Statistical analysis in tree-space
Dr. Tom Nye
(Universität Newcastle)

December 9, 2015

Data sets consisting of samples of evolutionary trees, or phylogenies, for some fixed set of species arise in many different contexts in evolutionary biology. Analysing such data sets is challenging, since the space of all phylogenetic trees is highly non-Euclidean, although it has a natural geodesic metric. It is very difficult to construct tractable distributions on tree space, so most analyses have relied on minimising sums of squared geodesic distances. Here I propose an alternative approach by constructing distributions via random walks in tree-space and their continuous-time limits. We consider particles undergoing Brownian motion in tree-space from some fixed starting point, and their distribution at subsequent times. I will then describe a Bayesian method for fitting distributions to samples of trees. This uses an Markov chain Monte Carlo scheme which relies heavily on a bridge construction between given start and end points for a random walk. These methods open up the possibility of fitting a variety of different models based on Brownian motion to samples of trees.