Tree of Life theory

Two Bachelor Project Proposals LIACS

This is a twin project consisting of the following two projects:

Project I. Tree of Life theory. Visualisation

Project II. Tree of Life theory. Algorithm development

Background & motivation

Darwin sometimes illustrated his 'origin of species' theory by sketching a tree of life (Fig 1).

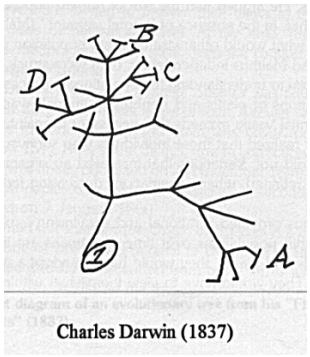

Figure 1

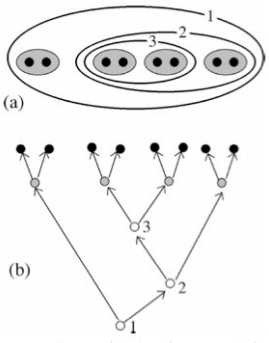

Figure 2. Hennig (1950)

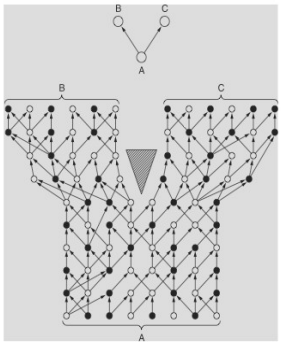

Figure 3. Hennig (1950)

A century later Hennig gave (as foundation of the hierarchical classification of species) a graph theoretical formalization of the tree of life (Fig 2), implying speciation as splitting events in the genealogical network (Fig 3). Kornet (1993) formalized the implied concept of species as equivalence classes of organisms. This work is illustrated in Fig. 4 showing a manually designed model of a genealogical network, its partitioning in species, and the implied tree of life (Kornet, 2012). The next step to take is to develop software for generating genealogical networks and for partitioning them into species, thus completing the tree of life theory.

Recently, a LIACS group headed by Kok, started collaboration with Kornet et al. in this field of bioinformatics, resulting in an interdisciplinary working group called *Tree of Life*.

Aim of the projects.

Project I. The goal is to build software to visualize and manipulate genealogical networks. The user should be able, with help of the software, to design alternative genealogical networks. Furthermore software should be developed to automatically generate such networks (subject to user definable constraints). The internal representation of these networks should allow to show the effects of the
algorithms to be developed in Project II.

Project II. The goal is to construct, apply and test algorithms to reduce such networks to a tree of life by partitioning the network along the lines of Kornet et al's theory (see also Fig. 4).

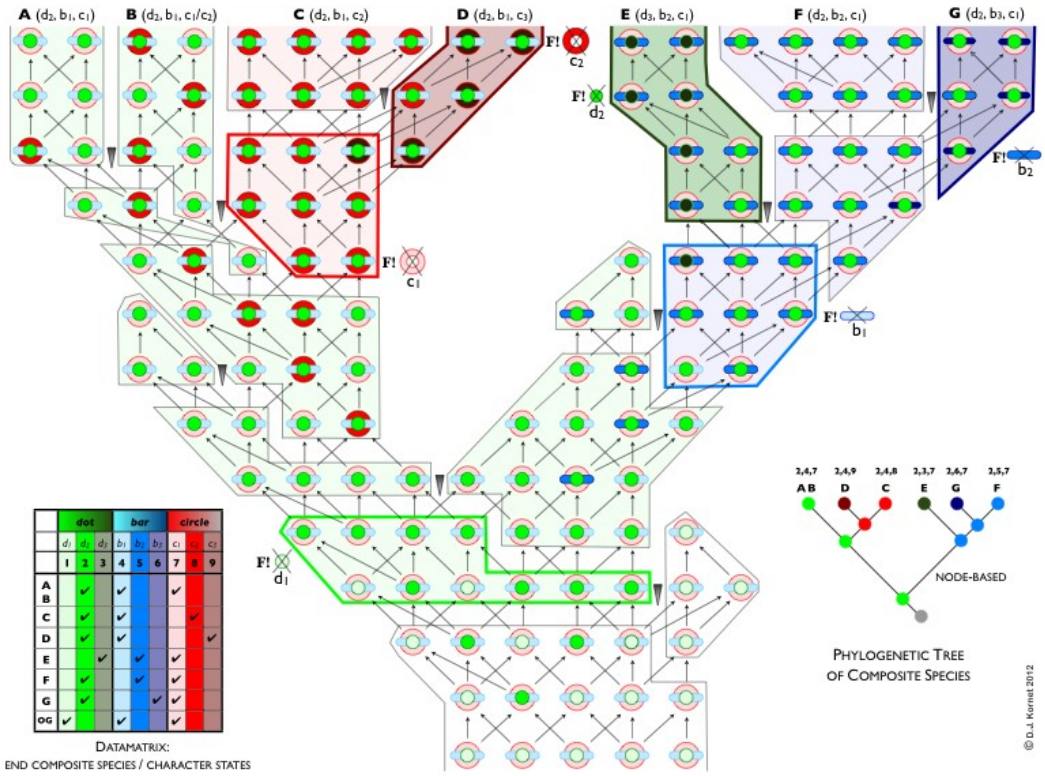

Figure 4. Kornet (2012)

Expected types of work

Projects I and II. Programming in C++, running experiments, software engineering, reading scientific literature both inside and outside computer science.
Project I. Building graphical user interfaces.
Project II. Developing and implementing algorithms, making mathematical proofs.

Context of the project

Projects I and II. Research papers are available that present the methods described above. The students are expected to participate in the LIACS working group Tree of Life.
Project I. Existing software: packages for building and manipulating graphs and graph visualization. Starting point of the work will be the network model of Fig.4.
Project II. One of the results of the working group is that we have discovered algorithms that implement the mathematical method partitioning networks like the one in Fig.4, resulting in a disjoint cover of the network as can be seen in the figure. Starting from an algorithm that was cubic in the number of nodes of the network we have trimmed it down to an almost linear one. The algorithm is based on a method published in a forthcoming research paper.

Required background.
Projects I and II. Software development in C++, software engineering.
Project I. User interface design.
Project II. Algorithmics and complexity theory.

Work plan.
The students in the twin projects will be closely cooperating.
Start of the project will be a literature study, and an implementation of the model network from Fig. 4. In *Project I* the focus will be on visualization aspects, in *Project II* the algorithmic side will be developed.
During the project the students will attend the working group. Requirement analysis will be done with members of the group.
From these requirements prototypes will be built. These prototypes will be used to experiment (e.g. in *Project II* to add and/or adapt the algorithms).
The projects will end up in a final presentation, and cooperation with the group members in a scientific publication is very well possible.

Names and email addresses:
A. de Bruin, arie2704@xs4all.nl,
D.J. Kornet, D.J.Kornet@biology.leidenuniv.nl
S. Nijssen, s.nijssen@liacs.leidenuniv.nl,
M. Zandee, rino.zandee@gmail.com