Reconstructibility of free trees from subtree size frequencies

Dénes Bartha and Péter Burcsi

Department of Computer Algebra, Eötvös Loránd University, Budapest, Hungary

denesb@gmail.com, bupe@compalg.inf.elte.hu

Let T be a tree on n vertices. The subtree frequency vector (STF-vector) of T, denoted by $\mathrm{stf}(T)$ is a vector of length n whose kth coordinate is the number of subtrees of T that have exactly k vertices. For example, if P_5 denotes a path of length 5 and S_4 a star with 4 leaves, then we have $\mathrm{stf}(P_5) = [6, 5, 4, 3, 2, 1]$ and $\mathrm{stf}(S_4) = [5, 4, 6, 4, 1]$. We present algorithms for calculating the subtree frequencies: a combinatorial one and an algorithm using generating polynomials. We give a combinatorial interpretation for the first few and last few entries of the STF-vector.

The main question we investigate – motivated by the problem of determining molecule structure from mass spectrometry data – is whether T can be reconstructed from $\mathrm{stf}(T)$. This problem falls in the broad family of combinatorial reconstruction problems. We show that there exist examples of non-isomorphic pairs of free (i.e. unlabled, unrooted) trees that are STF-equivalent, i.e. have identical subtree frequency vectors. Using exhaustive computer search, we determine all such pairs for small sizes. We show that there are infinitely many non-isomorphic STF-equivalent pairs of trees by constructing infinite families of examples. We also show that for special kinds of trees (e.g. paths, stars and trees containing a single vertex of degree larger than 2), the tree is reconstructible from the subtree frequencies.

We consider a version of the problem for rooted trees, where only subtrees containing the root are counted. We also show examples of equivalent pairs in this sense. Finally, we formulate some conjectures and open problems and outline further research directions.