MATH 239: DISCRETE MATH FOR THE LIFE SCIENCES
FINAL PROJECT - SPRING 2008
TROPICAL MIXTURES OF TREES OF THE SAME TOPOLOGICAL TYPE

MARIA ANGELICA CUETO

ABSTRACT. In this paper we analyze "tropical" mixtures of tree metrics of the same topological type. Although the resulting dissimilarity map need not be a tree metric (in fact, it need not be either tree additive) an interesting question is to characterize algebraically in which cases we get a tree metric of the same topology as the mixed ones and in which cases we get a different one. We also discuss about the geometric structure of these mixtures and we present several examples showing that unexpected results may appear in practice.

1. INTRODUCTION

In the present paper we study weighted trees on four taxa and the tropical mixture of two trees of the same topological type. This question was inspired by a recent work by E. Matsen and M Steel [2]. In that article, they study phylogenetic mixtures of two trees of the same topological type, and they characterize in which situations they obtain a tree of a different topology. For their work, they consider pylogenetic mixtures of trees, i.e. convex combinations of two tree metrics. Namely, a weighted average of site pattern frequencies derived from two phylogenetic trees. The importance of this concept relies on the fact that phylogenetic mixtures model the inhomogeneous molecular evolution commonly observed in experimental data.

Their motivation for studying this situation was that in certain parameter regimes, phylogenetic reconstruction methods tend to return a tree topology of a different type than the ones used to generate the mixture data. In particular, they concentrate on a phenomenon called *branch repulsion*, that correspond to the case where two neighboring pendant edges which alternate being long and short. After mixing the two trees, these edges may fail to be adjacent in the new tree. They also discuss this point to show that some common assumptions made for reconstructing methods may not hold. Namely, it is not correct to assume that a mixture of trees of one topological type gives site pattern frequencies than that of a (unmixed) tree of a different topology.

In our case, rather than being interested in phylogenetic mixtures we focus our attention on tropical mixtures. This concept is based on mixtures of probabilities. Suppose we are mixing two tree metrics $D^{(1)}$ and $D^{(2)}$ (or in general any finite number of tree metrics). As stated at the end of Chapter III in [1] (page 122), we can consider each tree metric $D^{(\nu)}$ as a random variable $X^{(\nu)}$ on the pairs of taxa, with probability distribution

$$\text{Prob}(X^{(\nu)} = \{i,j\}) \sim exp(-\tau D^{(\nu)}_{ij}), \tag{1}$$

where τ denotes the time parameter (i.e. we consider a multivalued Poisson distribution for $X^{(\nu)}$). Assume we have r random variables $X^{(\nu)}$, with mixing probabilities $p_1, \dots, p_r$ (i.e. $p_i \in [0,1]$ and $\sum_{i=0}^{r} p_i = 1$). Assume, moreover, that all $p_i > 0$. We want to analyze the behaviour of the mixture of these random variables for $\tau \gg 0$. Therefore, if we call X the random variable resulting from mixing all $X^{(\nu)}$ together, we get

$$\text{Prob}(X = \{i,j\}) = \sum_{\nu=1}^{r} p_\nu \text{Prob}(X^{(\nu)} = \{i,j\}) \sim \sum_{\nu=1}^{r} p_\nu exp(-\tau D^{(\nu)}_{ij}).$$

Assume that for all $i \neq j$ $D_{ij}^{(\nu)} \neq 0$ (we know that they are non-negative real numbers). Thus,

$$(2)\quad \sum_{\nu=1}^{r} p_\nu exp(-\tau D_{ij}^{(\nu)}) - exp(-\tau \max\{D_{ij}^{(\nu)} : \nu\}) = \sum_{\nu=1}^{r} p_\nu (exp(-\tau D_{ij}^{(\nu)}) - exp(-\tau \max\{D_{ij}^{(\nu)} : \nu\})) \longrightarrow 0$$

as $\tau \to \infty$. Thus, under this hypothesis (that hold for generic tree metrics) for $\tau \gg 0$ we get

$$\text{Prob}(X = \{i,j\}) \sim \sum_{\nu=1}^{r} p_\nu exp(-\tau D_{ij}^{(\nu)}) \sim exp(-\tau \max\{D_{ij}^{(\nu)} : \nu\}) = exp(-\tau \bigoplus_{\nu=1}^{r} D_{ij}^{(\nu)})$$

for all $i \neq j$, where $\oplus$ is the tropical addition in the *max-plus-algebra*. In case $i = j$ we know by definition of a dissimilarity map that $D_{ii}^{(\nu)} = 0$. Thus $\text{Prob}(X = \{i,i\}) \sim 1 = exp(-\tau \bigoplus_{\nu=1}^{r} D_{ii}^{(\nu)})$. Therefore, we see that tropical mixture arises in a natural way in this setting.

We should also mention that the distribution chosen in (1) and the approximation in (2) comes from the LogDet distances and the logdet transform defined in [3] in the context of Markov models on trees. We refer the interested reader to Theorem 8.4.3 for more details on this subject.

Let us go back to the treatment in [2]. To approach the characterization of mixed branch repulsion, the authors point out that this notion has a relative nature. Therefore, any algebraic characterization must rely on the difference between the edge weights of corresponding edges of each tree, instead of absolute branch lengths themselves. In addition, they rely on two of the main tools in theoretical phylogenetics: the Hadamard transform and phylogenetic invariants.

Inspired by this observation and by the four-point condition that is an algebraic characterization of tree metrics, we also approach the algebraic characterization of "tropical" mixtures of trees of the same topological type in a relative way. However, since we are provided with tree metrics we shouldn't expect to express everything in terms of the branch lenghts. Instead, we characterize the mixture in terms of the differences between the entries of the tree metrics, and the weight of the inner edge of one of the trees we're mixing together. On the other hand, since the tropical mixture is defined in terms of the tree metrics, we should always consider the entries of the matrices above the branch lenghts of the two trees.

In the next section we discuss the algebraic characterization of mixtures of trees on four taxa of the same topological type. We also discuss the geometric structure of the corresponding pairs of trees according to the resulting mixed topology. In particular, we discuss the convexity of this set. Finally, in Section §4 we provide a wide range of examples showing that the behaviour of mixtures is far from being predictable beforehand.

2. Dissimilarity maps, tree metrics and topological types

In the present and forthcoming sections we will assume all trees to be laminar families and, moreover, all labeled nodes in our trees must be leaves. This is not a restriction at all, since if we have a general laminar family with weighted edges and we have a labeled node, say x, with degree ≥ 3, we can attach an edge of weight *zero* to this node and label the corresponding leaf by x, removing the label on the internal node. As a consequence of this property, given a finite number of taxa X we have a finite number of trees with their leaves labeled by X and all internal nodes of degree ≥ 3. Therefore, when analyzing trees, we only have to deal with a finite number of posible tree topologies and labelings.

We now analyze some properties of dissimilarity maps and metrics on the taxa X. For basic definitions, we refer the reader to [4].

Remark 2.1. *An important remark to make is the following. If we have two metrics H, $\bar{H}$ we have that $H \vee \bar{H}$ is also a metric (i.e. $D_{ii} = 0$ for all i, $D_{ij} = D_{ji}$ for all i, j and the triangular inequality $D_{ij} \leq D_{ik} + D_{kj}$ holds). The only nontrivial point to check would be the triangular inequality, but this will be an immediate consequence of the triangular inequality for H and $\bar{H}$ plus the definition of a maximum between two real numbers.*

In addition to this, we may warn the reader on one point: a mixture of tree additive dissimilarity maps with respect to a fixed tree T need not be tree additive w.r.t. the same T or w.r.t. any other tree. Recall that the property of being tree additive can be checked with the weak four-point conditions. Moreover, by tautology, in the generic case, the result of such mixture won't be tree additive. In a similar way, mixing two generic tree metrics of the same or different topological type won't give a tree metric, nor even a tree additive dissimilarity map. We will see some examples of this in Section §4.

One final remark to make is the following. Trees can be described completely in many different ways. For example, we can present it as a set of nodes and edges (like any graph), or as the set of splitings on the n taxa. In this case, we choose to use another representation, by means of a set of quartets. There are several works in the literature that deal with the minimal number of quartets needed to describe completely a tree. In this case, we won't be interested in minimizing this number. Moreover, we will assume we are given the complete list of quartets. This representation will allow us to reduce our characterization of mixtures of trees on four taxa.

Definition 2.2. *Given a set* $X = \{1, \ldots, n\}$ *we define a set* $\mathscr{S} \subset \{(ij|kl) : i, j, k, l \in X, i < j, k < l, i < k\}$ *to be a* complete set of quartets *on* X *iff for all four different indices* $i < j < k < l$ *in* X *we have* $S \cap \{(ij|kl), (ik|jl), (il|jk)\} \neq \emptyset$.

Alternatively, we can relax our restrictions on a quartet by defining an equivalence relation on quartets as $(ij|kl) \sim (ij|lk) \sim (ji|lk) \sim (ji|kl) \sim (kl|ij) \sim (kl|ji) \sim (lk|ij) \sim (lk|ji)$.

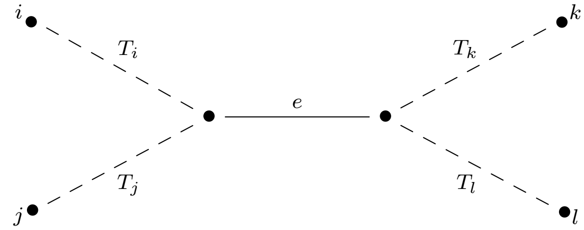
Proposition 2.3. *A complete set of quartets on a tree determines the topology on* T*, i.e. the map* $T \mapsto$*(complete set of quartets on* T*) is injective.*

Proof. By induction on the number of taxa ($n = |X|$). The base case will be $n = 4$. In this case quartets on T are splittings on T, so a complete list of quartets is given by a single quartet $(ij|kl)$ (if we have a tree with six nodes, four leaves and a unique inner node as in Figure 1, in case $i = 1, j = 2, k = 3, l = 4$), or by more that one quartet (if we have a tree on four taxa with the star topology).

Assume the result holds for $n > 1$ and consider a tree on $n + 1$ taxa. W.l.o.g., we may assume that $n + 1$ is a leaf in T. Consider the leaf labeled by $n + 1$ and the corresponding adjacent node. Since $n > 4$, we know that the adjacent node is an internal node (thus of degree ≥ 3) and so it must be unlabeled. If its degree is > 3, then by removing the edge corresponding to $n + 1$ we still get an honest tree (it will still be a tree satisfying the labeling restrictions). So we remove the leaf $n + 1$ and the adjacent edge. On the other hand, if the degree of this node is exactly three we need to perform a different operation. In this case, we remove the leaf $n + 1$, the adjacent edge and we delete also the inner node, glueing together the remaining two internal edges adjacent to this inner node. Call T' the resulting tree.

By inductive hypothesis, T' is determined uniquely by the subset of quartets of T not involving the index $n + 1$. To finish, we only need to show that the set of quartets $\mathscr{S}$ will allow us to reconstruct T in a unique way, i.e. they will give a unique way to attach the leaf $n + 1$ to T' to recover T.

Assume T' is not a star tree. Since we have $n > 4$, on X' we must have four indices s.t. T' has four leaves labeled i, j, k, l s.t. and $(ij|kl) \in \mathscr{S}$. Thus, T' is of the form:



where the dashed edges correspond to subtrees of T' and e is an honest edge in T'. We want to know where the leaf $n+1$ should be attached. We can attach it to a node in T' or to an edge in T', in which case we simply add an internal node to the edge and we attach the leaf $n + 1$ to this node, in correspondence with our construction of T' from the original tree T.

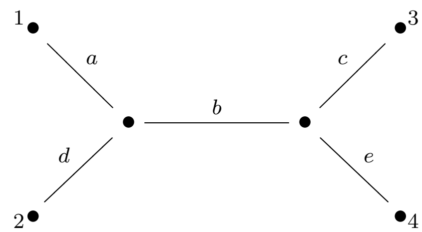


FIGURE 1. Tree on four taxa corresponding to the quartet (12|34)

The answer to this will be given by $\mathscr{S}$. We have five cases to analyze, but by symmetry we can reduce to two cases.

The first case corresponds to $\mathscr{S} \cap \{(i(n+1)|jk), (j(n+1)|ik), (k(n+1)|lj), (l(n+1)|jk)\} \neq \emptyset$. W.l.o.g., assume $(i(n+1)|jk) \in \mathscr{S}$. Then we know that the edge $n+1$ must be attached in the tree T_1. In this case, we have $(ik|j(n+1)) \in \mathscr{S}$ iff the leaf $n+1$ should be attached at the (LHS) vertex of edge e. If $(j(n+1)|ik) \notin \mathscr{S}$, then the leaf will be attached in T_1 but not on this vertex. Assume the latter. Since the number of taxa n_1 involved in T_1 is less than $n-3$, we need only to analyze the cases $n_1 = 1, 2, 3$ or $n_1 \geq 4$. This last case will be completely determined by the inductive hypothesis. Concerning the case $n_1 = 1$, we have only one choice since T_1 consists of one edge with two nodes. If $n_1 = 2$ or 3, by working with quartets involving all taxa in T_1 and j, k we can determine how to attach $n+1$ to T_1 (i.e., we reduce to the case of four taxa and how to add one leaf to this subtree).

Moreover, we have that the leaf $n+1$ must be attached to the tree T_s iff $(s(n+1)|\_\_) \in \mathscr{S}$ (for $s = i, j, k$ or l).

On the contrary, if $\mathscr{S} \cap \{(i(n+1)|jk), (ik|j(n+1)), (|j|k(n+1)), (jk|l(n+1))\} = \emptyset$, there is only one thing we can do: attach the leaf $n+1$ to the edge e by adding an internal node to e.

Thus, by construction there is a unique way of reconstructing T by means of T' and the quartets in $\mathscr{S}$. Since T' is completely determined by $\mathscr{S}$, the result follows. □

3. MIXTURES OF TREES

We will concentrate our attention on dissimilarity maps on four taxa $X = \{1, 2, 3, 4\}$. This will suffice our needs, since the tree topology for arbitrary number of taxa will be determined by quartets, as we explained in the previous section. And, in addition the property of being a tree metric is determined by the four-point condition. For simplicity, after relabelling if necessary, we will assume that our tree metrics H and $\overline{H}$ correspond to trees of topological type (12|34). For the moment, we will consider the star topology as a different case, although all the computations for this special case will follow from the previous one by considering weight zero in the middle edge.

Let us label our edge weights as follows. We will use variables a, b, c, d, e for H and $\bar{a}, \bar{b}, \bar{c}, \bar{d}, \bar{e}$ for the corresponding edges in $\bar{H}$, as in Figure 1.

In this way, our dissimilarity maps denote again by H and $\bar{H}$ are given by two 4×4 symmetric matrices with 0 entries in the diagonal, and with rows and columns labeled 1 through 4 in increasing order. For simplicity we will only show the values of the coefficientes above the diagonal (i.e. a_{ij} with $i < j$):

$$H := \begin{pmatrix} 0 & a+d & a+b+c & a+b+e \\ & 0 & d+b+c & d+b+e \\ & & 0 & c+e \\ & & & 0 \end{pmatrix}$$

$$\bar{H} := \begin{pmatrix} 0 & \bar{a}+\bar{d} & \bar{a}+\bar{b}+\bar{c} & \bar{a}+\bar{b}+\bar{e} \\ & 0 & \bar{d}+\bar{b}+\bar{c} & \bar{d}+\bar{b}+\bar{e} \\ & & 0 & \bar{c}+\bar{e} \\ & & & 0 \end{pmatrix}$$

As we said before, the mixture of H and $\bar{H}$ will correspond to taking the maximum of the corresponding entries on each matrix. Therefore if $D = H \vee \bar{H}$ is the resulting mixture, we get the upper right triangular component of D as follows

$$D_{12} = \max\{a+d, \bar{a}+\bar{d}\}\ ;\ D_{13} = \max\{a+b+c, \bar{a}+\bar{b}+\bar{c}\}\ ;\ D_{14} = \max\{a+b+e, \bar{a}+\bar{b}+\bar{e}\}\ ;$$

$$D_{23} = \max\{d+b+c, \bar{d}+\bar{b}+\bar{c}\}\ ;\ D_{24} = \max\{d+b+e, \bar{d}+\bar{b}+\bar{e}\}\ ;\ D_{34} = \max\{c+e, \bar{c}+\bar{e}\}\ .$$

As we said in the introduction, to determine if D is a tree metric it suffices to check the four-point condition. Since we have only four taxa, and we already know that D is a metric (i.e. it satisfies the triangular inequality), we need only the check the four point condition for our four indices $1, 2, 3$ and 4. This gives:

$$(*) \qquad \begin{cases} D_{12} + D_{34} &= \max\{a+d, \bar{a}+\bar{d}\} + \max\{c+e, \bar{c}+\bar{e}\} \\ D_{13} + D_{24} &= \max\{a+b+c, \bar{a}+\bar{b}+\bar{c}\} + \max\{d+b+e, \bar{d}+\bar{b}+\bar{e}\} \\ D_{14} + D_{23} &= \max\{a+b+e, \bar{a}+\bar{b}+\bar{e}\} + \max\{d+b+c, \bar{d}+\bar{b}+\bar{c}\} \end{cases}$$

Our goal will be to determine if the four point condition holds, and if so, determine the topological type of the underlying tree.

Proposition 3.1. *Let $D \in \mathbb{R}^{4\times 4}_{\geq 0}$ be a tree metric. Then the topological type of the corresponding tree T is determined as follows*

(i) $D_{14} + D_{23} = D_{13} + D_{24} > D_{12} + D_{34} \iff$ *T has topology* $(12|34)$,
(ii) $D_{12} + D_{34} = D_{13} + D_{24} > D_{14} + D_{23} \iff$ *T has topology* $(14|23)$,
(iii) $D_{12} + D_{34} = D_{14} + D_{23} > D_{13} + D_{24} \iff$ *T has topology* $(13|24)$,
(iv) $D_{12} + D_{34} = D_{13} + D_{24} = D_{14} + D_{23} \iff$ *T has the star topology.*

More concretely, we observe that the topology of T is determined by the pairing of the indices of $\min\{D_{12} + D34, D_{13} + D_{24}, D_{14} + D_{23}\}$. In case the three quantities are equal, the tree has topology $(12|34) = (13|24) = (14|23)$, i.e. the star topology.

Proof. It follows immediately from the previous observation, since the inner edge has non-negative weight, and it has weight zero iff T has the star topology □

Therefore, to determine the topology of D (in case it is a tree metric) we need to compare these quantities. Notice that we are given two tree metrics H, $\bar{H}$ with the same underlying tree topology (which we assume is $(12|34)$), so we don't know the weights of each one of the ten edges. We would like to derive an algebraic method for characterizing the topology of the mixture tree D by means of the entries of H and $\bar{H}$.

We will proceed by means of a very simple observation, although very useful. We will only need to deal with the differences among the entries H_{ij} and $\bar{H}_{ij}$ for $i < j$. Consider six new indeterminates s, t, x, y, u, w and the following equations relating the entries in H and $\bar{H}$.

$$(**) \qquad \begin{cases} \bar{a}+\bar{b}+\bar{c} &= a+b+c+s \\ \bar{d}+\bar{b}+\bar{e} &= d+b+e+t \\ \bar{a}+\bar{b}+\bar{e} &= a+b+e+x \\ \bar{d}+\bar{b}+\bar{c} &= d+b+c+y \\ \bar{a}+\bar{d} &= a+d+u \\ \bar{c}+\bar{e} &= c+e+w \end{cases}$$

Note that if we have two number $n_1, n_2 \in \mathbb{R}$, then $\max\{n_1, n_2\} = n_1 + \max\{n_2 - n_1, 0\}$. Therefore, we can rewrite the conditions (*) in terms of the entries in H and the new variables s, t, x, y, u, w as follows:

$$\begin{cases} D_{12} + D_{34} &= a+d+\max\{u,0\}+c+e+\max\{w,0\} \\ D_{13} + D_{24} &= a+b+c+\max\{s,0\}+d+b+e+\max\{t,0\} \\ D_{14} + D_{23} &= a+b+e+\max\{x,0\}+d+b+c+\max\{y,0\} \end{cases}$$

Notice that the summand $a+c+d+e$ is present in the three equations above. Since we need to compare these three expressions, we can substract $a+c+d+e$ without changing the order relation. Thus, we get

$$(***)\qquad \begin{cases} \tilde{D}_{12}+\tilde{D}_{34} &= \max\{u,0\}+\max\{w,0\} \\ \tilde{D}_{13}+\tilde{D}_{24} &= 2b+\max\{s,0\}+\max\{t,0\} \\ \tilde{D}_{14}+\tilde{D}_{23} &= 2b+\max\{x,0\}+\max\{y,0\} \end{cases}$$

Notice that in the previous equations, we are only dealing with the differences $\bar{H}_{ij}-H_{ij}$ for $i<j$ and also with the weight of the inner node in the tree H. However, this quantity is not so difficult to obtain since we have the following formula:

$$b = \frac{1}{2}(H_{13}+H_{24}-(H_{12}+H_{34})).$$

Therefore, we can characterize the tree topology of $H\vee\bar{H}$ by rewriting Proposition 3.1 in terms of the new variables:

Proposition 3.2. *Consider $H,\bar{H}\in\mathbb{R}^{4\times 4}_{\geq 0}$ be tree metric of the same topology. Consider the expressions in (***). Then the $H\vee\bar{H}$ is a tree metric iff it satisfies the four-point condition. Moreover, topological type of the corresponding tree T given by the mixture $H\vee\bar{H}$ is determined as follows:*

(i) $\tilde{D}_{14}+\tilde{D}_{23}=\tilde{D}_{13}+\tilde{D}_{24}>\tilde{D}_{12}+\tilde{D}_{34} \iff T$ *has topology* $(12|34)$,
(ii) $\tilde{D}_{12}+\tilde{D}_{34}=\tilde{D}_{13}+\tilde{D}_{24}>\tilde{D}_{14}+\tilde{D}_{23} \iff T$ *has topology* $(14|23)$,
(iii) $\tilde{D}_{12}+\tilde{D}_{34}=\tilde{D}_{14}+\tilde{D}_{23}>\tilde{D}_{13}+\tilde{D}_{24} \iff T$ *has topology* $(13|24)$,
(iv) $\tilde{D}_{12}+\tilde{D}_{34}=\tilde{D}_{13}+\tilde{D}_{24}=\tilde{D}_{14}+\tilde{D}_{23} \iff T$ *has the star topology.*

Namely, the topology is determined by the quartet $(ij|kl)$ realizing the minimum value $\min\{\tilde{D}_{12}+\tilde{D}_{34},\tilde{D}_{13}+\tilde{D}_{24},\tilde{D}_{14}+\tilde{D}_{23}\}$.

Therefore, by replacing the variables by their expressions in terms of the entries of H and $\bar{H}$ we obtain:

Theorem 3.3. *With the previous notation, consider $D=H\vee\bar{H}$ the mixture of two trees on four taxa of the same topology $(12|34)$ or the star topology. Consider the following seven expressions:*

$$\begin{cases} b &= \frac{1}{2}(H_{13}+H_{24}-(H_{12}+H_{34})), \\ s &= \bar{H}_{13}-H_{13}, \\ t &= \bar{H}_{23}-H_{23}, \\ x &= \bar{H}_{14}-H_{14}, \\ y &= \bar{H}_{24}-H_{24}, \\ u &= \bar{H}_{12}-H_{12}, \\ w &= \bar{H}_{34}-H_{34}. \end{cases}$$

Note that the star topology case corresponds to $b=0$.

Then D is a tree metric iff the four-point condition holds, namely

$$\max\{2b+\max\{x,0\},\max\{y,0\},\max\{u,0\}+\max\{y,0\},2b+\max\{s,0\}+\max\{t,0\}\}$$

is attained at least twice.

Moreover in case D is a tree metric, we have

(i) D *has topology* $(12|34) \iff 2b+\max\{s,0\}+\max\{t,0\} = 2b+\max\{x,0\}+\max\{y,0\} > \max\{u,0\}+\max\{w,0\}$,
(ii) D *has topology* $(13|24) \iff 2b+\max\{x,0\}+\max\{y,0\} = \max\{u,0\}+\max\{w,0\} > 2b+\max\{s,0\}+\max\{t,0\}$,
(iii) D *has topology* $(14|23) \iff 2b+\max\{s,0\}+\max\{t,0\} = \max\{u,0\}+\max\{w,0\} > 2b+\max\{x,0\}+\max\{y,0\}$,
(iv) D *has the star topology* $\iff \max\{u,0\}+\max\{w,0\} = 2b+\max\{s,0\}+\max\{t,0\} = 2b+\max\{x,0\}+\max\{y,0\}$.

An interesting remark to make is that although H and $\bar{H}$ may have the star topology, D can have a non-star topological type. We will see several examples illustrating this and others unexpected cases on Section §4.

At this point, the reader might be troubled by the equations described on Therorem 3.3 and the commutativity of the mixture operation. Namely, what happens to the equations if we invert the order of H and $\bar{H}$. For this, we need to describe the algebraic relations among the variables $s, t, x, y, u, w, b, \bar{b}$.

Let us take a look at the system of equations (**). If we sum up the first two equations, we get $\bar{a}+2\bar{b}+\bar{c}+\bar{e}+\bar{d} = a+d+2b+c+e+s+t$. Using the last two equations, we get that $2\bar{b}+u+w = 2b+s+t$.

Likewise, if we sum up the third and forth equation, and we use the last two equations as we did before, we get $2\bar{b}+u+w = 2b+x+y$.

Therefore, the indeterminates $s, t, x, y, u, w, b, \bar{b}$ verify two linear relations:

$$2(b-\bar{b})+s+t = 2(b-\bar{b})+x+y = u+w.$$

One question we might ask at this point is if these conditions suffice to have two tree metrics H and $\bar{H}$ of the same topological type given values for the eight variables. It turns out that this is always the case. We refer to this discussion in the first part of Section §4.

Another important question one can formulate is the following. How do these quantities change if we switch the roles of H and $\bar{H}$. The new variables will be $s' = -s$, $t' = -t$, $x' = -x$, $y' = -y$, $u' = -u$, $w' = -w$, $b' = \bar{b}$ and $\bar{b}' = b$, i.e. switch signs on the first six variables and change the role of b and $\bar{b}$.

We need to show that the relations among the tree expressions $\max\{u,0\}+\max\{w,0\}, \max\{u,0\}+\max\{w,0\}, 2b+\max\{x,0\}+\max\{y,0\}$ are preserved if we replace the variables by the new prime variables. But this follows by construction, since we would have

$$\begin{cases} D_{12}+D_{34} &=& \bar{a}+\bar{d}+\max\{-u,0\}+\bar{c}+\bar{e}+\max\{-w,0\} \\ D_{13}+D_{24} &=& \bar{a}+\bar{b}+\bar{c}+\max\{-s,0\}+\bar{d}+\bar{b}+\bar{e}+\max\{-t,0\} \\ D_{14}+D_{23} &=& \bar{a}+\bar{b}+\bar{e}+\max\{-x,0\}+\bar{d}+\bar{b}+\bar{c}+\max\{-y,0\} \end{cases}$$

And so by substracting $\bar{a}+\bar{d}+\bar{c}+\bar{e}$ from the three expressions, the inequalities provided in Proposition 3.1 are preserved, and so we obtain the same characterization as in Proposition 3.2, but with all variables replaced as have we already explained.

An interesting question to formulate at this point is to try to determine the geometric structure of the space

$$\begin{aligned} \mathscr{S}_{T,T'} := \{(H,\bar{H}) : H, \bar{H} \text{ trees on 4 taxas with same tree topology } T, \\ H \vee \bar{H} \text{ is tree metric with a fixed topological type } T'\}. \end{aligned} \tag{3}$$

For example, one could ask if this set has a cone structure and moreover, a convex structure. As we have seen, the set $\mathscr{S}_{T,T'}$ is characterized by Theorem 3.3. And from this statement it is straightforward to check that $(H,\bar{H}) \in \mathscr{S}_{T,T'}$ and $\lambda > 0$, then $(\lambda H, \lambda\bar{H}) \in \mathscr{S}_{T,T'}$, since all new variables involved will correspond to the old variables scaled by λ. In particular, we will have $\lambda(H \vee \bar{H}) = (\lambda H) \vee (\lambda\bar{H})$, so the four-point condition for $\lambda(H \vee \bar{H})$ will also hold. Moreover, the inequalities on the theorem satisfied by the new variables will be the same as the old ones, but scaled by the factor λ. Since this number is positive, the new inequalities will be the same as the old ones. So $\lambda(H \vee \bar{H})$ will have topological type T'.

Remark 3.4. *Since the mixture of trees is a commutative operation, it follows that $(H,\bar{H}) \in \mathscr{S}_{T,T'}$ iff $(\bar{H},H) \in \mathscr{S}_{T,T'}$. This observation will be the key to discuss convexity.*

Let us concentrate on the convex structure. The first guess about the convexity of the set $\mathscr{S}_{T,T'}$ would be a positive answer. However, this turns out to be far from being true. The key-point in this case is given by the following lemma

Lemma 3.5. *The mixture of the convex sum (coordinatewise) of two elements in $\mathscr{S}_{T,T'}$ doesn't coincide with the convex combination of the corresponding mixtures.*

Proof. Consider $(H, \bar{H})$ and $(\bar{H}, H)$. As we discused on Remark 3.4, both points belong to the same $\mathscr{S}_{T,T'}$. Pick any $\alpha \in (0,1)$. Then we want to show that

$$H \vee \bar{H} = \alpha(H \vee \bar{H}) + (1-\alpha)(\bar{H} \vee H) \neq (\alpha H + (1-\alpha)\bar{H}) \vee (\alpha \bar{H} + (1-\alpha)H). \tag{4}$$

Assume that on the (i,j)-th entry, the maximum corresponds to $H_{ij} = a > b = \bar{H}_{ij}$. In this case, we want to show that both sides on the previous expression differ on the (i,j)-th. entry. Namely (LHS)$= \alpha a + (1-\alpha)a = a$, whereas (RHS)$= \max\{\alpha a + (1-\alpha)b, \alpha b + (1-\alpha)a\}$. In any case for $\alpha \in (0,1)$:

$$a\alpha + (1-\alpha)b = a \iff b = a \iff \alpha b + (1-\alpha)a = a.$$

Therefore, if $H_{ij} = a > b = \bar{H}_{ij}$, the inequality on (4) does hold. □

In addition to the inconvenience of Lemma 3.5, the question about convexity, although clearly stated, has an obscure nature. As we saw on Remark 3.4, the mixture of trees is a commutative operation, but in this case we are choosing an ordering $(H, \bar{H})$. The sum in $\mathscr{S}_{T,T'}$ will be given coordinatewise. This subtle point about the ordering will allow us to find examples s.t. $(H, \bar{H}), (H', \bar{H}') \in \mathscr{S}_{T,T'}$ but $(\alpha H + (1-\alpha)H') \vee (\alpha \bar{H} + (1-\alpha)\bar{H}')$ doesn't have T' as the underlying topological type for some $\alpha \in (0,1)$, that is $\alpha(H, \bar{H}) + (1-\alpha)(H', \bar{H}') \notin \mathscr{S}_{T,T'}$. Moreover, we will find examples where the mixture of the convex combination will *never* give a point in $\mathscr{S}_{T,T'}$ but if we reverse the order of the first point, then the convex combination of $(\bar{H}, H)$ and $(H', \bar{H}')$ always lies in $\mathscr{S}_{T,T'}$. We refer to Section §4 for details, in particular Example 4.

4. Examples

In the present section we present several examples that show that all situations can be achived. As we said on previous sections, in the generic case we will obtain a non-tree metric as a mixture of two tree metrics of the same topological type: the closed sets where this fails will be given by the union of three hyperplanes each one of them given by the equality of two out of three linear forms appearing in the four-point condition.

We explain briefly how we build all the examples in this sections. Recall from Section §3 that if we have H and $\bar{H}$ of topology $(12|34)$ then the eight variables $s, t, x, y, u, w, b, \bar{b}$ satisfy three linear equations:

$$2(b-\bar{b}) + s + t = 2(b-\bar{b}) + x + y = u + w \tag{5}$$

For example, we fix values for u, w and $b - \bar{b}$. For each situation described in Theorem 3.3 we consider a suitable value for b, s, t, x and y giving the right inequalities and satisfying $s+t = x+y = (u+w-2(b-\bar{b}))$, which is already fixed value.

After this, we need to pick positive values for $a, c, d, e, \bar{a}, \bar{c}, \bar{d}, \bar{e}$ and non-negative values for $b, \bar{b}$ (recall that we've fixed the value $b - \bar{b}$). To do this, we use equations in (**), namely:

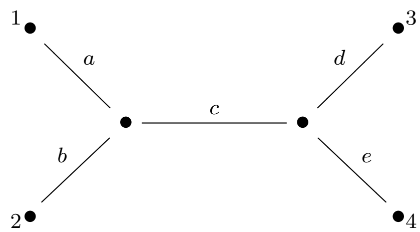
$$\begin{cases} \bar{a} + \bar{c} &= a + c + (b-\bar{b}) + s \\ \bar{d} + \bar{e} &= d + e + (b-\bar{b}) + t \\ \bar{a} + \bar{e} &= a + e + (b-\bar{b}) + x \\ \bar{d} + \bar{c} &= d + c + (b-\bar{b}) + y \\ \bar{a} + \bar{d} &= a + d + u \\ \bar{c} + \bar{e} &= c + e + w \end{cases}$$

We obtain:

$$\begin{cases} \bar{a} &= a + (b-\bar{b}) + \frac{s+x-w}{2} \\ \bar{d} &= d + (b-\bar{b}) + \frac{t+y-w}{2} \\ \bar{c} &= c + (b-\bar{b}) + \frac{s+y-u}{2} \\ \bar{e} &= e + (b-\bar{b}) + \frac{t+x-u}{2} \end{cases}$$

If we replace these values on the linear system, we get three equalities $2(b-\bar{b}) + x + y = u + w$, $2(b-\bar{b}) + s + t = u + w$ and $4(b-\bar{b}) + s + t + x + y = 2(u+w)$ that hold by hypothesis. Therefore, we *always* have values for the weights for the edges of H and $\bar{H}$. Moreover, by picking convenient values for the entries in H, the corresponding entries in $\bar{H}$ are also positive, thus we can obtain tree metrics of the same topological type for *any* set of variables $s, t, x, y, u, w, b, \bar{b}$ satisfying (5).

In the remainder of this section, we present several examples and the corresponding MAPLE code used to obtain the results. Recall that we will always work with tree metrics H, $\bar{H}$ on four taxa with the same topological type ((12|34) or the star topology). For simplicity, we will call $\bar{H}$ by K in our examples. First we present the general Maple code used for all the computations. An important remark to say is that we've changed notation with respect to the previous sections. In this case the weights on each tree correspond to



i.e. $w(e_1) = a$, $w(e_2) = b$, $w(e_0) = c$, $w(e_3) = d$ and $w(e_4) = e$ (e_0 denotes the middle node).

```
> A := matrix([[0,a+b,a+c+d, a+c+e],[a+b,0,b+c+d,b+c+e],[a+c+d,b+c+d,0,d+e],
[a+c=e,b+c+e,d+e,0]]);
```

So the matrix corresponding to the tree metric of topological type (12|34) or the start topology has the form:

$$\begin{bmatrix} 0 & a+b & a+c+d & a+c+e \\ a+b & 0 & b+c+d & b+c+e \\ a+c+d & b+c+d & 0 & d+e \\ a+c+e & b+c+e & d+e & 0 \end{bmatrix}$$

Example 4.1 In this case, we show that given H and $\bar{H}$ tree metrics with topology $T = (12|34)$ we might get a dissimilarity map $H \vee \bar{H}$ that is not even tree additive w.r.t. any tree, since it doesn't satisfy the weak four-point condition.

```
> H := matrix(4, 4); for i to 4 do for j to 4 do
H[i, j] := subs({a = 2, b = 3, c = 4, d = 1, e = 6}, A[i, j]) end do end do;
> H:
> eval(H);
> K := matrix(4, 4); for i to 4 do for j to 4 do
K[i, j] := subs({a = 4, b = 38, c = 1, d = 1, e = 6}, A[i, j]) end do end do;
> K:
> eval(K);
```

We get

$$H = \begin{bmatrix} 0 & 5 & 7 & 12 \\ 5 & 0 & 8 & 13 \\ 7 & 8 & 0 & 7 \\ 12 & 13 & 7 & 0 \end{bmatrix}; \; K = \begin{bmatrix} 0 & 42 & 6 & 11 \\ 42 & 0 & 40 & 45 \\ 6 & 40 & 0 & 7 \\ 11 & 45 & 7 & 0 \end{bmatrix}.$$

After this, we compute the mixture of H and K and check the three expressiones involved in the four point condition.

```
> D := matrix(4, 4); for i to 4 do for j to 4 do
D[i, j] := max(H[i, j], K[i, j]) end do end do;
> D: eval(D);
> G:=vector(3,[D[1,2]+D[3,4],D[1,3]+D[2,4],D[1,4]+D[2,3]]);
```

And we get

$$D = \begin{bmatrix} 0 & 42 & 7 & 16 \\ 42 & 0 & 40 & 45 \\ 7 & 40 & 0 & 11 \\ 16 & 45 & 11 & 0 \end{bmatrix};$$

$$G = \begin{bmatrix} 53 & 52 & 56 \end{bmatrix}.$$

Thus, we see that this is not tree additive, and so in particular the mixture is not a tree metric. □

Now we present the examples that give tree metrics. In this case, we will the check conditions on Theorem 3.3. More concretely, after computing the mixture D and the vector G (see Example 4) we get

- If $G[1] = G[2] > G[3]$ then the topology is $(14|23)$;
- If $G[1] < G[2] = G[3]$ then the topology is $(12|34)$;
- If $G[1] = G[3] > G[2]$ then the topology is $(13|24)$;
- If $G[1] = G[2] = G[3]$ then the topology is the star topology.

In the cases in which we obtain a tree metric, we can recover the corresponding weights using different formulas according to each topology.

Example 4.2 In this case, we present an example where we obtain the *same* topology $(12|34)$.

```
> H := matrix(4, 4); for i to 4 do for j to 4 do
H[i, j] := subs({a = 2, b = 3, c = 4, d = 1, e = 6}, A[i, j]) end do end do;
> H:
> eval(H);
> K := matrix(4, 4); for i to 4 do for j to 4 do
K[i, j] := subs({a = 4, b = 3, c = 1, d = 1, e = 6}, A[i, j]) end do end do;
> K:
> eval(K);
```

We get

$$H = \begin{bmatrix} 0 & 5 & 7 & 12 \\ 5 & 0 & 8 & 13 \\ 7 & 8 & 0 & 7 \\ 12 & 13 & 7 & 0 \end{bmatrix} \quad ; \quad K = \begin{bmatrix} 0 & 7 & 6 & 11 \\ 7 & 0 & 5 & 10 \\ 6 & 5 & 0 & 7 \\ 11 & 10 & 7 & 0 \end{bmatrix}.$$

The corresponding mixture and the vector G turn out to be

$$D := \begin{bmatrix} 0 & 7 & 7 & 12 \\ 7 & 0 & 8 & 13 \\ 7 & 8 & 0 & 7 \\ 12 & 13 & 7 & 0 \end{bmatrix};$$

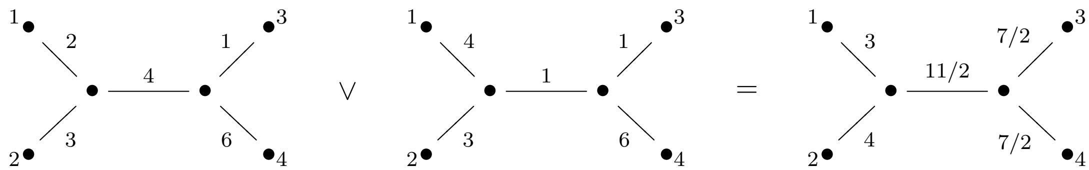
$$G := \begin{bmatrix} 14 & 20 & 20 \end{bmatrix}.$$

Let us compute the corresponding weights for the case of the topology $(12|34)$:

```
> e5 := -(1/2)*(D[1, 2]+D[3, 4]-D[1, 3]-D[2, 4]);
                                    3
> e1 := 1/2*(D[1, 4]+D[1, 3]-2*e5-D[3, 4]);
                                    3
> e2 := 1/2*(D[2, 4]+D[2, 3]-2*e5-D[3, 4]);
                                    4
> e3 := 1/2*(D[3, 1]+D[3, 2]-2*e5-D[1, 2]);
                                    1
```

```
> e4 := 1/2*(D[4, 1]+D[4, 2]-2*e5-D[1, 2]);
                                          6
```

Therefore:



□

Example 4.3 In this case, we present an example where we obtain the topology (13|24). The values for the parameters are: $u = 6, w = -5, s = 2, t = -1, x = 4, y = -3, (b - \bar{b}) = 0, b = 1$.

```
> H := matrix(4, 4); for i to 4 do for j to 4 do
H[i, j] := subs({a = 1, b = 2, c = 1, d = 9/2, e = 7/2}, A[i, j]) end do end do;
> H:
> eval(H);
> K := matrix(4, 4); for i to 4 do for j to 4 do
K[i, j] := subs({a = 13/2, b = 5/2, c = 1, d = 1, e = 2}, A[i, j]) end do end do;
> K:
> eval(K);
```

Thus:

$$H = \begin{bmatrix} 0 & 3 & 13/2 & 11/2 \\ 3 & 0 & 15/2 & 13/2 \\ 13/2 & 15/2 & 0 & 8 \\ 11/2 & 13/2 & 8 & 0 \end{bmatrix} \quad ; \quad K = \begin{bmatrix} 0 & 9 & 17/2 & 19/2 \\ 9 & 0 & 9/2 & 11/2 \\ 17/2 & 9/2 & 0 & 3 \\ 19/2 & 11/2 & 3 & 0 \end{bmatrix}$$

$$D = \begin{bmatrix} 0 & 9 & 17/2 & 19/2 \\ 9 & 0 & 15/2 & 13/2 \\ 17/2 & 15/2 & 0 & 8 \\ 19/2 & 13/2 & 8 & 0 \end{bmatrix};$$

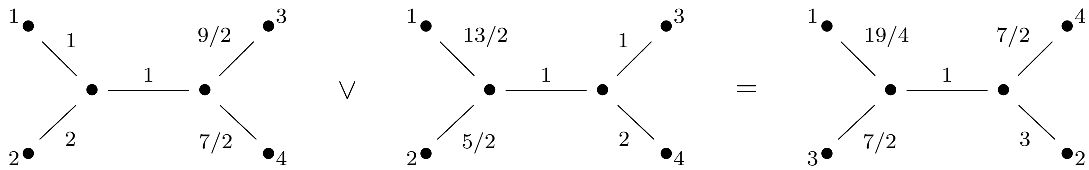
$$G = \left[\begin{array}{ccc} 17 & 15 & 17 \end{array}\right].$$

Let us compute the corresponding weights for the case of the topology (13|24):

```
> e5 := (1/2)*(D[1, 2]+D[3, 4]-D[1, 3]-D[2, 4]):
                                          1
> e1 := 1/2*(D[1, 4]+D[1, 3]-2*e5-D[2, 4]):
                                         19
                                         --
                                         4
> e2 := 1/2*(D[2, 3]+D[1, 2]-2*e5-D[1, 3]):
                                          3
> e3 := 1/2*(D[3, 4]+D[3, 2]-2*e5-D[2, 4]);
                                          7
                                          -
                                          2
> e4 := 1/2*(D[4, 1]+D[3, 4]-2*e5-D[1, 3]);
                                          7
                                          -
```

```
                                       2
```

Therefore:



□

Example 4.4 Now an example where we get topology (13|24) and the inner edge has different weight on each tree. In this case, the parameters are: $u = 6, w = -5, b = 3/2, (b - \bar{b}) = 1, s = 1, t = -2, x = 3$ and $y = -4$.

```
> H := matrix(4, 4); for i to 4 do for j to 4 do
H[i, j] := subs({a = 11/2, b = 3/2, c = 3/2, d = 11/2, e = 5/2}, A[i, j]) end do end do;
> H:
> eval(H);
> K := matrix(4, 4); for i to 4 do for j to 4 do
K[i, j] := subs({a = 11, b = 2, c = 1/2, d = 2, e = 1}, A[i, j]) end do end do;
> K:
> eval(K);
```

Thus:

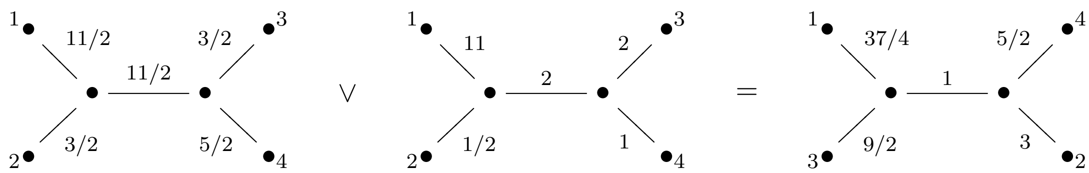
$$H = \begin{bmatrix} 0 & 7 & \frac{25}{2} & 19/2 \\ 7 & 0 & 17/2 & 11/2 \\ \frac{25}{2} & 17/2 & 0 & 8 \\ 19/2 & 11/2 & 8 & 0 \end{bmatrix} \quad ; \quad K = \begin{bmatrix} 0 & 13 & \frac{27}{2} & \frac{25}{2} \\ 13 & 0 & 9/2 & 7/2 \\ \frac{27}{2} & 9/2 & 0 & 3 \\ \frac{25}{2} & 7/2 & 3 & 0 \end{bmatrix}.$$

$$D = \begin{bmatrix} 0 & 13 & \frac{27}{2} & \frac{25}{2} \\ 13 & 0 & 17/2 & 11/2 \\ \frac{27}{2} & 17/2 & 0 & 8 \\ \frac{25}{2} & 11/2 & 8 & 0 \end{bmatrix};$$

$$G = \begin{bmatrix} 21 & 19 & 21 \end{bmatrix}.$$

```
> e5 := (1/2)*(D[1, 2]+D[3, 4]-D[1, 3]-D[2, 4]);
                                       1
> e1 := 1/2*(D[1, 4]+D[1, 3]-2*e5-D[2, 4]);
                                      37
                                      --
                                      4
> e2 := 1/2*(D[2, 3]+D[1, 2]-2*e5-D[1, 3]);
                                       3
> e3 := 1/2*(D[3, 4]+D[3, 2]-2*e5-D[2, 4]);
                                       9
                                       -
                                       2
> e4 := 1/2*(D[4, 1]+D[3, 4]-2*e5-D[1, 3]);
                                       5
                                       -
                                       2
```

Therefore:



□

Example 4.5 Now we present one example where H has the *star* topology whereas K has topology (12|34). We obtain a mixture with topology (13|24). In this case we have $b = 0$ and $\bar{b} > 0$, for example $\bar{b} = 1$. The other parameters are $u = 6, w = -5, s = 4, t = -1, x = 6$ and $y = -3$.

```
> H := matrix(4, 4); for i to 4 do for j to 4 do
H[i, j] := subs({a = 1/2, b = 5/2, c = 0, d = 4, e = 9/2}, A[i, j]) end do end do;
> H:
> eval(H);
> K := matrix(4, 4); for i to 4 do for j to 4 do
K[i, j] := subs({a = 7, b = 2, c = 1, d = 1/2, e = 3}, A[i, j]) end do end do;
> K:
> eval(K);
```

Thus:

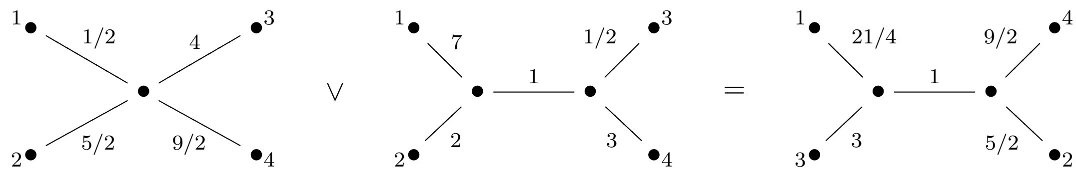
$$H = \begin{bmatrix} 0 & 3 & 9/2 & 5 \\ 3 & 0 & 13/2 & 7 \\ 9/2 & 13/2 & 0 & 17/2 \\ 5 & 7 & 17/2 & 0 \end{bmatrix} \quad ; \quad K = \begin{bmatrix} 0 & 9 & 17/2 & 11 \\ 9 & 0 & 7/2 & 6 \\ 17/2 & 7/2 & 0 & 7/2 \\ 11 & 6 & 7/2 & 0 \end{bmatrix}.$$

$$D = \begin{bmatrix} 0 & 9 & 17/2 & 11 \\ 9 & 0 & 13/2 & 7 \\ 17/2 & 13/2 & 0 & 17/2 \\ 11 & 7 & 17/2 & 0 \end{bmatrix};$$

$$G = \left[\begin{array}{ccc} \frac{35}{2} & \frac{31}{2} & \frac{35}{2} \end{array}\right].$$

```
> e5 := (1/2)*(D[1, 2]+D[3, 4]-D[1, 3]-D[2, 4]);
                                        1
> e1 := 1/2*(D[1, 4]+D[1, 3]-2*e5-D[2, 4]);
                                       21
                                       --
                                       4
> e2 := 1/2*(D[2, 3]+D[1, 2]-2*e5-D[1, 3]);
                                        5
                                        -
                                        2
> e3 := 1/2*(D[3, 4]+D[3, 2]-2*e5-D[2, 4]);
                                        3
> e4 := 1/2*(D[4, 1]+D[3, 4]-2*e5-D[1, 3]);
                                        9
                                        -
                                        2
```

Thus,



□

Example 4.6 In this case, H has the star topology whereas K has topology $(12|34)$. We obtain a mixture with topology $(12|34)$. As in the previous example, we pick a value for $\bar{b}$, say $\bar{b} = 2$. The other parameters are $b = 0, u = 1, w = -2, s = 4, t = 1, x = 3$ and $y = 2$.

```
> H := matrix(4, 4); for i to 4 do for j to 4 do
H[i, j] := subs({a = 1/2, b = 7/2, c = 0, d = 3/2, e = 4}, A[i, j]) end do end do;
> H:
> eval(H);
> K := matrix(4, 4); for i to 4 do for j to 4 do
K[i, j] := subs({a = 4, b = 5, c = 1, d = 3, e = 9/2}, A[i, j]) end do end do;
> K:
> eval(K);
```

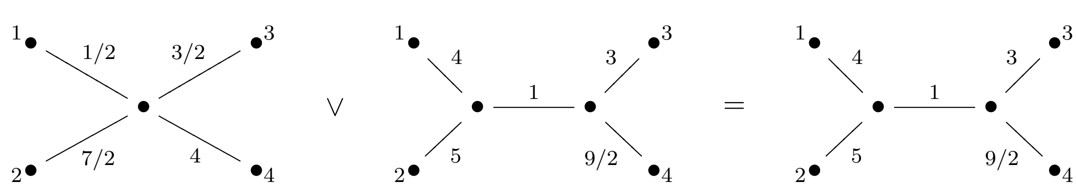
Thus:

$$H = \begin{bmatrix} 0 & 4 & 2 & 9/2 \\ 4 & 0 & 5 & 15/2 \\ 2 & 5 & 0 & 11/2 \\ 9/2 & 15/2 & 11/2 & 0 \end{bmatrix} \quad ; \quad K = \begin{bmatrix} 0 & 9 & 8 & 19/2 \\ 9 & 0 & 9 & 21/2 \\ 8 & 9 & 0 & 15/2 \\ 19/2 & 21/2 & 15/2 & 0 \end{bmatrix}.$$

$$D = \begin{bmatrix} 0 & 9 & 8 & 19/2 \\ 9 & 0 & 9 & 21/2 \\ 8 & 9 & 0 & 15/2 \\ 19/2 & 21/2 & 15/2 & 0 \end{bmatrix};$$

$$G = \left[\ \tfrac{33}{2} \quad \tfrac{37}{2} \quad \tfrac{37}{2}\ \right].$$

```
> e5 := -(1/2)*(D[1, 2]+D[3, 4]-D[1, 3]-D[2, 4]);
                                        1
> e1 := 1/2*(D[1, 4]+D[1, 3]-2*e5-D[3, 4]);
                                        4
> e2 := 1/2*(D[2, 4]+D[2, 3]-2*e5-D[3, 4]);
                                        5
> e3 := 1/2*(D[3, 1]+D[3, 2]-2*e5-D[1, 2]);
                                        3
> e4 := 1/2*(D[4, 1]+D[4, 2]-2*e5-D[1, 2]);
                                        9
                                        -
                                        2
```



□

Example 4.7 Now, one example where we obtain the star topology and $H, \bar{H}$ have topology $(12|34)$. In this case, the parameters are $b = 3, (b - \bar{b}) = 2, u = 8, w = 0, s = 2, t = -4, x = -4$ and $y = -2$.

```
> H := matrix(4, 4); for i to 4 do for j to 4 do
H[i, j] := subs({a = 1, b = 5, c = 3, d = 7, e = 8}, A[i, j]) end do end do;
> H:
> eval(H);
> K := matrix(4, 4); for i to 4 do for j to 4 do
K[i, j] := subs({a = 5, b = 9, c = 1, d = 7, e = 2}, A[i, j]) end do end do;
> K:
> eval(K);
```

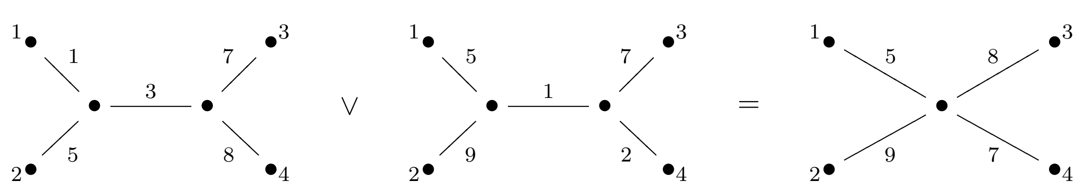
Thus:

$$H = \begin{bmatrix} 0 & 6 & 11 & 12 \\ 6 & 0 & 15 & 16 \\ 11 & 15 & 0 & 15 \\ 12 & 16 & 15 & 0 \end{bmatrix} \quad ; \quad K = \begin{bmatrix} 0 & 14 & 13 & 8 \\ 14 & 0 & 17 & 12 \\ 13 & 17 & 0 & 9 \\ 8 & 12 & 9 & 0 \end{bmatrix}$$

$$D = \begin{bmatrix} 0 & 14 & 13 & 12 \\ 14 & 0 & 17 & 16 \\ 13 & 17 & 0 & 15 \\ 12 & 16 & 15 & 0 \end{bmatrix};$$

$$G = \begin{bmatrix} 29 & 29 & 29 \end{bmatrix}.$$

```
> e5 := -(1/2)*(D[1, 2]+D[3, 4]-D[1, 3]-D[2, 4]);
                                    0
> e1 := 1/2*(D[1, 4]+D[1, 3]-2*e5-D[3, 4]);
                                    5
> e2 := 1/2*(D[2, 4]+D[2, 3]-2*e5-D[3, 4]);
                                    9
> e3 := 1/2*(D[3, 1]+D[3, 2]-2*e5-D[1, 2]);
                                    8
> e4 := 1/2*(D[4, 1]+D[4, 2]-2*e5-D[1, 2]);
                                    7
```



□

Now we will present some examples dealing with the cone structure mentioned on the previous section. By the construction, if we have parameters $s, t, u, w, x, y, b, \bar{b}$ for (H, K), and $s', t', u', w', x', y', b', \bar{b}'$ for (H', K'), then the parameters corresponding to $(\alpha H + (1-\alpha)H') \bigvee (\alpha K + (1-\alpha)K')$ will be convex combinations of the first group variables and the second group variables, namely:

$$s'' = \alpha s + (1-\alpha)s' \quad ; \quad t'' = \alpha t + (1-\alpha)t' \quad ; \quad u'' = \alpha u + (1-\alpha)u' \quad ; \quad w'' = \alpha w + (1-\alpha)w' \quad ;$$
$$x'' = \alpha x + (1-\alpha)x' \quad ; \quad y'' = \alpha y + (1-\alpha)y' \quad ; \quad b'' = \alpha b + (1-\alpha)b' \quad ; \quad \bar{b}'' = \alpha \bar{b} + (1-\alpha)\bar{b}' \quad .$$

As we mentioned in the previous section, we present an example that gives totally different results if we consider convex combination of (H, K) and (H', K') versus convex combinations of (K, H) and (H', K'), i.e. when we reverse the mixing order in the first pair.

Before this example, we present another one where $(H, K), (H', K') \in \mathscr{S}_{(12|34),(12|34)}$ and $\alpha(H, K) + (1-\alpha)(H', K') \in \mathscr{S}_{(12|34),(13|24)}$ for all $\alpha \in (0, 1)$.

Example 4.8 In this case, we pick T, T' as $(12|34)$ and $(13|24)$, and we consider (H, K) and (H', K') as in Examples 4 and 4 respectively. Thus $(H, K), (H', K') \in \mathscr{S}_{T,T'}$.

Recall from these two examples that we have $u = 6, w = -5, s = 2, t = -1, x = 4, y = -3, b = \bar{b} = 1$ and $u' = 6, w' = -5, s' = 1, t' = -2, x' = 3, y' = -4, (b' - \bar{b}') = 1, b' = 3/2$. Thus:

$$s'' = 1 + \alpha \ ; \ t'' = -2 + \alpha \ ; \ u'' = 6 \ ; \ w'' = 5 \ ; \ x'' = \alpha + 3 \ ; \ y'' = \alpha - 4 \ ; \ b'' = \frac{3}{2} - \frac{1}{2}\alpha \ ; \ \bar{b}'' = \frac{1}{2} + \frac{1}{2}\alpha.$$

Since $\alpha \in (0, 1)$ we get $y'' < 0$ and $s > 0$ for all α. As before, we need to compare the expressions in Proposition 3.2:

$$\begin{cases} \max\{u'', 0\} + \max\{w'', 0\} & = & 6 \\ 2b'' + \max\{x'', 0\} + \max\{y'', 0\} & = & 3 - \alpha + \alpha + 3 = 6 \\ 2b'' + \max\{s'', 0\} + \max\{t'', 0\} & = & 3 - \alpha + 1 + \alpha = 4 \end{cases}$$

So in this case, we always get topology $(13|24)$.

□

Example 4.9 In this case, we will consider (H, K) as in Example 4 and we take a new pair (H', K') with parameters $u' = -7, w' = 10, s' = 7, t' = -1, x' = -2, y' = 8, b' = 1, (b' - \bar{b}') = -3/2$, so $\bar{b}' = 5/2$. Note that $10 = max\{u', 0\} + \max\{w', 0\} = 2b' + \max\{x', 0\} + \max\{y', 0\} > 9 = 2 + \max\{s', 0\} + \max\{t', 0\}$ so the topology of the mixture is $(13|24)$, i.e. $(H', K') \in \mathscr{S}_{(12|34),(13|24)}$.

Notice that $2(b' - \bar{b}') + s' + t' = 2(b' - \bar{b}') + x' + y' = u' + w' = 3$, so the parameters where chosen correctly. We omit the computation of valid weights for the edges of H' and K', but this is not a problem since we've already proved they exist.

We consider the convex combination:

$$s'' = 7 - 6\alpha \quad ; \quad t'' = -\alpha - 1 \quad ; \quad u'' = 13\alpha - 7 \quad ; \quad w'' = -15\alpha + 10 \quad ;$$

$$x'' = 5\alpha - 2 \quad ; \quad y'' = -12\alpha + 8 \quad ; \quad b'' = \frac{1}{2}\alpha + 1 \quad ; \quad \bar{b}'' = -2\alpha + \frac{5}{2} \quad .$$

As before, we have $s > 0, t < 0, b > 0$ for all α. The breaking points will be given by the zeros of u'', w'', x'' and y'', That is $\alpha = \frac{7}{13}, \frac{2}{3}$ and $\frac{2}{5}$:

Therefore,

$$\max\{u'', 0\} + \max\{w'', 0\} = \begin{cases} -15\alpha + 10 & \text{if } \alpha < \frac{7}{13} \\ -2\alpha + 3 & \text{if } \frac{7}{13} \leq \alpha < \frac{2}{3} \\ 13\alpha - 7 & \text{if } \frac{2}{3} \leq \alpha \end{cases}$$

$$2b'' + \max\{x'', 0\} + \max\{y'', 0\} = \begin{cases} 10 - 11\alpha & \text{if } \alpha < \frac{2}{5} \\ 8 - 6\alpha & \text{if } \frac{2}{5} \leq \alpha < \frac{2}{5} \\ 6\alpha & \text{if } \frac{2}{3} \leq \alpha \end{cases}$$

$$2b'' + \max\{s'', 0\} + \max\{t'', 0\} = 9 - 5\alpha.$$

By simple inspection we see that the only values of α that give tree metrics are $\alpha = \frac{1}{6}$ and $\alpha = \frac{9}{11}$. In both cases we obtain a tree with topology $(12|34)$).

For example, the values for the corresponding matrices $H'' = \alpha H + (1-\alpha)H'$ and $K'' = \alpha K + (1-\alpha)K'$ in the case $\alpha = \frac{1}{6}$ are:

```
> H'' := matrix(4, 4); for i to 4 do for j to 4 do
H''[i, j] := subs({a = 106/24, b = 77/12, c = 13/2, d = 7/2, e = 10}, A[i, j]) end do end do;
> H'':
> eval(H'');
> K'' := matrix(4, 4); for i to 4 do for j to 4 do
K''[i, j] := subs({a = 2, b = 1, c = 13/6, d = 3, e = 4}, A[i, j]) end do end do;
> K'':
> eval(K'');
```

Thus:

$$H'' = \begin{bmatrix} 0 & \frac{65}{6} & \frac{173}{12} & \frac{251}{12} \\ \frac{65}{6} & 0 & \frac{197}{12} & \frac{275}{12} \\ \frac{173}{12} & \frac{197}{12} & 0 & \frac{27}{2} \\ \frac{251}{12} & \frac{275}{12} & \frac{27}{2} & 0 \end{bmatrix} \quad ; \quad K'' = \begin{bmatrix} 0 & 3 & \frac{43}{6} & \frac{49}{6} \\ 3 & 0 & \frac{37}{6} & \frac{43}{6} \\ \frac{43}{6} & \frac{37}{6} & 0 & 7 \\ \frac{49}{6} & \frac{43}{6} & 7 & 0 \end{bmatrix}$$

$$D'' = \begin{bmatrix} 0 & \frac{65}{6} & \frac{173}{12} & \frac{251}{12} \\ \frac{65}{6} & 0 & \frac{197}{12} & \frac{275}{12} \\ \frac{173}{12} & \frac{197}{12} & 0 & \frac{27}{2} \\ \frac{251}{12} & \frac{275}{12} & \frac{27}{2} & 0 \end{bmatrix};$$

$$G'' = \begin{bmatrix} \frac{73}{3} & \frac{112}{3} & \frac{112}{3} \end{bmatrix}.$$

```
> e5'' := -(1/2)*(P[1, 2]+P[3, 4]-P[1, 3]-P[2, 4]);
                                  13
                                  --
                                  2
> e1'' := 1/2*(P[1, 4]+P[1, 3]-2*e5-P[3, 4]);
                                  53
                                  --
                                  12
> e2'' := 1/2*(P[2, 4]+P[2, 3]-2*e5-P[3, 4]);
                                  77
                                  --
                                  12
> e3'' := 1/2*(P[3, 1]+P[3, 2]-2*e5-P[1, 2]);
                                  7
                                  -
                                  2
> e4'' := 1/2*(P[4, 1]+P[4, 2]-2*e5-P[1, 2]);
                                  10
```

□

Example 4.10 To finish, we provide an example where we use the same pairs as in Example 4 but we reverse the order of the first one, that is we take $(K, H), (H', K') \in \mathscr{S}_{(12|34),(13|24)}$. In this case, we'll see that infinitely many values of α will give tree metrics and for all these values we obtain a tree with topology $(12|34)$, that is different from the original one. Moreover, all other values of α (also infinitely many) won't give tree metrics.

Recall that when changing the order of the tree metrics we need to modify the parameters. In our case, we get $u = -6, w = 5, s = -2, t = 1, x = -4, y = 3, b = \bar{b} = 1$ and $u' = 6, w' = -5, s' = 1, t' = -2, x' =$

$3, y' = -4, (b' - \bar{b}') = 1, b' = 3/2$. And so the convex combination gives:

$$s'' = 1 - 3\alpha \quad ; \quad t'' = -2 + 3\alpha \quad ; \quad u'' = -12\alpha + 6 \quad ; \quad w'' = 10\alpha - 5 \quad ;$$

$$x'' = -7\alpha + 3 \quad ; \quad y'' = 7\alpha - 4 \quad ; \quad b'' = -\frac{1}{2}\alpha + \frac{3}{2} \quad ; \quad \bar{b}'' = \frac{1}{2}\alpha + \frac{1}{2} \quad .$$

Analyzing the breaking points we get:

$$\max\{u'', 0\} + \max\{w'', 0\} = \begin{cases} -12\alpha + 6 & \text{if } \alpha < \frac{1}{2} \\ 10\alpha - 5 & \text{if } \frac{1}{2} \leq \alpha \end{cases}$$

$$2b'' + \max\{x'', 0\} + \max\{y'', 0\} = \begin{cases} -8\alpha + 6 & \text{if } \alpha < \frac{3}{7} \\ -\alpha + 3 & \text{if } \frac{3}{7} \leq \alpha < \frac{4}{7} \\ 6\alpha - 1 & \text{if } \frac{2}{3} \leq \alpha \end{cases}$$

$$2b'' + \max\{s'', 0\} + \max\{t'', 0\} = \begin{cases} -4\alpha + 4 & \text{if } \alpha < \frac{1}{3} \\ -\alpha + 3 & \text{if } \frac{1}{3} \leq \alpha < \frac{2}{3} \\ 2\alpha + 1 & \text{if } \frac{2}{3} \leq \alpha \end{cases}$$

In this case for all $\alpha \notin \{\frac{3}{7}\} \cup [\frac{1}{2}; \frac{4}{7}]$, $0 < \alpha < 1$, the mixture of H'' and K'' is not a tree metric (although in some cases we do obtain tree additive dissimilarity maps).

If $\alpha \in \{\frac{3}{7}\} \cup [\frac{1}{2}; \frac{4}{7}]$ then $H'' \vee K''$ has topology $(12|34)$. If $\alpha = 0, 1$, then $(H'', K'') \in \mathscr{S}_{(13|24),(13|24)}$ since they correspond to the mixtures $K \vee H = H \vee K$ and $H' \vee K'$).

In this case we see that although $\alpha = \frac{3}{7}$ and $\frac{1}{2}$ give mixtures in $\mathscr{S}_{(12|34),(12|34)}$, all values of α between these two give mixtures that are not tree metrics. This behaviour differs from the one exposed in [2] for the case of phylogenetic mixtures, where if the parameters $\beta < \beta'$ give trees of a fixed topology, then all intermediate values in $[\beta, \beta']$ give the same result. □

Thus, as we see from the examples, the behaviour of convex combinations of tropical mixtures is far from being predictable and in general one needs to treat each case separately.

References

[1] L. Pachter and B. Sturmfels (editors). *Algebraic statistics for computational biology*. Cambridge University Press, 2005.

[2] F.A. Matsen and M. Steel. *Phylogenetic mixtures on a single tree can mimic a tree of another topology*. Sys. Bio., 56:767-775, Oct 2007.

[3] Ch. Sempel and M. Steel. *Phylogenetics*. Oxford University Press, 2003.

[4] L. Pachter (editor) *Lecture Notes for Math 239 - Spring 2008*. UC Berkeley. Available at: *http://math.berkeley.edu/~lpachter/239/*

E-mail address: macueto@math.berkeley.edu