

Maximum Rooted Triplets Consistency

Jesper Jansson
Kyoto University, Japan
jj@kuicr.kyoto-u.ac.jp

2013-09-18
(updated on 2015-10-01)

In this note, we define a *phylogenetic tree* as a rooted, unordered, distinctly leaf-labeled tree in which every internal node has at least two children. A *rooted triplet* is a binary phylogenetic tree with exactly three leaves. The unique rooted triplet with leaf label set $\{x, y, z\}$ in which the lowest common ancestor (lca) of x and y is a proper descendant of the lca of x and z (or equivalently, where the lca of x and y is a proper descendant of the lca of y and z) is denoted by $xy|z$. For any rooted triplet $xy|z$ and phylogenetic tree T which includes three leaves labeled by x, y, z , if the lca in T of x and y is a proper descendant of the lca of x and z , then $xy|z$ and T are said to be *consistent* with each other; otherwise, $xy|z$ and T are *inconsistent*. A set \mathcal{R} of rooted triplets is *consistent* if there exists a phylogenetic tree T such that every $xy|z \in \mathcal{R}$ is consistent with T . See Figure 1 for an example. Now, define the following optimization problem:

- **Maximum Rooted Triplets Consistency (MaxRTC):** Given a set \mathcal{R} of rooted triplets, output a phylogenetic tree that is consistent with as many rooted triplets from \mathcal{R} as possible.

For a survey on MaxRTC, see [2]. In short, MaxRTC is NP-hard and has a polynomial-time 3-approximation algorithm. The main open problem related to MaxRTC is:

Does MaxRTC admit a polynomial-time approximation algorithm whose worst-case approximation ratio is asymptotically better than 3?

Motivation: The phylogenetic tree is a 150-year-old data structure used by scientists and scholars to describe treelike evolutionary history [3]. For various reasons, inferring an accurate phylogenetic tree from experimental data can be a difficult task. For example, computationally expensive methods that yield good trees may be impractical for very large data sets. One potential remedy is the divide-and-conquer approach: first apply some expensive method to obtain a collection

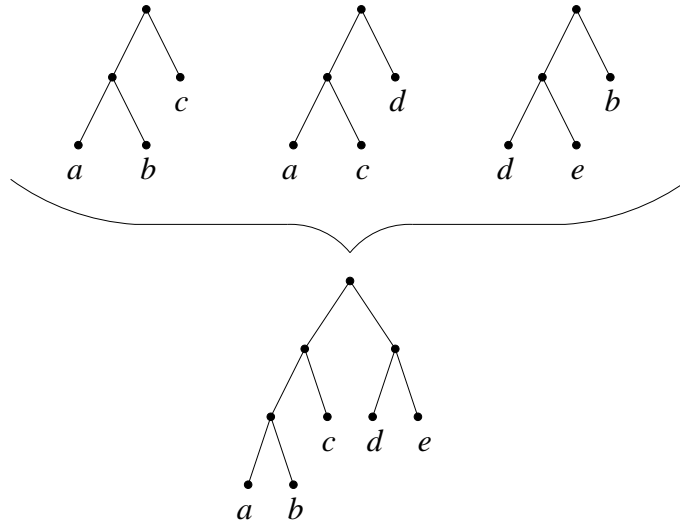


Figure 1: An example from [2]. The set $\mathcal{R}_1 = \{ab|c, ac|d, de|b\}$ of rooted triplets is consistent. However, it can be shown by contradiction that there exists no phylogenetic tree consistent with all four rooted triplets in the set $\mathcal{R}_2 = \{ab|c, ac|d, de|b, ce|b\}$.

of reliable trees for small, overlapping subsets of the leaf labels, and then use a computationally cheaper method to merge these small trees into a so-called *phylogenetic supertree*. A fundamental concept that captures the combinatorial aspects of the smallest meaningful building blocks of a phylogenetic supertree is *rooted triplets consistency*, described above.

In 1981, Aho *et al.* [1] presented an elegant algorithm for determining if an input set \mathcal{R} of rooted triplets is consistent. The algorithm runs in polynomial time and outputs a phylogenetic tree that is consistent with every rooted triplet in \mathcal{R} , if one exists. If \mathcal{R} is inconsistent, Aho *et al.*'s algorithm outputs *null*. However, in practical applications, when \mathcal{R} contains contradictory branching information due to errors in the experimental data, one might still want some kind of tree. It is therefore natural to consider optimization versions such as MaxRTC. More information on what is currently known about MaxRTC and other related problems can be found in [2] and [4].

References

- [1] A. V. Aho, Y. Sagiv, T. G. Szymanski, and J. D. Ullman. Inferring a tree from lowest common ancestors with an application to the optimization of relational expressions, *SIAM Journal on Computing*, Vol. 10, No. 3, pp. 405–421, 1981.
- [2] J. Byrka, S. Guillemot, and J. Jansson. New results on optimizing rooted triplets consistency, *Discrete Applied Mathematics*, Vol. 158, No. 11, pp. 1136–1147, 2010.
- [3] J. Felsenstein. *Inferring Phylogenies*, Sinauer Associates, Inc., Sunderland, 2004.
- [4] J. Jansson, A. Lingas, and E.-M. Lundell. The approximability of maximum rooted triplets consistency with fan triplets and forbidden triplets, Proceedings of CPM 2015, *Lecture Notes in Computer Science*, Vol. 9133, pp. 272–283, 2015.