

## **Parallelized Monte Carlo for Biological Network Inference**

Complexity mini-project proposal.

Supervisors: Sach Mukherjee (Netherlands Cancer Institute, Amsterdam), Anthony Lee (Statistics) & Adam Johansen (Statistics)

Network-based approaches are now widely used in biological applications to probe regulatory relationships between molecular components such as genes or proteins [1,2]. Markov chain Monte Carlo (MCMC) approaches can be used to allow Bayesian treatments of diverse network-related inference problems. However, the computationally demanding nature of these analyses has limited their applicability, both in terms of scaling to larger problems (higher dimensions, richer models) and facilitating the interactive use that is invaluable in exploratory analysis.

At the same time, the use of many-core computation, of which GPUs are a popular option, has seen increased interest in recent years, as it emerges as a focus of the high performance computing community for a variety of reasons. Tasks in statistical inference that are computationally challenging are therefore likely to benefit from the use of many-core architectures in their implementation, and attention is being paid to this area [4,5]. This architecture, however, requires a degree of specific types of parallelization to be exhibited by algorithms, which can motivate the development of new algorithms and models that are amenable to parallel implementation.

This project will investigate the use of many-core computation for the purpose of inference regarding biological networks. The specific focus will be networks rooted in dynamical models of signaling (see e.g. [3] for an example). Ongoing work in the Mukherjee group focuses on dynamical approaches for network inference in cancer, and a successful project will contribute to these efforts. Good computational skills and an interest in inference are required, but *no "wet" experimental work is involved and no prior knowledge of biology is needed.*

The project offers the opportunity to spend time (part or all of the mini-project) visiting 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, with the potential to exploit new computational ideas to address fundamental questions in biology.

### References

- [1] N. Friedman. Inferring Cellular Networks Using Probabilistic Graphical Models. *Science*, 303(5659):799-805, 2004.
- [2] C. J. Oates & S. Mukherjee. Network inference and biological dynamics. *Annals of Applied Statistics*, to appear (<http://arxiv.org/abs/1112.1047>)
- [3] T. R. Xu *et al.* Inferring signaling pathway topologies from multiple perturbation measurements of specific biochemical species, *Science Signaling*, 3(113), ra20, 2010
- [4] A. Lee *et al.* On the utility of graphics cards to perform massively parallel simulation of advanced Monte Carlo methods. *Journal of Computational and Graphical Statistics* 19(4): 769-789, 2010
- [5] Y. Zhou *et al.* GPU accelerated biochemical network simulation. *Bioinformatics* 27(6):874-6, 2011