Complexity Retreat 2017 Booklet

Schedule

Tuesday, 2nd May

Time	Event	Location
08:30	Coach departs	Behind the department
	Arrival	YHA Ravenstor
11:30 – 12:00	Welcome	Conference room
12:00 – 13:00	Lunch	
13:00 – 14:00	MSc Poster Session	Lounge
14:00 – 14:15	Coffee Break	
14:15 – 15:15	Dan Pearce: Post-doc in Theoretical Physics, Physics of Condensed and Biological Matter Group, University of Leiden. **Controlling an active fluid with geometry and topology**	Conference Room
15:15 – 15:45	Coffee Break	
15:45 – 16:45	Spectra Analytics: Marcus Ong - CEO and founder of Spectra Analytics. Dan Sprague - Chief Science Officer for Spectra Analytics. **Data Science in Industry - A Short Practitioner's Guide by Spectra Analytics**	Conference Room
16:45 – 18:00	Time to work on WARPs	
18:00 – 19:30	Dinner	Dining room
19:30 onwards	Free Time	

Wednesday, 3rd May

07:30-09:00	Breakfast	Dining room
09:00-09:30	6x5min talks (2 streams)	Conference/Lounge
09:30-09:35	Coffee Break	
09:35-10:05	6x5min talks (2 streams)	Conference/Lounge
10:05-10:20	Coffee Break	
10:20-10:40	4x5min talks (2 streams)	Conference/Lounge
10:40-12:00	Time to work on WARPs	
12:00-13:00	Lunch	Dining room
13:00-14:00	Tobias Galla: Senior Lecturer in Theoretical Physics, Complex Systems and Statistical Physics Group, University of Manchester. **You are a young and aspiring physicist. Is working at the interface with economics or biology a good idea?**	Conference room
14:00-18:00	Free time	
18:00-19:30	Dinner	Dining room
19:30 onwards	Free time	

Thursday, 4th May

07:30-09:00	Breakfast	Dining room
09:00-09:30	6x5min talks (2 streams)	Conference/Lounge
09:30-09:35	Break	
09:35-10:05	6x5min talks (2 streams)	Conference/Lounge
10:05-10:20	Break	
10:20-10:40	4x5min talks (2 streams)	Conference/Lounge
10:40-11:00	Coffee Break	
11:00-12:00	Workshop choice: 1) Starting your PhD – Sam Johnson 2) Writing PhD thesis as you go – Fede Botta 3) Life after PhD/Postdoc – Martine Barons	Conference/Lounge/Dining room
12:00-13:00	Lunch	Dining room
13:00-14:45	Director's Talks + SSLC Meeting + Prizes	Conference/Lounge
14:45-15:00	Group Photo	
15:00-16:00	Time to work on WARPS	
16:00-16:15	Coffee Break	
16:15-18:30	Time to work on WARPS	
18:30-20:00	Dinner	Dining room
20:00-onwards	Free Time	

Friday, 5th May

07:30-09:00	Breakfast	Dining room
10:00	Coach departs :(	

Student Talk Schedule

Wednesday, 3rd May

Time	Stream 1 (Lounge)	Stream 2 (Conference Room)
9:00-9:30	Rob Eyre	Sophie Meakin
	Michael Pearce	Cameron Lack
	Jeremy Reizenstein	Katherine Broadfoot
9:35-10:05	Iliana Peneva	Chris Davis
	Jim Skinner	Joe Hilton
	Ayman Boustati	Emma Davis
10:20-10:40	Jessie Liu	Jack Binysh
	Matt Groves	Janis Klaise

Thursday, 4th May

Time	Stream 1 (Lounge)	Stream 2 (Conference Room)
9:00-9:30	Tim Pollington	Robert Gowers
	Alex Bishop	Yihe Lu
	Giovanni Mizzi	
9:35-10:05	Alvaro Cabrejas-Egea	Henry Charlesworth
	Roger Hill	Arthur King
	Guillem Mosquera	Jason Lewis
10:20-10:40	Luke Whincop	Ben Atkins
	Laura Guzman-Rincon	Gian Lorenzo Spisso

Student Abstracts

- Alex Bishop
    - **Introduction to probabilistic programming with Stan** I will give an introduction to probabilistic programming with Stan (http://mc-stan.org/) and demonstrating how it can be used with Python/R/MATLAB to specify a probabilistic model and perform full Bayesian inference with Hamiltonian Monte Carlo (HMC) methods.
- Alvaro Cabrejas Egea
    - **STL: A Seasonal-Trend Decomposition Based on Loess** I will be presenting the STL algorithm for decomposing a time series into trend, seasonal and remainder components and its recommended diagnostic tools. STL has a simple design that consists of a sequence of applications of the loess (local regression) smoother. It allows fast computation, even for very long time-series and large amounts of trend and seasonal smoothing, while being robust enough to prevent the components from being distorted by aberrant behavior in the data.
- Arthur King
    - **Extreme Human Collective Motion at Heavy Metal Concerts** Human collective behaviour can vary from calm to panicked depending on social context. Observational studies of such motion generally relies on volunteers following simple behavioural rules or simulating situations of panic. While these studies are guided by ethical principles for human studies, extreme forms of collective motion arising in real-world scenarios can occur when the participants are in atypical and highly stressed psychological states. As a consequence, our empirical understanding of the most dangerous forms of human collective motion are limited by a scarcity of data. Here I present a past study [1] of the highly energised collective motion of attendees at heavy metal concerts. This unique group of people offers an ethical testbed for probing the most extreme forms of human collective motion. The phenomenology found in this social context sheds new insights on how groups of people move, and suggests new strategies to minimize genuine harm occurring in riots, protests, and panicked evacuations.

      [1] Silverberg, Jesse L., et al. "Collective motion of humans in mosh and circle pits at heavy metal concerts." Physical review letters 110.22 (2013): 228701.
- Ayman Boustati
    - **Inference in Implicit Generative Models (Working Title)** In this talk I will give an overview of inference methods for cases where we do not have an explicit model for our data. Instead, we only have a simulator for the data.
- Benjamin Atkins
    - **Mastering the game of Go** The game of Go is a particularly difficult game for artificial intelligence to master due to an enormous search space of possible moves and the inherent difficulty of evaluating board positions and moves. I will introduce a recently developed program, named AlphaGo, that employs 'value networks', 'policy networks' and a combination of supervised and reinforcement learning to efficiently and accurately predict the best move at each point in the game. When compared to previous models, AlphaGo achieved a 99.8% win rate. Even more impressive, it was able to defeat the human European Go champion by five games to none on a full sized board, a feat that has never been achieved before even once. Finally, I will give an overview as to how these methods can be applied to my work in the control of disease outbreaks.
- Cameron Lack
    - **Strategies for efficient numerical implementation of hybrid multi-scale agent-based models to describe biological systems** Clifone et al. review different methods by which to efficiently simulate complex biological systems which are

operating across multiple scales. Since biologically related systems often operate across multiple spatio-temporal scales, individual scale models must be linked in a way that allows for a dynamic transfer of information between the scales. A powerful methodology to achieve this, is to combine discrete approaches, agent-based models (ABMs), with continuum models.
In this talk I will discuss the types of systems these methods can be applied to and give an overview of the methods themselves.

- Christopher Davis
    - **Maps and disease** An overview of plotting data on maps with examples from my work on sleeping sickness in the DRC.
- Emma Davis
    - **Modelling seasonality in round worm infections** Using an ODE model to simulate the within-host and external aspects of the round worm life-cycle it is possible to explicitly consider the effects of seasonal conditions on the infective eggs and hence onwards transmission. Different seasonal mass drug administration timings can then be simulated to reveal the optimal time of year for highest intervention impact.
- Gian Lorenzo Spisso
    - **Job Automation: How long before a computer can do your PhD better than you?**
      Mine? Two days.
      An economic perspective on the advances in Machine Learning.
- Giovanni Mizzi
    - **The MapReduce Algorithm**
      MapReduce is a programming model and an associated implementation for processing and generating big data sets with a parallel, distributed algorithm on a cluster.
      It was originally proposed by Google, and it has been the basis of a whole series of open-source applications.
      The most well known of these is Hadoop, which exists in many different distributions, usually customised to the needs of whoever is using it.
      Everybody has got one, Amazon has got one, Linkedin has got one. Google probably does not have it any more because it is probably already obsolete for them.
      The University of Warwick has got one, and it is on Chiron.
      In the best case, this could be useful for your research and you'll start using it.
      In the worst case, you'll know that it exists, if you ever are in need of something like this.
- Guillem Mosquera Donate
    - **Detecting Communities with Markov Chains on Graphs** Community Detection is an open research field in Network Science. Here I introduce a recent approach to such problem that uses dynamical Markov processes to unfold community structure across different scales. Additionally, I'll show how this method can be regarded as a generalization of popular techniques such as Modularity Optimization or Spectral Clustering.
- Henry Charlesworth
    - **Freedom Maximisation and Collective Motion** In the absence of any specific goal to be achieved perhaps a sensible principle to follow is that it is best to be prepared and to keep one's options open as much as is possible - that is you should seek to maximise the number of potential futures that are available to you because by doing so you are in a sense remaining in control of your environment as much as you can. I will briefly discuss an attempt to formalise this idea in the language of information theory called "empowerment" and then go on to talk about an application of this principle to a model of flocking which produces motion which is highly-aligned and cohesive under a wide range of conditions and which is robust to noise.
- Iliana Peneva

    - **Image-to-image translation with pix2pix**
      Deep learning has a wide range of applications: from brain tumour detection and self-driving cars to sorting cucumbers and differentiating between cats and dogs. In this talk I will present another of its applications – image-to-image translation, and will illustrate how it works with examples generated by the pix2pix method (Isola et al, 2016).
- Jack Binysh
    - **The theory of hitches** I'll talk about the mathematical theory of hitches - i.e. things you tie around a pole so your horse doesn't get away. If I can get my hands on some bits of rope and some poles, it will be a hands on talk!
- Janis Klaise
    - **Aspects of network robustness** I will survey some classical results on component robustness in random network models and discuss my observations in some special cases, namely degree-regular networks, networks with a non-negligible number of triangles and potential for discontinuous transitions.
- Jason Lewis
    - **Swarming and Art** Most people are aware of the visual impact of a swarm of thousands of fish or birds moving in unison, splitting and recombining, and producing an endless array of shapes and structures. In this talk, I will present a few simulations of this type of collective behaviour in which the swarm has transitioned from art to become the artist.
- Jeremy Reizenstein
    - **Signatures as memory in RNNs** Recurrent neural networks are among the machine learning algorithms of choice when you want to learn a function whose inputs are sequences. Making them remember stuff from a long sequence of input requires tweaks, the most popular of which is Long Short Term Memory, which can be slow to learn but very effective. I'll explain why it might be a good idea to use rough path signatures as an extended way to give RNNs memory, and show some results on how well this works.
- Jessie Liu
    - **Improving Stability of Imaging-Behavioral CCA with Supervised Dimension Reduction** Using self-developed dimension reduction method, Supervised Dimension Reduction (SDR), to show the stability of results of Canonical Correlation Analysis (CCA) are improved. The analysis is done on the Human Connectome Project 900 data release.
- Jim Skinner
    - **Bayes' Rules!** I will discuss Machine Learning from a Bayesian perspective, what it means to be Bayesian and what all this "Bayesian methods don't overfit" hype is about.
- Joe Hilton
    - **Modelling emerging diseases** Emerging diseases arise when new pathogens adapt to human physiology. They are characterised by high mortality, low transmissibility, and discrete well-defined outbreaks which tend to go extinct in their early stages. Early epidemic dynamics are well-approximated by branching processes, which can be quickly and efficiently fitted to outbreak size data. However, standard branching process approaches are susceptible to inaccuracies because they do not take into account population constraints on disease transmission. In this talk I introduce these models and explore potential modifications which could improve their accuracy in fitting to data.
- Katherine Broadfoot
    - **Measles outbreaks in Swansea** I will talk a bit about my research - mainly the use of MLE and MCMC techniques with data from a measles outbreak in Swansea in 2013, and focus at the end on the importance of keeping in touch with external partners (if you have one!)

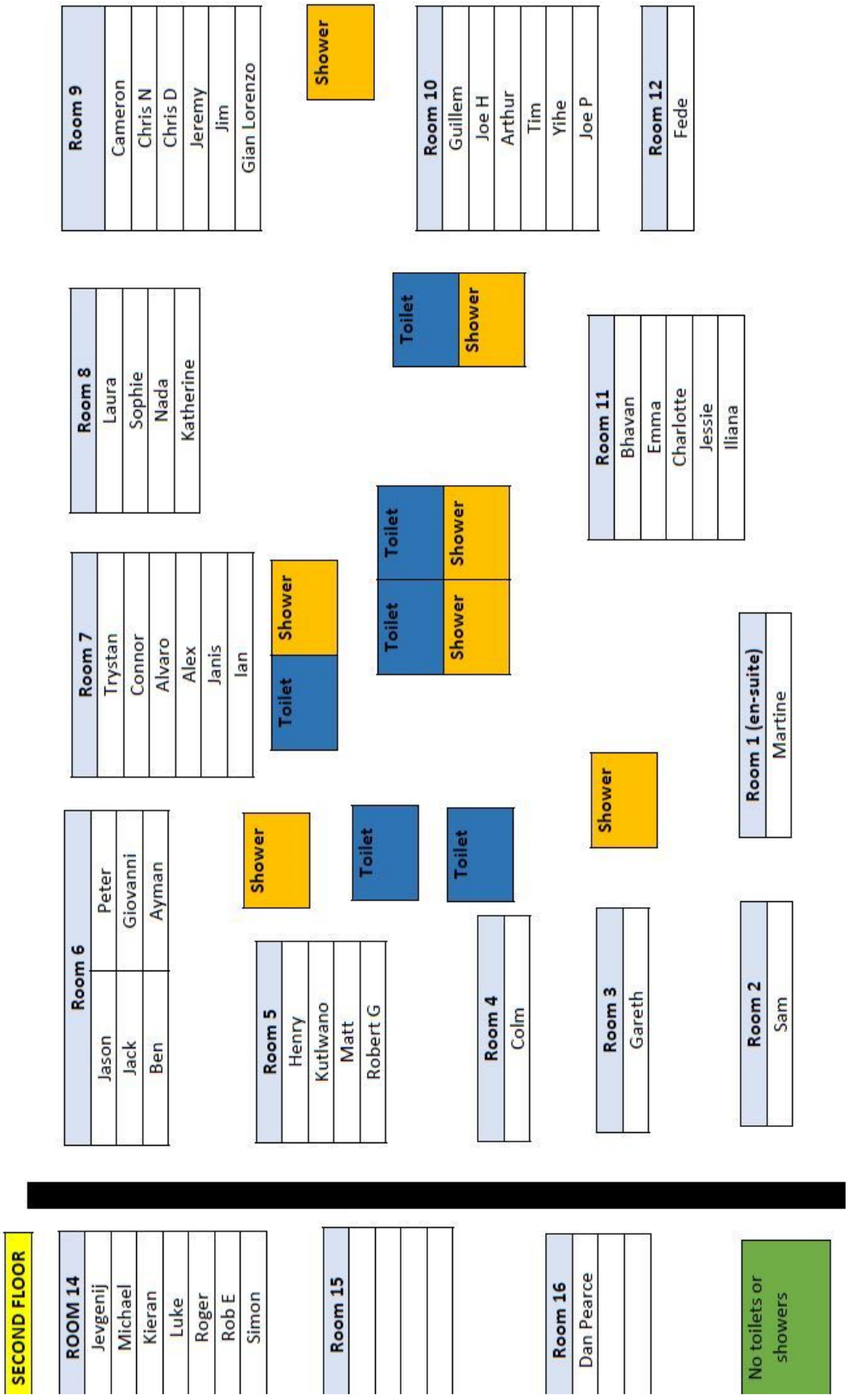
- Luke Whincop
    - **Improving the attribution of sources of human infection** The primary question of my research asks: how can we best attribute the sources of human infection? The uncertainty regarding the answer to this has inhibited effective public health intervention by government agencies and industry; therefore, there is a pressing need to address the research question. One research approach is to analyse attribution algorithms and to use statistical methods, to improve the accuracy and reliability of source attribution. I analysed the accuracy of the STRUCTURE algorithm on two datasets, one containing isolates of Campylobacter coli, the other containing isolates of Campylobacter jejuni; I used cross-validation, taking tenths of the isolates of each source, to improve reliability. STRUCTURE produces reasonable attribution accuracies on each dataset; taking subsamples of the datasets, with an even number of isolates per source, as well as removing sequence types that generate error, can improve overall accuracy. In addition, taking all isolates of one source can lead to over-attribution to that source, but not always. To conclude, modifying the number of isolates you choose to take in an analysis of STRUCTURE can affect overall attribution accuracy. Future research aims to compare different attribution algorithms, and improve their accuracy and reliability, to help further answer the above research question.
- Matthew Groves
    - **Optimal information collection and learning in games** In this talk, I will discuss a range of pairwise Bayesian myopic sampling techniques and how they can be used to improve the efficiency of evolutionary learning methods applied to stochastic two-player games.
- Michael Pearce
    - **"Deep" Linear Regression and Wide Learning** We've all heard of deep learning and how amazing it supposedly is, we've all heard of Linear Regression and how overused and terrible it supposedly is. Surely we can combine these to make something that is deep and terrible and unused? I will give an introductory talk starting with linear regression and ending with artificial neural networks and hopefully everyone will know what the hype is about.
- Rob Eyre
    - **A very very brief intro to Bayesian networks** This talk will give a shallow overview of Bayesian networks for application in areas of Social Science such as Public Health. I will describe (in brief) what Bayesian networks are, how they can be learned, and how inferences can be performed on them. At the very least I will give you a bunch of terms to google. This will include an example from my own research on household food security in rural South Africa (hopefully you do not need to google South Africa).
- Robert Gowers
    - **Linking Neuronal Microcircuits to Disease** Pyramidal neurons contain an array of different excitatory and inhibitory synapses along their length. Activation of these synapses can give rise to complex firing patterns, oscillations and synchrony between neurons. Understanding how these patterns emerge is crucial to understanding various neuropathologies, such as schizophrenia and epilepsy. By modelling the synaptic bombardment of pyramidal cells as a time-varying and spatially dependent stochastic process, insight can be gained into these firing patterns.
- Roger Hill
    - **Parameter optimisation using CMA-ES** I will be giving a brief overview of the algorithm I am using to optimise the parameters of my model to fit the data I want to represent.
      This will include a short background of my problem and then the ideas that are used to estimate the parameters using the CMA-ES.
- Sophie Meakin

    - **Moment closure approximations in stochastic epidemic models** Moment closure is an approximation method that allows us to write down a finite system of ordinary differential equations for the moments of a stochastic process. This allows us to derive (approximate) analytic expressions for the mean and variance of a stochastic process, avoiding the need for computationally-intensive stochastic simulations. I will describe the method by which we derive these equations for the first and second order moments of a continuous time Markov chain epidemic model.
- Tim Pollington
    - **My first GPU card** Sharing my first coding adventures in CUDA with C to use a graphics card to speed up evaluations of a likelihood function.
- Yihe Lu
    - **Response functions on a tapering dendritic tree** Neuronal dendrite branches and tapers. The branching tree structure fascinates scientists since the very beginning of the study in neuroscience by Ramon y Cajal, while the tapering structure remains relatively ignorable particularly in mathematical modelling, even though very fundamental theory derived by Rall has brought both branching and tapering into the PDE describing the membrane voltage (the general cable equation). Although many observations and theoretical works have suggested a tapering structure helps control signal attenuation and filtration, computational models often overlook tapering structures as they are small, local, and replacable by densive discretisation anyway. To better quantify the effects by tapering, Poznanski has built up a set of useful transformations which allows some tapering cable equation to be solved analytically. Based on that, we have generalised the sum-over-trips approach which worked on a branching network with all cylinders only to a method capable of acquiring response functions of a dendritic tree with both tapering and branching structures.

External Speakers

- Tobias Galla
    - Senior Lecturer in Theoretical Physics, Complex Systems and Statistical Physics Group, University of Manchester.
    - Talk: **You are a young and aspiring physicist. Is working at the interface with economics or biology a good idea?**
- Dan Pearce
    - Post-doc in Theoretical Physics, Physics of Condensed and Biological Matter Group, University of Leiden
    - Talk: **Controlling an active fluid with geometry and topology**
      Active nematic liquid crystals are achieved experimentally by combining high concentrations of microtubules, kinesin and ATP. These active nematic suspensions can then be localized to a 2D interface, such as the surface of a water droplet suspended in another medium. We will discuss the distribution of topological defects within an active nematic liquid crystal confined to the surface of a toroidal water droplet. Due to the topology of the surface, it is possible to coat a torus in a nematic material with no topological defects. When the activity of the nematic is sufficiently high, topological defects will spontaneously form, but they must preserve the net topological charge of the system, which is zero. Through experiment and simulation we confirm that the local distribution of topological defects is proportional to the Gaussian curvature. As the activity of the system is increased the reliance on the Gaussian curvature is diminished and the available area on different regions of the torus dictates the behavior.
- Spectra Analytics: Dan Sprague & Marcus Ong
    - Marcus Ong: CEO and founder of Spectra Analytics. PhD in Complexity Science.
    - Dan Sprague: Chief Science Officer for Spectra Analytics. PhD in Complexity Science.
    - Talk: **Data Science in Industry - A Short Practitioner's Guide by Spectra Analytics**
      This talk will discuss pursuing a career as a data scientist, the options available and the typical skills required. It will outline the differences between working in industry and academia to try to help you better choose the right option for you.
      We will then discuss how Spectra came into being, our projects and some of the lessons we have learned on the way.

Room Plan


Room 9
Cameron
Chris N
Chris D
Jeremy
Jim
Gian Lorenzo
Shower
Room 10
Guillem
Joe H
Arthur
Tim
Yihe
Joe P
Room 12
Fede
Room 8
Laura
Sophie
Nada
Katherine
Toilet
Shower
Room 11
Bhavan
Emma
Charlotte
Jessie
Iliana
Room 7
Trystan
Connor
Alvaro
Alex
Janis
Ian
Toilet
Shower
Toilet
Shower
Toilet
Shower
Room 6
Jason
Jack
Ben
Peter
Giovanni
Ayman
Shower
Toilet
Toilet
Shower
Room 1 (en-suite)
Martine
Room 5
Henry
Kutlwano
Matt
Robert G
Room 4
Colm
Room 3
Gareth
Room 2
Sam
SECOND FLOOR
ROOM 14
Jevgenij
Michael
Kieran
Luke
Roger
Rob E
Simon
Room 15
Room 16
Dan Pearce
No toilets or showers