
CAROLINE COLIJN, Simon Fraser University

The forests and the trees: new metrics on some flavours of trees

Trees are discrete, tractable objects that arise in many applications, from biology to machine learning. These naturally give rise to several distinct kinds of trees: trees labelled at the tips, trees labelled uniquely or non-uniquely at all nodes, trees labelled at some nodes, unlabelled trees, bifurcating trees and so on. Comparing large-scale trees can help researchers to explore the dimension and shape of a suitably-defined space of trees, to characterize the behaviour of a simulation model or to analyze data. As the number of trees rises combinatorially with the number of leaves, for larger trees it makes sense to explore representing sets of trees with metrics and embedding them in simpler (Euclidean) spaces. We have developed metrics, in the sense of distance functions, on several of the above flavours of trees. For rooted, tip-labelled and partly-labelled trees, we embed the tree in a high-dimensional Euclidean space and use the Euclidean metric. Metrics on unlabelled trees are not easily obtained the same way; we describe two different approaches to comparing unlabelled trees. The second of these can be generalized to obtain polynomials that distinguish a quite general class of trees.