Developing a metric on phylogenetic trees
based on the concept of specific quartets

From a graph-theoretic viewpoint, trees are connected acyclic graphs (connected because there exists at least a path between any two vertices in the tree, and acyclic because for any two vertices such a path is unique). They are used in many domains of science or engineering, to represent a wide variety of things. In particular, in phylogenetics, trees provide for the common representation of evolutionary relationships between taxa (be they species or homologous genes or proteins, for instance). It is therefore important as well as very common to be able to compare two trees and yield some kind of distance measure.

Several tree distances have been proposed in the past, for instance the Robinson-Foulds metric (https://en.wikipedia.org/wiki/Robinson%E2%80%93Foulds_metric), the Maximum Agreement Subtree co-distance (http://arxiv.org/abs/1201.5168) or the quartet distance (https://en.wikipedia.org/wiki/Quartet_distance). Unfortunately, it can be proved that the quartet distance contains essentially some amount of redundancy, as some quartets of taxa agree (or "are defined by") several distinct branches. Interestingly, every single branch of the tree defines a non-empty set of quartets that are specific to it and shared with no other branch.

The aim of this project is to investigate on how to use this concept of specific quartets in order to compute efficiently the distance between two binary trees initially, and possibly between two non-binary trees eventually. If carried out seriously, this project will lead to a publication.

Project highlights: graph theory, algorithmics, combinatorics. Student willing to take on this must be both inventive and have strong algorithmics (and mathematics) skills. Full guidance and appropriate supervision will be provided by me, though.

For any queries or to apply, please contact me: Dr Jean-Baka Domelevo Entfellner (domelevo@gmail.com), CS dpt, room 1.24.