Algorithmic Decomposition of Most Parsimonious Sets

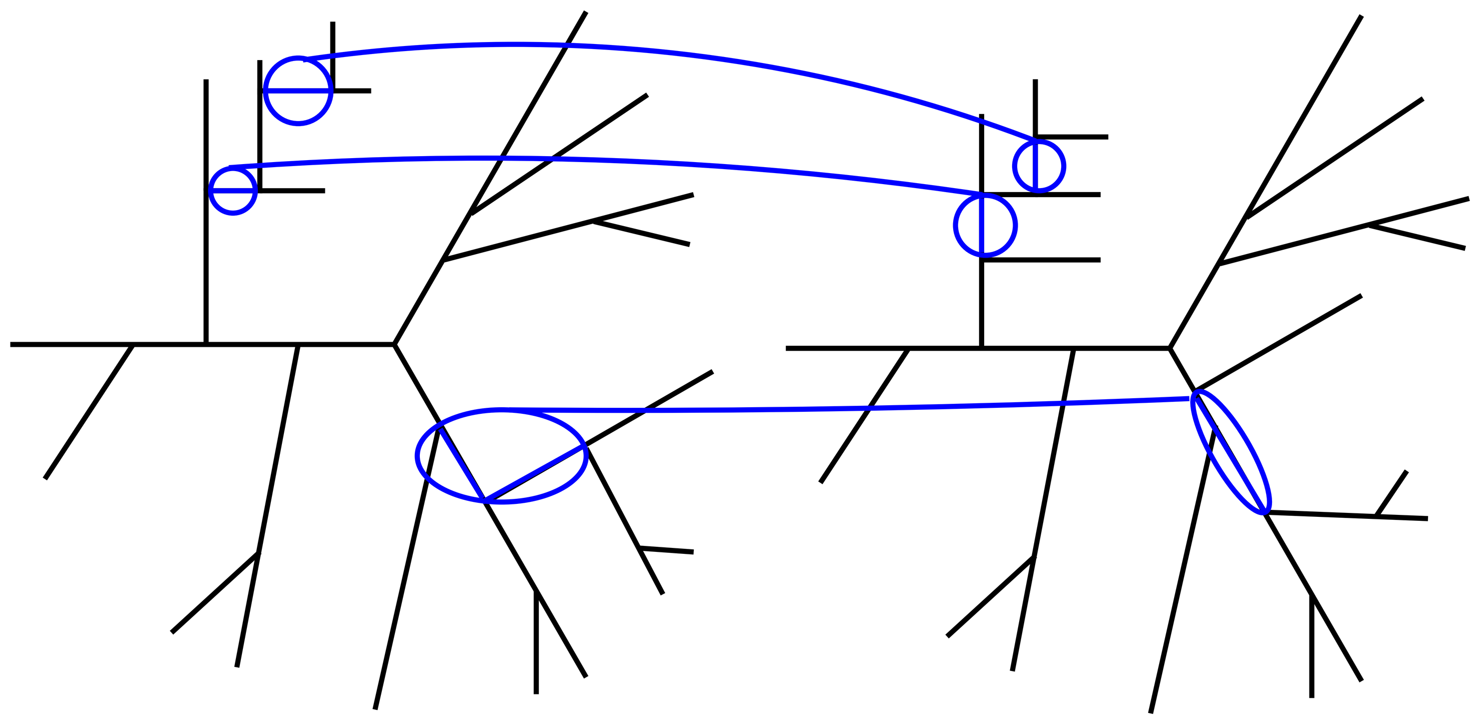
Author: Nicholas D. Pattengale
<npcomplete@gmail.com>

Advisor: Bernard M.E. Moret

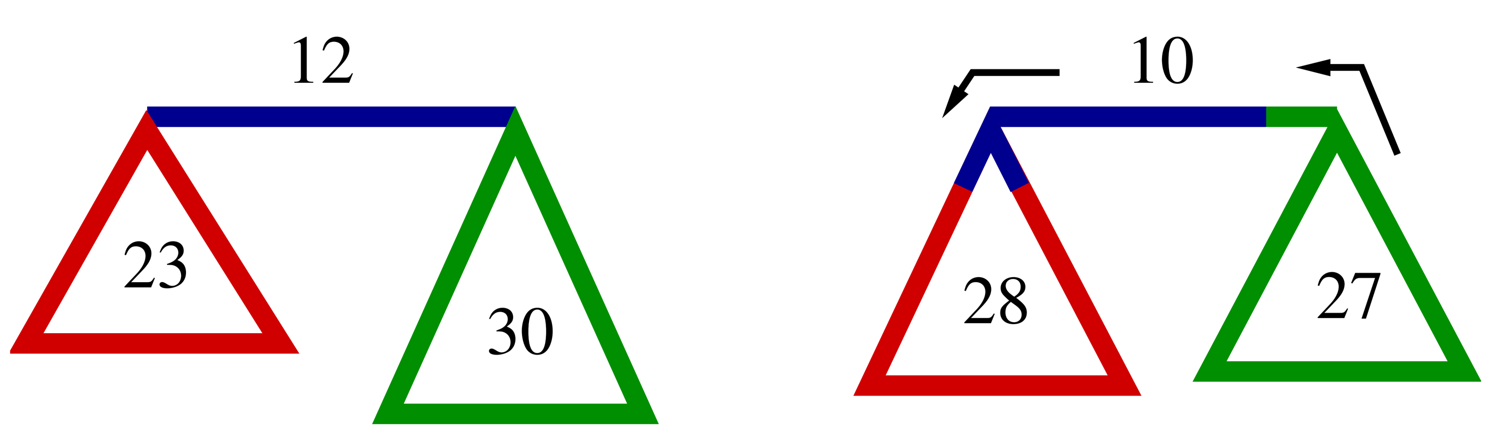
Department of Computer Science

The University of New Mexico

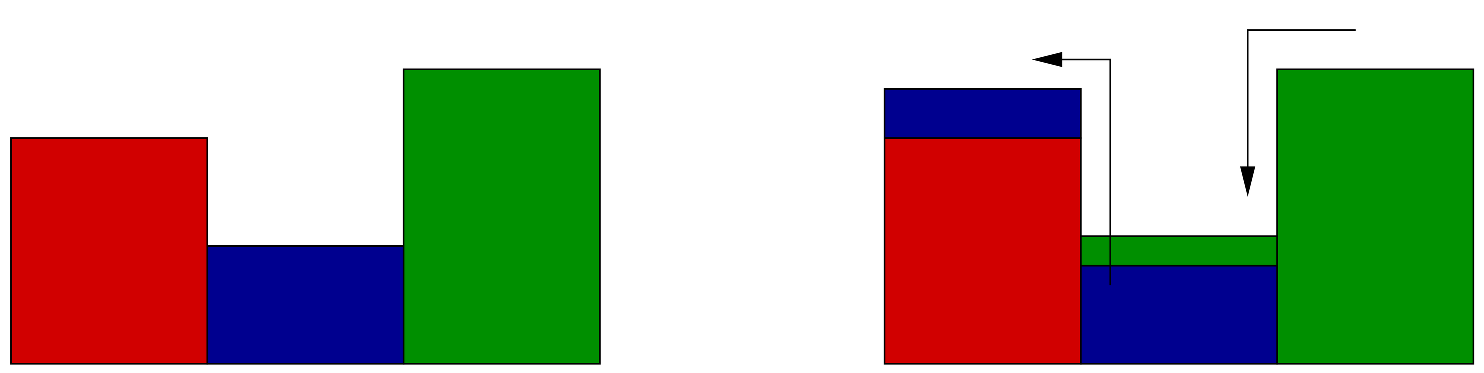
Concepts



Decomposition



Weight Exchange



Decomposition + Theorem

Theory

Decomposition

Input: Phylogenetic trees $S = (V, E_S, w_S)$ and $T = (V, E_T, w_T)$
Output: A function $g : \mathcal{P}(E_S) \leftrightarrow \mathcal{P}(E_T)$

1. $E^* \leftarrow E_S \cap E_T$
2. $E'_S \leftarrow E_S - E^*$
3. $E'_T \leftarrow E_T - E^*$
   S' and T' are now forests with equal cardinality in the number of subtrees.
4. map subtrees from S' to T' based on adjacency to E^*
5. also $\forall e \in E^*$, $g(\{e\}) \leftrightarrow \{e\}$

"Weight Exchange" Theorem

Theorem $\forall e^* \in E^*$,

$$w_S(e^*) - w_T(e^*) = \left(\sum_{t \in c(T^1_{e^*}, E^*)} \sum_{e \in E_t} w_T(e) - \sum_{t \in c(S^1_{e^*}, E^*)} \sum_{e \in E_t} w_S(e) \right) + \left(\sum_{t \in c(T^2_{e^*}, E^*)} \sum_{e \in E_t} w_T(e) - \sum_{t \in c(S^2_{e^*}, E^*)} \sum_{e \in E_t} w_S(e) \right)$$

This theorem arises because S and T have equal parsimony scores. Now observe that the formula commutes in the following fashion:

$$w_S(e^*) - w_T(e^*) = \sum_{t \in c(T^1_{e^*}, E^*)} \left(\sum_{e \in E_t} w_T(e) - \sum_{e \in E_{g^{-1}(t)}} w_S(e) \right) + \sum_{t \in c(T^2_{e^*}, E^*)} \left(\sum_{e \in E_t} w_T(e) - \sum_{e \in E_{g^{-1}(t)}} w_S(e) \right)$$

Decomposition + Theorem

We use the theory to algorithmically derive functional relationships between two trees with equal parsimony scores. The relationship characterizes for each subtree in T (as defined by the decomposition) the unique weighted set of subtrees from S that sum to the suitable weight.

Application

Comparison with Consensus Methods

Consensus methods are generally

- constrained to producing one tree
- not sensitive to edge weights

We produce sets of functions, not an output tree. Our method can be used to add edge-weight sensitivity to many traditional consensus methods.

Also notice that computing the decomposition resembles computing a strict consensus tree. We plan to investigate alternative decompositions that resemble other consensus methods.

Interpolating Stable Weight Distribution

How might we infer a representative tree? Use the output of our algorithm as a dissimilarity measure, $d(T_1, T_2)$, by taking the sum of squares. We propose the optimization problem:

- Given a set of phylogenetic trees $\{T_1, T_2, ..., T_n\}$ with homogeneous parsimony scores, return a phylogenetic tree T_R that minimizes $d(T_R, T_k)$ for all $1 \leq k \leq n$.

We strongly suspect that this problem is intractable and plan to investigate approximation algorithms.

Sets With Heterogeneous Scores

We propose two methods for handling sets with heterogeneous parsimony scores.

1. Add a constant-valued correction factor into the theorems such that the trees now have equal scores.
2. Normalize such that the edge lengths always sum to one.

We can carry a positive correction factor into the algorithm by depositing it as spontaneously arising weight in the second tree. Alternatively, we can carry a negative correction factor *mutatis mutandis*.