**Causal inference in dynamical systems**

Complexity mini-project proposal.
Supervisors: Sach Mukherjee (Netherlands Cancer Institute, Amsterdam) & Robert MacKay
(Maths)

What does it mean to say that *X* causes *Y*? How should one infer and test causal relationships from data? For isolated, low-dimensional settings, causality may be obvious and these questions may seem of largely philosophical interest. However, for high-dimensional, poorly understood systems (e.g. in biology or the social sciences) questions of causality are hugely important in practice. For example, in modern biology, time-varying data on molecular components such as genes and proteins are used to infer networks that aim to show regulatory relationships between such components [e.g. 1,3]. In effect, such networks encode causal claims concerning a multivariate dynamical system. Yet basic questions about causality in such settings remain poorly understood. Pearl [2] has written extensively on causality and probabilistic models, placing the emphasis on intervention on the system, but there remain many open questions in the continuous-time dynamical case. How should the influence of one variable on another be quantified? What are the implications of stochasticity or measurement noise? How should interventional data (where one or more variables are externally manipulated) be handled?

The aim of this project is to explore causal inference for dynamical systems, with biological networks providing motivation and an important application area. The project will comprise two related goals. Initially, to fix ideas, the candidate should gain some familiarity with biological networks and associated inference methods, see e.g. [3] for an overview. Such understanding will provide a concrete focus and enable the student to test ideas on real data from cancer cells. Then, the student will explore the treatment of causality in dynamical systems, informed by recent work on metrics for stochastic dynamical systems [4], relating these ideas back to questions that arise in the biological context. *The project will not involve any "wet" experimental work and does not require any prior experience of biology.*

The project will be based mainly at the Netherlands Cancer Institute (NKI) in Amsterdam. We are a world-leading biological research institute, and offer a stimulating scientific environment, with expertise in diverse areas of biology. The Mukherjee Lab in particular focuses on statistical and computational approaches in molecular and cancer biology. The project could be extended into a PhD, offering the opportunity to spend part of the time in Amsterdam, and the potential to make a substantive contribution to cancer biology from the theoretical side.

References
[1] S. Mukherjee and T. P. Speed, Network inference using informative priors. *PNAS* 105(38): 14313–14318, 2008
[2] J. Pearl. *Causality*. CUP, 2009
[3] C. J. Oates & S. Mukherjee, Network inference and biological dynamics. *Annals of Applied Statistics*, to appear (http://arxiv.org/abs/1112.1047)
[4] R. S. MacKay, Robustness of Markov processes on large networks, Journal of Difference Equations & Applications 17:1155–67, 2011