

Parameterised algorithms to compute distances between phylogenetic trees

Simone Linz

School of Computer Science, University of Auckland, Auckland, New Zealand

In evolutionary biology, phylogenetic trees are widely used to unravel the ancestral history of entities such as species, viruses, and cancer cells. A central challenge in the field of phylogenetics is to accurately infer such trees from molecular sequence data for a set of present-day species. In this context, it is not uncommon to obtain different trees for the same set of species. This can happen for methodological reasons or due to the fact that some species have multiple distinct tree signals woven into their genome. This motivates the use of distance measures in phylogenetics to quantify the dissimilarities between two phylogenetic trees. One popular distance, which is based on local rearrangement operations on a phylogenetic trees, is called the tree bisection and reconnection (TBR) distance. Although this distance is NP-hard to compute, it is also fixed-parameter tractable and computable in reasonable time for many practical instances. In this talk, we describe a series of results on the size of the TBR kernel, i.e. the size of two phylogenetic trees after pre-processing. We introduce novel data reduction rules that complement two existing rules to reduce the TBR kernel from $28k$ to $9k$, where k is the TBR distance. Along the way, we also explore the practical impact of the new data reduction rules and how they can be used to compute other distances between phylogenetic trees. The talk is based on ongoing joint work with Steven Kelk and Ruben Meuwese (Maastricht University).