

3 The Alpha Model

Supervisor: Matthias Winkel

Prerequisites: Some of the possible directions would benefit from B10a Martingales through measure theory. Part A Graph Theory useful but not essential.

Type: Project with possibilities for theoretical study, simulation and probably fitting models to data

Description: The Alpha Model was proposed by Daniel Ford as a model for phylogenetic trees that describe the genealogical tree relating species. Ford fitted the model to published phylogenies as available on the internet in a database called TreeBASE.

An Alpha Model tree is a random rooted binary combinatorial tree T_n with n leaves (cladogram) constructed inductively as follows. For $n = 1$ and $n = 2$ there is only one such tree, looking like an I and a Y , respectively, the vertex at the bottom being the root, vertices at the top being called leaves. Given T_n , a tree T_{n+1} is constructed as follows. For some parameter $\alpha \in (0, 1)$, associate weights with edges, weight $1 - \alpha$ each with edges between a leaf and another vertex, weight α each with all other edges. Pick an edge at random with probability mass function proportional to the weights. Add two vertices. Replace the chosen edge by two edges connecting the end points to one of the new vertices and add a further edge between the new vertices.

The plan of this project is to carefully set up the model, study its properties, possibly simulate and/or fit the model to some data. There is a wide range of possibilities on the theoretical side including combinatorial and asymptotic properties, with some recent literature, to a large part at an accessible level, challenging in other parts. The data analysis side would repeat Ford's study on some more trees and compare with other models such as Aldous's beta-splitting model; further questions will arise while doing this.

References:

- D. Aldous: Probability distributions on cladograms. In *Random discrete structures (Minneapolis, MN, 1993)*, volume 76 of *IMA Vol. Math. Appli.*, pages 1017. Springer, New York, 1996.
- D. Ford: Probabilities on cladograms: introduction to the alpha model. 2005. *Preprint*, arXiv:math.PR/0511246
- J. Pitman: *Combinatorial Stochastic Processes*. Ecole d'été de Probabilités de St-Flour XXXII, to appear in *Lecture Notes in Mathematics*, Springer 2006