymptotic regime where the total arrival rate and number of servers N grow large in proportion and the diversity parameter $d(N)$ depends on N. We show that the fluid limit and the diffusion limit correspond to those for the ordinary JSQ policy when $d(N)$ tends to infinity and $d(N)/(\sqrt{(N)}\log(N))$ tends to infinity, respectively, as the number of servers N grows large. Thus, the optimality of the JSQ policy can be retained at fluid level and diffusion level while reducing the amount of communication overhead by nearly a factor N and $\sqrt{(N)}/\log(N)$, respectively. In addition, we analyze load balancing mechanisms which leverage memory at the dispatcher in order to reduce the amount of communication overhead further while maintaining low delay.

A key facet of the supermarket model is that any task can be handled equally efficiently by any server, which provides analytical tractability but does not cover situations where the processing speeds of a specific task at the various servers may be different. These scenarios may arise due to data locality, task-server affinity relations, or broader compatibility constraints. In order to capture such heterogeneity features, we broaden the attention to network scenarios where the various servers are interconnected by some underlying graph topology G_N. Tasks arrive with rate lambda at each of the N servers, and each task is assigned to the server with the shortest queue among the one where it appears and its neighbours in the graph G_N. We establish conditions - in terms of the average degree and structure of the graph G_N - for the fluid-scaled and diffusion-scaled occupancy processes to be equivalent to those for the case of a clique. The results demonstrate that the optimality of a clique can be asymptotically achieved with far fewer connections, provided the graph topology G_N is suitably random. (Based on joint work with Mark van der Boor, Johan van Leeuwaarden, Debankur Mukherjee and Phil Whiting (Macquarie))

Dai, Jim

Steady-state approximations

Diffusion models and mean-field models have been used to approximate many stochastic dynamical systems. A functional strong law of large numbers or a functional central limit theorem justifies such an approximation. Such a result, however, does not justify the convergence of the equilibria of pre-limit systems to the equilibrium of a limit system. In this talk, I will touch on three recently developed methods for justifying equilibrium convergence in the setting of bandwidth sharing networks and multiclass queueing networks, with a focus on moment generating function method and the state-space-collapse. (Based on joint works with Anton Braverman, Chang Cao, Masakiyo Miyazawa, and Xiangyu Zhang).

Duffy, Ken

DNA coded randomised programs

Did you know that your immune system, and that of all jawed vertebrates, is underpinned by a randomised programme coded into your DNA? The aim of that vignette, and this talk, is to tempt you into studying a set of significant stochastic networks about which quantitative detail is currently being revealed by experimental advances. The talk will first motivate these investigations in broad terms to coax you into the subject. As a specific example, it will then describe the mathematics behind a randomised programme Tom Weber, Leila Perie and I developed for inference of the average depth of a directed tree. That algorithm is presently being coded into DNA by the Schumacher Lab. at the Netherlands Cancer Institute. No prior biological knowledge of the audience will be assumed.

Ferrari, Pablo

Soliton decomposition of the Box-Ball-System

The Box-Ball System (BBS) is a cellular automaton introduced by Takahashi and Satsuma (TS) as a discrete counterpart of the Korteweg & de Vries (KdV) differential equation. Both systems exhibit solitons, solitary waves that conserve shape and speed even after collision with other solitons. The BBS has state space $\{0,1\}^Z$ where each integer represents a box which may contain one ball or be empty. One iteration of the automaton consists of a carrier visiting successively all boxes from left to right, picking balls from occupied boxes and depositing one ball, if carried, at each visited empty box. Building on the TS identification of solitons, we provide a soliton decomposition of the ball configurations, show that the dynamics reduces to a hierarchical translation of the components and prove that shift stationary measures with independent soliton components are invariant for the dynamics.

We also prove that the asymptotic speed of a tagged soliton of size k converges to a positive real number v_k and exhibit the equations satisfied by the speeds $(v_k)_{k\geq 1}$. A detailed analysis shows that among many others, product measures and Ising measures have independent soliton components. (Joint work with Chi Nguyen, Minmin Wang, Leonardo Rolla and Davide Gabrielli).

Gamarnik, David

A lower bound on the queueing delay in resource constrained load balancing

We consider a large scale distributed service model with n servers. We assume that the dispatching decisions are made by a central dispatcher endowed with a bounded memory, and with the ability to exchange messages with the servers. We study the fundamental resource requirements (memory bits and message exchange rate), in order to drive the expected queueing delay in steady-state of a typical job to zero, as n increases. We develop a novel approach to show that, within a certain broad class of "symmetric" policies, appropriately defined, every dispatching policy with a message rate of the order of n, and with a memory of order O(log) n bits, results in an expected queueing delay which is bounded away from zero, uniformly as the number of servers n diverges to infinity. (Joint work with John Tsitsiklis (MIT) and Martin Zubeldia (MIT))

Hajek, Bruce

On the structure of preferential attachment networks with community structure

An extensive theory of community detection has developed within the past few years. The goal is to discover clusters of vertices in a graph based on the edges of the graph. Much work has focused on evaluation of algorithms, including message passing algorithms, for a particular generative model, namely, the stochastic block model. The stochastic block model is a variant of Erdos-Renyi type graph. In this talk we focus on the problem of community detection for the Barabasi-Albert preferential attachment model with communities, defined by J. Jordan. In such model, vertices are sequentially attached to the graph, with preference to attach more edges to existing vertices with larger degrees, multiplied by affinities based on community membership. While more complicated than SBM, it is shown that the model has sufficient structure to formulate approximate belief propagation algorithms for community detection. Based on joint work with S Sankagiri.

Kapodistria, Stella

Prescribed maintenance under uncertainty

Motivated by industrial practice, we consider several models for a dynamically adapting lifetime distribution (e.g., fluid models with jumps) and for the monitoring of the condition (e.g., Brownian motion with drift with bayesian parameters) of an asset and study the structural properties and the asymptotic behaviour of the (near) optimal cost based policy for the purpose of prescribed maintenance. For the numerical calculation of the policies, we propose some fast converging algorithms. We apply these models to a wind farm.

Martin, James

Multi-type queues and exclusion processes

I'll describe some old and some new results about multi-type queues and multi-type exclusion processes.

Consider for example a system of N particles with labels $1,2,\ldots,N$ jumping on a ring of size N. Each of the N sites contains a single particle. Whenever particle i is immediately to the right of particle j with $i < j$, the two particles exchange places at rate 1. I'll review various descriptions of the stationary distribution of such "multi-type totally asymmetric simple exclusion processes", for example in terms of outputs of systems of Markovian queues in series, or in terms of traces of matrix products. Some interesting facts about queues in series emerge.

Suppose we now add jumps in the reverse direction (i.e. when particle i is to the left of particle j for $i < j$) with some rate q in $(0,1)$. I'll describe how the constructions above generalise to this "partially asymmetric" case. A new queueing description has various nice consequences, both probabilistic (descriptions of the formation of "convoys" of nearby particles with similar labels for large N) and algebraic (e.g. a common denominator for the stationary probabilities of the various configurations as rational functions of q).

If time permits I'll mention some other variants (zero-range processes, Busemann functions for last-passage percolation) and some open problems.

Minca, Andreea C

Inhomogeneous financial networks and contagious links

We propose a framework for testing the possibility of large cascades in financial networks. This framework accommodates a variety of specifications for the probabilities of emergence of 'contagious links' conditional on a macroeconomic shock, where a contagious link leads to the default of a bank following the default of its counterparty. Under general contagion mechanisms and incomplete information, the financial network is modelled as an inhomogeneous random graph, where the conditional probabilities of contagious links depend on banks' characteristics. We give bounds on the size of the first order contagion and testable conditions for it to be small. (Joint with Hamed Amini)

Moerters, Peter

The age-dependent random connection model

In this talk I propose a simple network model which is scale-free with spatially induced clustering. Vertices arrive according to a Poisson process, are randomly placed on a d-dimensional torus and connected to existing vertices with a probability depending on their spatial distance and birth times. I show that this model has a weak local limit, which is an age-dependent random connection model on a Poisson process in d-dimensional space and report on ongoing projects related to this model in collaboration with P Gracar, A Grauer and L Luechtrath (Cologne).

Puha, Amber

Asymptotic behaviour of a critical fluid model for a processor sharing queue via relative entropy

In this talk, we discuss a new approach to studying the asymptotic behaviour of fluid model solutions for critically loaded processor sharing queues. For this, we introduce a notion of relative entropy associated with measure-valued fluid model solutions. In contrast to the approach used in our prior work, which does not readily generalize to networks of processor sharing queues, we expect the relative entropy approach to be more robust. Indeed, we anticipate that similar notions involving relative entropy may be helpful for understanding the asymptotic behaviour of critical fluid model solutions for stochastic networks operating under various resource sharing protocols naturally described by measure-valued processes. This work is joint with Ruth J. Williams (UCSD).

Salez, Justin

Mixing time and cutoff for the random walk on random digraphs

A finite ergodic Markov chain exhibits cutoff if its distance to equilibrium remains close to 1 over a certain number of iterations and then abruptly drops to 0 on a much shorter time scale. Originally discovered in the context of card shuffling (Aldous-Diaconis, 1986), this remarkable phenomenon is now rigorously established for many reversible chains. Here we consider the non-reversible case of random walks on sparse directed graphs, for which even the equilibrium measure is far from being understood. We work under the configuration model, allowing both the in-degrees and the out-degrees to be freely specified. We establish the cutoff phenomenon, determine its precise window and prove that the cutoff profile approaches a universal shape. We also provide a detailed description of the equilibrium measure. This is joint work with Charles Bordenave and Pietro Caputo.

Stolyar, Alexander

A queueing system with on-demand servers

In the basic model a random flow of customers is served by servers invited on-demand. Invited servers do not show up immediately, but arrive after random delays. One customer and one server are matched for service, after which they both leave the system. The basic objective is to minimize both customer and server waiting times. The model has several applications, including call/contact centres, inventory systems, telemedicine. We study two feedback-based adaptive server invitation algorithms, and derive fluid and diffusion limits (including steady-state limits) as the customer arrival rate becomes large. The results imply, in particular, that in the classical single-item inventory system with exponentially distributed lead times and order crossovers, our algorithm provides potentially infinite average inventory cost improvement over the standard constant-base-stock algorithm. We also consider some extensions of the basic model. (Based on joint works with G.Pang (Pennstate), Q.Wang (UIUC), L.Nguyen (Lehigh Univ.)

Tezcan, Tolga

Yardstick competition to improve access and quality in health care systems

Yardstick competition is a regulatory scheme for local monopolists (e.g., hospitals), where the monopolist's reimbursement is linked to her performance relative to other equivalent monopolists. This regulatory scheme is known to work well in providing cost-reduction incentives and serves as the theoretical underpinning behind the hospital prospective reimbursement system used throughout the developed world. Recent initiatives in healthcare regulation focus on improving quality by

modifying the standard prospective payment systems while containing costs. In this talk we use a game-theoretic queueing model to investigate how these policies and yardstick competition perform in service systems, where, in addition to incentivizing cost reduction, the regulator's goal is to incentivize the provider to reduce customer waiting times and (unnecessary) patient readmissions.

Tran, Ngoc
Neuron spike sorting
Neural activity in a brain region forms a stochastic network. When a neuron `spikes', it sends out a sharp waveform recorded by neighbouring electrodes. Spike sorting is the problem of assigning spikes to individual neurons. Over a recording session, the neuron may move and its waveform may change, posing unique challenges to obtaining long-term recordings. In this talk, we review existing techniques, showcase recent advances and discuss open problems on stochastic networks based on spike sorting challenges. No prior knowledge on neuroscience is required. (Joint work with Hanlin Zhu and Chong Xie, Department of Biomedical Engineering, University of Texas at Austin

van der Hofstad, Remco
Citation networks as a window to science: a case study
Citation patterns between papers form a window to how science works. In citation networks, the nodes are papers and the directed edges are formed by one paper citing the other. Citation networks give us a wealth of information about differences in citation cultures between subfields, how scientists collaborate, etc. These insights are useful to interpret citation statistics of papers and scientists beyond simply counting citations. How long does it take publications to be cited? How long does it take papers to be forgotten, and how much does this depend on the citation patterns of the papers early on? How many citations do papers get, and how do these develop in time? How variable are citation patterns when comparing papers within a field, as well as between fields? Can one quantify what a 'good' paper is based on citation network structure?

In this talk, we present empirical data of citation patterns obtained from the Web of Science database. We investigate aging of citation patterns, the presence of dynamical power laws in them, as well as how predictable citations behave after having observed them for a while. We then continue to describe a possible model for them based on preferential attachment models with aging and fitness that, on a qualitative level, shows similar connectivity patterns as in citation networks. We also discuss PageRank, which is an algorithm that gives an order to vertices in a directed network, and show how local weak convergence can be used to analyse PageRank in large graph limits.

We close with open problems.

We assume no prior knowledge in graph theory, probability or otherwise.

This is joint work with Alessandro Garavaglia (TU/e), Nelly Litvak (TU/e and Twente University), and Gerhard Woeginger (RWTH Aachen).

Williams, Ruth
Queues with reneging and random order of service: fluid limits and their asymptotic behaviour
Random order of service (ROS) is a natural scheduling policy for queueing systems where there is no implicit ordering of jobs. Applications in which ROS arises include the processing of biological molecules in a cell with limited resources. Indeed, multiclass queues with ROS and reneging have been used to gain insights into correlations between different types of proteins processed by a common enzyme in E. Coli cells.

To date, this analysis has relied on exact results that assume exponential distributions for processing and patience times. It is desirable to have methods for analysing such systems when these distributions are more general, especially to accommodate lengthy biological operations such as binding/unbinding, folding, transcription and translation.

In this talk we will consider a multi-class queueing model with reneging operating under ROS and with generally distributed interarrival, processing and patience times. We will use measure-valued processes to describe the dynamic evolution of the model, and establish a fluid approximation for this representation. Obtaining a fluid limit for this network requires a multi-scale analysis of its fast and slow components. We further study the asymptotic behaviour of fluid model solutions as time goes to infinity. (This talk is based on joint work with Reza Aghajani.)

Zwart, Bert
Heavy tails: asymptotics, algorithms, applications
Recently, we have developed a sample-path large deviations principle for random walks with regularly varying jumps. In this talk we (i) apply this result to a class of queueing networks; (ii) we utilize this result to design a universally applicable importance sampling method, and (iii) extend this result to a class of Markov-additive processes.

We also present complementary result for random walks with Weibullian jumps, and apply these results to obtain large-deviations results for the queue length in the queue with multiple servers. The number of big jobs to create a large queue length is shown to be fundamentally different compared to the regularly varying case.

The talk is based on joint work with Mihail Bashba, Jose Blanchet, Bohan Chen, and Chang-Han Rhee.

Public Lecture
Tuesday 26 June 2018

Taylor, Peter (University of Melbourne)
Modelling block arrivals in the Bitcoin blockchain
In 2009 the pseudononymous Satoshi Nakamoto published a short paper on the Internet, together with accompanying software, that proposed an `electronic equivalent of cash' called Bitcoin. At its most basic level, Bitcoin is a payment system where transactions are verified and stored in a distributed data structure called the blockchain. The Bitcoin system allows electronic transfer of funds without the presence of a trusted third party. It achieves this by making it `very hard work' to create the payment record, so that it is not computationally-feasible for a malicious player to subsequently repudiate a transaction and create the forward history without it.

The Nakamoto paper contained a simple stochastic model, used to show that the above-mentioned malicious player would be very unlikely to succeed. Unfortunately, this calculation contained an error, which I shall discuss and show how to correct.

The Bitcoin payment record is stored in a data structure called the blockchain. Blocks are added to this structure by `miners' working across a distributed peer-to-peer network to solve a computationally difficult problem. With reference to historical data, I shall describe the block mining process, and present a second stochastic model that gives insight into the block arrival process.

Finally, I shall make some brief comments about how stochastic modelling can be used to address the current concerns that the transaction processing rate of the Bitcoin system is not high enough.