A Note on the Approximability of the Balanced Minimum Evolution Problem

Daniele Catanzaro[a], Raffaele Pesenti[b] and Francesco Pisanu[a,*]

[a]CORE-Université Catholique de Louvain, Voie du Roman Pays 34, Ottignies-Louvain-la-Neuve, 1348, Belgium
[a]Venice School of Management, Università Ca' Foscari di Venezia, Cannaregio 873, Venezia, 30121, Italy

ARTICLE INFO

Keywords:
balanced minimum evolution problem
cross-entropy minimization
unrooted binary trees
path-length matrices
approximation algorithms

ABSTRACT

The *Balanced Minimum Evolution Problem (BMEP)* is an $\mathcal{APX}$-hard nonlinear network design problem that has received attention from the bioinformatics and mathematical programming communities over the past two decades. In this work, we show that if all input distances are positive and bounded, this problem admits a polynomial-time approximation algorithm. We also identify some polynomially solvable instances of the BMEP, including additive dissimilarity matrices, and derive tight lower and upper bounds on the optimal solution to the problem.

1. Introduction

Consider a set $\Gamma = \{1, 2, \dots, n\}$ of $n \geq 3$ items and let $\mathbf{D} = \{d_{ij}\}$ be a *dissimilarity* matrix on Γ, i.e., a symmetric matrix with all diagonal entries equal to 0 and nonnegative off-diagonal entries. An *Unrooted Binary Tree* (UBT) T on Γ is a tree having Γ as its leaf set and internal vertices of degree 3 [10]. For example, Figure 1 shows a possible UBT for a set Γ of five items. Let Θ_n denote the set of all UBTs on Γ, and for a given UBT $T \in \Theta_n$, let τ_{ij} denote the number of edges on the unique path in T between items i and j. Then, the *Balanced Minimum Evolution Problem* (BMEP) seeks a UBT $T^*_{\mathbf{D}} \in \Theta_n$ that solves the following nonlinear network design problem:

$$\min_{T \in \Theta_n} L_{\mathbf{D}}(T) = \sum_{i \in \Gamma} \sum_{\substack{j \in \Gamma \\ j \neq i}} \frac{d_{ij}}{2^{\tau_{ij}}}. \tag{1}$$

The BMEP arises from the literature on molecular phylogenetics [9]. In this setting, Γ represents a collection of distinct aligned molecular sequences (commonly referred to as *taxa*), such as DNA, RNA, codon sequences, or entire genomes; $\mathbf{D}$ encodes estimated evolutionary distances between pairs of taxa [3, 4, 5, 13]; and a UBT on Γ represents a candidate *phylogeny*, i.e., a tree hypothesizing the evolutionary relationships among the considered taxa. The BMEP aims to identify, among an exponential number of such candidate trees, the one that minimizes the cross-entropy across taxa, thus selecting the phylogeny with the highest likelihood of having occurred over the course of their evolution [7].

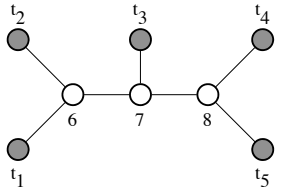
Beyond its biological motivation, recently reviewed in [9], the BMEP has also become an object of independent interest in algorithmic and combinatorial optimization. Research efforts have focused on its statistical foundations [12, 17, 18], its combinatorial structure [6, 8, 11, 15, 16, 19, 22], and algorithmic solution approaches, including both exact methods [1, 6, 21] and heuristics [14, 17, 21]. From a complexity-theoretic perspective, Fiorini and Joret [14] proved that the BMEP is $\mathcal{NP}$-hard, via a reduction from the 3-coloring problem, and inapproximable within any factor c^n, for some constant $c > 1$, unless $\mathcal{P} = \mathcal{NP}$. The authors also showed that when the dissimilarity matrix $\mathbf{D}$ is metric (i.e., when its entries satisfy the triangle inequality), the problem admits a polynomial-time approximation algorithm with an approximation factor of two. These hardness and inapproximability results rely on reductions that construct instances of the BMEP in which $\mathbf{D}$ contains zero-valued off-diagonal entries. Such cases, however, are rarely encountered in practical phylogenetic analyses, where taxa are typically assumed to be all distinct (leading so to dissimilarity matrices having strictly positive off-diagonal entries). Motivated by this observation, we investigate here the broader conditions under which practical instances of the BMEP become efficiently approximable. We show that when all pairwise distances are strictly positive and confined within a prescribed interval, the problem admits a polynomial-time approximation algorithm, whose performance guarantee is explicitly linked to the width of this interval. In addition, we identify specific subclasses of instances that are solvable in polynomial time and establish tight lower and upper bounds on the optimal solution to the problem. Altogether, these results provide new insight into the approximability of the BMEP in settings that are directly relevant to phylogenetic inference. Before proceeding, we introduce in the next section some notation and definitions that will prove useful throughout the article, and we briefly recall some prior results from the literature that will be instrumental in the subsequent sections.

2. Notation, Definitions, and Prior Work

Consider a UBT $T \in \Theta_n$. We say that two distinct leaves i and j of T form a *cherry* if the length of the unique path in T connecting i to j is equal to 2, i.e., $\tau_{ij} = 2$. For example, the leaves t_1 and t_2 in the UBT in Figure 1 form a cherry.

*Corresponding author
daniele.catanzaro@uclouvain.be (D. Catanzaro); pesenti@unive.it (R. Pesenti); francesco.pisanu@uclouvain.be (F. Pisanu)
ORCID(s): 0000-0001-9427-1562 (D. Catanzaro); 0000-0001-5890-4238 (R. Pesenti); 0000-0003-0799-5760 (F. Pisanu)

Taxon	Sequence		
t_1	A	A	A
t_2	A	C	C
t_3	C	G	C
t_4	C	C	G
t_5	G	A	G



$$\begin{array}{c} \\ t_1 \\ t_2 \\ t_3 \\ t_4 \\ t_5 \end{array} \begin{array}{c} \begin{array}{ccccc} t_1 & t_2 & t_3 & t_4 & t_5 \end{array} \\ \begin{bmatrix} 0 & 2 & 3 & 3 & 2 \\ 2 & 0 & 2 & 2 & 3 \\ 3 & 2 & 0 & 2 & 3 \\ 3 & 2 & 2 & 0 & 2 \\ 2 & 3 & 3 & 2 & 0 \end{bmatrix} \end{array} = \{d_{ij}\} \qquad \begin{array}{c} \\ t_1 \\ t_2 \\ t_3 \\ t_4 \\ t_5 \end{array} \begin{array}{c} \begin{array}{ccccc} t_1 & t_2 & t_3 & t_4 & t_5 \end{array} \\ \begin{bmatrix} 0 & 2 & 3 & 4 & 4 \\ 2 & 0 & 3 & 4 & 4 \\ 3 & 3 & 0 & 3 & 3 \\ 4 & 4 & 3 & 0 & 2 \\ 4 & 4 & 3 & 2 & 0 \end{bmatrix} \end{array} = \{\tau_{ij}\}$$

Figure 1: Above: a simple example of a set Γ of 5 taxa and the associated hypothetical DNA sequences (left), and a possible phylogeny of Γ encoded as a UBT (right). Below: an example of a *distance matrix* associated with Γ (left), obtained by assuming the Hamming distance as a measure of dissimilarity between pairs of sequences, along with the corresponding path-length matrix τ encoding the above UBT (right).

We say that $T \in \Theta_n$ is a *caterpillar* if, for some pair of distinct leaves $i, j \in \Gamma$, $\tau_{ij} = n - 1$ [10]. Stated otherwise, a caterpillar is a particular UBT characterized by precisely two cherries.

Given a dissimilarity matrix $\mathbf{D}$ on Γ, we define $d^+ = \max_{i \neq j} d_{ij}$ and $d^- = \min_{i \neq j} d_{ij}$. In addition, for a fixed $i \in \Gamma$, we denote $\mathbf{D}^i$ as the (dissimilarity) matrix obtained from $\mathbf{D}$ by removing its i-th row and column. Finally, whenever appropriate, we shall refer to a dissimilarity matrix as an *instance* of the BMEP.

We recall that a dissimilarity matrix $\mathbf{D}$ on Γ is said to be *metric* if its entries satisfy the following *triangle inequality*

$$d_{ij} + d_{jk} \geq d_{ik} \qquad \forall \text{ distinct } i, j, k \in \Gamma.$$

A metric dissimilarity matrix $\mathbf{D}$ on Γ is to be *additive* if for every distinct $i, j, p, q \in \Gamma$, precisely one of the following *four point conditions* holds [2]:

$$d_{ij} + d_{pq} \leq d_{ip} + d_{jq} = d_{iq} + d_{jp}, \tag{2}$$
$$d_{ip} + d_{jq} \leq d_{ij} + d_{pq} = d_{iq} + d_{jp}, \tag{3}$$
$$d_{iq} + d_{jp} \leq d_{ij} + d_{pq} = d_{ip} + d_{jq}. \tag{4}$$

For metric dissimilarity matrices the following result holds:

Proposition 2.1 (Fiorini and Joret [14]). *Any metric instance of the BMEP admits a 2-approximation algorithm.*

The approximation algorithm proposed by Fiorini and Joret [14] proceeds in two steps. First, it computes a *Minimum Spanning Tree* (MST) over the complete graph having the items in Γ as vertices and with the generic weight for the edge (i, j) being the entry d_{ij} of $\mathbf{D}$. Subsequently, the algorithm transforms this MST into a valid UBT on Γ with a total cost no greater than that of the MST by recursively replacing each internal node of degree greater than 3 with a binary subtree. Fiorini and Joret [14] show that the cost of the resulting UBT is at most twice the optimal value of the BMEP, thus yielding the 2-approximation guarantee.

Consider a dissimilarity matrix $\mathbf{D}$ on Γ and any tree T having Γ as a leaf set. Let w denote a function that associates a nonnegative weight $w(e)$ to each edge e of T. In addition, let P_{ij}^T denote the unique path in T between the distinct leaves $i, j \in \Gamma$ and define

$$w(P_{ij}^T) = \sum_{e \in P_{ij}^T} w(e).$$

Then, we say that T is *additive* for $\mathbf{D}$ (or *fits* $\mathbf{D}$, for short) if the following *additivity constraint* holds:

$$d_{ij} = w(P_{ij}^T) \qquad \forall \text{ distinct } i, j \in \Gamma. \tag{5}$$

A classical result from Buneman [2] shows that the set of *additive* dissimilarity matrices on Γ corresponds to the set of $W = \{w_{p_{ij}}^T\}$ matrices associated with additive trees on Γ and whose entries satisfy (5). Waterman et al. [23] shows that the additive tree T that fits a given additive dissimilarity matrix $\mathbf{D}$ always exists. Such a tree is unique if all four-point conditions hold with strict inequalities and can be constructed in polynomial time by means of the iterative algorithm described in [23]. A natural question is whether this algorithm could also serve as a certificate of optimality for additive instances of the BMEP. However, there is no clear evidence supporting this possibility. Indeed, the algorithm of Waterman et al. [23] is designed to address a feasibility problem based on the additivity condition (5), whereas the BMEP is an optimization problem that neither enforces nor assumes additivity, whether explicitly or implicitly. Thus, the tree returned by the Waterman et al. [23] algorithm may, in general, differ from the optimal solution to an additive instance $\mathbf{D}$ of the BMEP. Nonetheless, we will see in Section 3 that the foundational results of Buneman and Waterman et al. [23] still provide valuable insights for understanding the relationship between additive trees and the optimal solutions of additive BMEP instances. Before moving to the next section, however, we briefly recall here that any UBT $T \in \Theta_n$ can be encoded by means of a *path-length matrix* $\tau = \{\tau_{ij}\}$ whose combinatorial properties have been extensively discussed in [10]. Some properties that will be frequently invoked throughout the article are the following:

Proposition 2.2 (Catanzaro et al. [6]). *Let T be a UBT in Θ_n. Then, the path-length matrix $\tau = \{\tau_{ij}\}$ encoding T satisfies the following* Kraft equality*:*

$$\sum_{j\in\Gamma\setminus\{i\}} 2^{-\tau_{ij}} = \frac{1}{2} \qquad \forall\, i \in \Gamma.$$

Proposition 2.3 (Catanzaro et al. [6]). *Let T be a UBT in Θ_n. Then, the path-length matrix $\tau = \{\tau_{ij}\}$ encoding T satisfies the following* manifold condition*:*

$$\sum_{i\in\Gamma}\sum_{j\in\Gamma\setminus\{i\}} \tau_{ij} 2^{-\tau_{ij}} = 2n-3.$$

3. Efficiently solvable instances of the BMEP

In this section, we present some instances of the BMEP that can be solved efficiently. We start by considering the simplest case in which $\mathbf{D}$ is additive and has all equal non-diagonal entries. In such a case, the following result holds:

Proposition 3.1. *Let $\mathbf{D}$ be a dissimilarity matrix such that $d_{ij} = k$, for some real $k > 0$. Then, for every UBT $T \in \Theta_n$, $L_{\mathbf{D}}(T) = k \cdot n/2$.*

Proof. Let $T \in \Theta_n$. Then, we have that

$$L_{\mathbf{D}}(T) = \sum_{i\in\Gamma}\sum_{j\in\Gamma\setminus\{i\}} k 2^{-\tau_{ij}} = k\sum_{i\in\Gamma}\sum_{j\in\Gamma\setminus\{i\}} 2^{-\tau_{ij}} \overset{\text{Kraft}}{=} k\sum_{i\in\Gamma}\frac{1}{2} = k\,\frac{n}{2}.$$

□

This proposition shows that, in the context of the BMEP, the additivity of $\mathbf{D}$ does not guarantee the uniqueness of the optimal solution for the BMEP as for Waterman et al. [23].

Remark 1. Proposition 3.1 can be straightforwardly extended to the case in which $d_{ij} = (k_i + k_j)/2$, for some reals $k_i > 0$, $i \in \Gamma$. In this setting, for every UBT $T \in \Theta_n$, it holds that $L_{\mathbf{D}}(T) = \frac{1}{2}\sum_{i\in\Gamma} k_i$. An analogous extension can be formulated for all subsequent propositions that assume the dissimilarities d_{ij} contain a constant term k.

As a second case, consider a dissimilarity matrix $\mathbf{D}$ such that $d_{ij} \gg d_{pq}$ for some item $i \in \Gamma$ and for every distinct $j, p, q \in \Gamma \setminus \{i\}$. We refer to these cases as *nearly rooted*. Without loss of generality, assume that $i = 1$, that the values d_{1j} are sorted in increasing order, and that $d_{pq} = k$, for all distinct $p, q \in \{2, \dots, n\}$ and some $0 < k < \min_{j\in\Gamma\setminus\{1\}} d_{1j}$ (as $\max_{p,q,k,\ell\in\Gamma\setminus\{i\}}\{(d_{pq} - d_{k\ell})/d_{1n} \approx 0\}$). A simple example satisfying this condition occurs when $d_{ij}/d_{pq} > 2^{n-2}$. Indeed, in this case, the penalty incurred by reducing the path length τ_{ij} by one unit outweighs the maximum possible benefit obtained by increasing up to $n-1$ other path lengths τ_{pq}. Then, the following result holds:

Proposition 3.2. *Let $\mathbf{D}$ be any dissimilarity matrix such that $d_{12} \le d_{13} \le \dots \le d_{1n}$, and $d_{pq} = k$, for all distinct $p, q \in \{2, \dots, n\}$ and for some $0 < k < \min_{j\in\Gamma\setminus\{1\}} d_{1j}$. Then, the value of the optimal solution to the instance of the BMEP $\mathbf{D}$ is*

$$L_{\mathbf{D}}(T^*_{\mathbf{D}}) = \frac{n}{2}k + \frac{d_{12}}{4} + \dots + \frac{d_{1j}}{2^j} + \dots + \frac{d_{1\,n-1}}{2^{n-1}} + \frac{d_{1n}}{2^{n-1}}.$$

Proof. Consider the matrix $\mathbf{D}' = \{d'_{ij}\}$, such that $d'_{ij} = d_{ij} - k$, for every distinct $i, j \in \Gamma$, and 0 otherwise. Then, by (1) and Proposition 3.1 we have that:

$$\begin{aligned} L_{\mathbf{D}}(T) &= \sum_{i\in\Gamma}\sum_{j\in\Gamma\setminus\{i\}} d_{ij}2^{-\tau_{ij}} = \\ &= \sum_{i\in\Gamma}\sum_{j\in\Gamma\setminus\{i\}} d'_{ij}2^{-\tau_{ij}} + \sum_{i\in\Gamma}\sum_{j\in\Gamma\setminus\{i\}} k\cdot 2^{-\tau_{ij}} = \\ &= L_{\mathbf{D}'}(T) + k\cdot n/2. \end{aligned}$$

As this holds for every UBT $T \in \Theta_n$, it also holds for the optimal solution $T^*_{\mathbf{D}}$ with respect to $\mathbf{D}$, and hence the optimal solution does not depend on k. Thus, suppose, without loss of generality, that $k = 0$. We proceed by induction. For $n = 4$, a direct computation shows that as $d_{12} \le d_{13} \le d_{14}$, the optimal solution is T^*_4, satisfying $\tau_{12} = 2$, $\tau_{13} = \tau_{14} = 3$. This configuration coincides with the claim that the UBT must be a caterpillar. Now, assume that the statement holds true for $n = r > 4$; then, we show that it also holds true for $n = r + 1$. To this end, for each $j \in \Gamma \setminus \{1\}$, denote by

$$L^j(T) = \sum_{\ell\in\Gamma\setminus\{j\}} d_{i\ell}\, 2^{-\tau_{i\ell}},$$

i.e., $L^j(T)$ represents the BMEP objective function $L_{\mathbf{D}}(T)$ with all the terms involving j removed. Minimizing the BMEP on the instance $\mathbf{D}^j$ is therefore equivalent to minimizing $L^j(T)$ over all UBTs on r items. Now, consider $\sum_{j\in\Gamma} L^j(T)$. Each pair $(1, \ell)$ appears in all addends except when $\ell = 1, j$. Hence, this sum equals $(r-1)L_{\mathbf{D}}(T)$. Since removing any item j does not alter the order of the other items in the sequence $d_{i1}, \dots, d_{in}$, we can use induction over $\mathbf{D}^j$. By the induction hypothesis, for every subset of r items, the optimal solution is the one that has been claimed. Moreover, this tree coincides with the one claimed, say $T' \in \Theta_{r+1}$, after deleting j and contracting the two edges adjacent to the unique (internal) node adjacent to j in T' by construction. Since $(r-1)L_{\mathbf{D}}(T')$ decomposes as $\sum_{j\in\Gamma\setminus\{i\}} L^j(T')$, and the minimum of each term is achieved by T', we conclude that $(r-1)L_{\mathbf{D}}(T')$ is minimal. Thus, $L_{\mathbf{D}}(T')$ itself is minimal by linearity. □

Remark 2. Consider a caterpillar T'' obtained by reversing the label order with respect to Proposition 3.2, i.e.,

$$\begin{aligned} L_{\mathbf{D}}(T'') = \frac{(n-2)}{2}k + \frac{d_{12}}{2^{n-1}} + \frac{d_{13}}{2^{n-1}} + \frac{d_{14}}{2^{n-2}} + \dots \\ + \frac{d_{1j}}{2^{n-j+2}} + \dots + \frac{d_{1\,n-1}}{8} + \frac{d_{1n}}{4}. \end{aligned}$$

Then, $L_{\mathbf{D}}(T'') \geq L_{\mathbf{D}}(T''')$ for every UBT $T''' \in \Theta_n$, i.e., T'' is the UBT with the greatest cost.

Now, consider the case where $\mathbf{D}$ is an additive dissimilarity matrix of order $n \geq 3$. Then, we present an inductive algorithm, outlined in Algorithm 1, that constructs, in polynomial-time, the optimal solution to this specific instance of the BMEP. This approach can be viewed as a reformulation of the algorithm introduced by Waterman et al. [23], tailored to exploit the particular combinatorial nature of the BMEP. The algorithm iteratively extends an initial UBT with four leaves until a UBT with n leaves is obtained. We start by considering the following base case:

Case $|\Gamma| = n = 3$**.** There exists a unique UBT in Θ_3; hence, the solution to this instance of the BMEP is trivial.

Case $|\Gamma| = n = 4$**.** Suppose that

$$d_{12} + d_{34} < d_{14} + d_{23} = d_{13} + d_{24}.$$

A direct computation shows that the tree T with $\tau_{12} = \tau_{34} = 2$ and $\tau_{13} = \tau_{14} = \tau_{23} = \tau_{24} = 3$ minimizes $L_{\mathbf{D}}(T)$ and therefore constitutes the optimal solution. If equality holds, i.e.

$$d_{12} + d_{34} = d_{14} + d_{23} = d_{13} + d_{24},$$

then all UBTs on four leaves are optimal.

Case $|\Gamma| = n = 5$**.** We first identify the best configuration for items $1, 2, 3$, and 4. Assume, without loss of generality, that

$$d_{12} + d_{34} < d_{14} + d_{23} = d_{13} + d_{24}, \tag{6}$$
$$d_{15} + d_{24} < d_{14} + d_{25} = d_{12} + d_{45}. \tag{7}$$

In this case, the optimal topology groups items 1 and 5 so that they form a cherry. This follows from two observations. First, inequalities (6)-(7) imply that

$$\begin{aligned} &d_{12}2^{-\tau_{12}} + d_{34}2^{-\tau_{34}} + d_{14}2^{-\tau_{14}} + d_{23}2^{-\tau_{23}} \\ &+d_{13}2^{-\tau_{13}}+d_{24}2^{-\tau_{24}}+d_{15}2^{-\tau_{15}}+d_{25}2^{-\tau_{25}}+d_{45}2^{-\tau_{45}} \end{aligned}$$

is minimized, as the smallest dissimilarities are associated with shorter paths and the largest with longer ones, consistent with (6) and (7). Second, any alternative inequality, such as

$$d_{12} + d_{45} < d_{15} + d_{24} = d_{14} + d_{25}, \tag{8}$$

would contradict the previous construction, as it would require $\tau_{12} + \tau_{45} < \tau_{15} + \tau_{24}$, which is incompatible with (6)-(7). Moreover, substituting d_{14} and d_{25} from (6)-(7) into (8) yields $d_{35} + d_{12} = d_{15} + d_{23}$, violating so the strict inequality in (7). This observation is consistent with Buneman [2] result: for any additive matrix $\mathbf{D}$, there exists a weighted tree T associated with $\mathbf{D}$. In particular, consider distinct $i, j, p, q \in \Gamma$ that satisfy the four-point condition, and assume that one of the two connected components of

Algorithm 1: BMEP Solver for additive instances (adjacency-based incremental construction)

Input: An additive dissimilarity matrix $\mathbf{D} = \{d_{ij}\}$ on the set of taxa Γ
Output: A UBT T on Γ

$n \leftarrow |\Gamma|$;
Select three distinct elements $i_0, j_0, p_0 \in \Gamma$ and set $S \leftarrow \{i_0, j_0, p_0\}$;
Construct the unique UBT T on S;
Build its adjacency list $\mathcal{A}(T)$;
while $|S| < n$ **do**
    Select an element $x \in \Gamma \setminus S$;
    Initialize $S_{\text{tmp}} \leftarrow S$, $T_{\text{tmp}} \leftarrow T$, $\mathcal{A}_{\text{tmp}} \leftarrow \mathcal{A}(T)$;
    while *position of x not yet determined* **do**
        Choose three distinct leaves $i, j, p \in S_{\text{tmp}}$ forming a minimal subtree of T_{tmp} (using $\mathcal{A}_{\text{tmp}}$), preferably such that $|\mathcal{L}(i, j, p)|$ and $|\mathcal{R}(i, j, p)|$ are as balanced as possible;
        Evaluate the four-point condition on $\{i, j, p, x\}$;
        if $d_{ix} + d_{jp} < d_{ip} + d_{jx} = d_{ij} + d_{px}$ **then**
            Remove all leaves in $\mathcal{R}(i, j, p)$ from S_{tmp}, T_{tmp}, and $\mathcal{A}_{\text{tmp}}$;
        else if $d_{ij} + d_{px} < d_{ip} + d_{jx} = d_{ix} + d_{pj}$ **then**
            Remove all leaves in $\mathcal{L}(i, j, p)$ from S_{tmp}, T_{tmp}, and $\mathcal{A}_{\text{tmp}}$;
        else
            x forms a cherry with j;
            Mark the position of x as determined;
            break;
        end
    end
    if $|S_{tmp}| \leq 4$ **then**
        $T_{\text{tmp}} \leftarrow$ SolveInstanceUpTo5($S_{\text{tmp}} \cup \{x\}, \mathbf{D}$);
        Mark the position of x as determined;
    end
    Update T by inserting x in the position found;
    Update $\mathcal{A}(T)$ accordingly;
    $S \leftarrow S \cup \{x\}$;
end
return T

$T \setminus \{\hat{i}\}$, where $\hat{i}$ is the adjacent vertex of i in T, contains the leaves j, p, and q. Then, one can deduce which of the three inequalities of the four-point condition holds for all triples j', p', q' within the same connected component such that $\min\{\tau_{ij'}, \tau_{ip'}, \tau_{iq'}\} \leq \max\{\tau_{ij}, \tau_{ip}, \tau_{iq}\}$.

Case $|\Gamma| = n > 5$**.** Consider a subset $S \subset \Gamma$, such that $|S| = s$, for some $5 \leq s < n$. Assume that the optimal UBT T for S is known and denote by x an item in $\Gamma \setminus S$. Select three distinct leaves i, j, and p of T such that, if T' is the subtree of T obtained by the union of the paths P_{ij}^T, P_{ip}^T, and P_{jp}^T, then no other subtree with exactly three leaves i', j', and p' in Γ is such that $T' \setminus \{i', j', p'\} \subset T \setminus \{i, j, p\}$. This ensures that the remaining leaves lie entirely either on one side, say $\mathcal{L}(i, j, p)$, or the other, say $\mathcal{R}(i, j, p)$, of T, when T is drawn on the plane (see Figure 2). In particular, the sets of items $\mathcal{L}(i, j, p)$ and $\mathcal{R}(i, j, p)$ lying on the left and right,

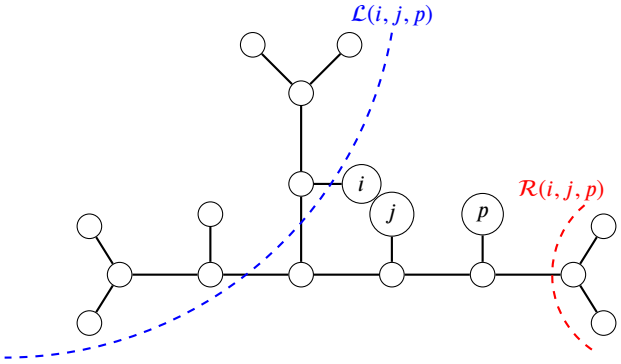


Figure 2: A UBT in Θ_{13} and the subpartition $(\mathcal{L}(i,j,p), \mathcal{R}(i,j,p))$ on the leaves associated with i, j, and p. Every leaf in $\mathcal{L}(i,j,p)$ is topologically closer to i, while every leaf in $\mathcal{R}(i,j,p)$ is topologically closer to p.

respectively, of this triplet are such that

$$\tau_{ia} \leq \tau_{ja},\ \tau_{ia} \leq \tau_{pa} \text{ for all } a \in \mathcal{L}(i,j,p),$$
$$\tau_{pb} \leq \tau_{ib},\ \tau_{pb} \leq \tau_{jb} \text{ for all } b \in \mathcal{R}(i,j,p).$$

Assume that i and p have the maximal path-length in T', as shown in Figure 2. If $d_{ix} + d_{jp} < d_{ip} + d_{jx} = d_{ij} + d_{px}$ holds true, the quantity

$$d_{ix}2^{-\tau_{ix}} + d_{jp}2^{-\tau_{jp}} + d_{ip}2^{-\tau_{ip}} + d_{jx}2^{-\tau_{jx}} + d_{ij}2^{-\tau_{ij}} + d_{px}2^{-\tau_{px}}$$

is smaller when x is placed either in $\mathcal{L}(i,j,p)$ or forms a cherry with i or j, consistent with the case of four leaves, than when it is placed in $\mathcal{R}(i,j,p)$ or forms a cherry with p. Conversely, if $d_{ij} + d_{px} < d_{ip} + d_{jx} = d_{ix} + d_{pj}$ holds, then either x belongs to $\mathcal{R}(i,j,p)$ or forms a cherry with p or j. In either case, one can decide if x forms a new cherry by simply testing a four-point condition on the corresponding group of four leaves involving x and its nearest leaf, as done in the case $|\Gamma| = 5$.

At each iteration, the leaves in $\mathcal{L}(i,j,p)$ or $\mathcal{R}(i,j,p)$ are excluded from further consideration since the corresponding inequalities have already been determined at previous steps, consistent with Buneman's result (i.e., this local optimal choice is also optimal for all the leaves in $\mathcal{L}(i,j,p)$ or $\mathcal{R}(i,j,p)$). Finally, if $d_{ij} + d_{px} = d_{ip} + d_{jx} = d_{ix} + d_{pj}$ holds, then x forms a cherry with any of i, j, or p, all yielding equivalent solutions. This procedure is iterated until all items in S have been compared to x, either directly or by exclusion, as belonging to one of the elements of the subpartition defined $\mathcal{L}(i'', j'', p'')$ by $\mathcal{R}(i'', j'', p'')$ for some i'', j'', and p'' in S generated during the process. Now, denote by

$$\mathcal{A}(T) = \{(i,j,p,\mathcal{L}(i,j,p),\mathcal{R}(i,j,p)) : i,j,p \text{ are leaves of } T\}$$

the collection of all triplets corresponding to the minimal subtrees described above, together with their associated bipartitions. This structure is used by the algorithm to efficiently identify, in $O(n^2)$ time, suitable triplets (i,j,p) that allow the leaf partitioning to be as balanced as possible, ensuring that the subsequent exclusion of $\mathcal{L}(i,j,p)$ or $\mathcal{R}(i,j,p)$ is performed efficiently. Moreover, maintaining and updating $\mathcal{A}(T)$ as leaves are added or removed is straightforward, which preserves the overall efficiency of the incremental construction.

In conclusion, at each iteration, the algorithm updates the adjacency structure $\mathcal{A}(T)$ while progressively removing from it the subsets of leaves $\mathcal{L}(\cdot,\cdot,\cdot)$ or $\mathcal{R}(\cdot,\cdot,\cdot)$ that are excluded by the four–point condition. Since $\mathcal{A}(T)$ contains $\mathcal{O}(n^2)$ entries and each update (removal or contraction) affects only a constant number of them, maintaining and querying this structure throughout the process requires overall $\mathcal{O}(n^2)$ time, dominating the operation of removing the vertices from $\mathcal{L}(\cdot,\cdot,\cdot)$ or $\mathcal{R}(\cdot,\cdot,\cdot)$. Therefore, the incremental construction of T via adjacency updates achieves total complexity $\mathcal{O}(n^3)$ and yields the optimal solution for the BMEP, as we show in the following result

Proposition 3.3. *Algorithm 1 yields the optimal solution to any additive instance of the BMEP.*

PROOF. The proof is similar to that of Proposition 3.2, as both instances preserve their properties upon removing a leaf. We therefore proceed by induction, indicating only the necessary differences. If $n = 4$, a direct computation shows that if $d_{12} + d_{34} \leq d_{13} + d_{24} = d_{14} + d_{23}$, the optimal solution is T_4^*, satisfying $2 + 2 = \tau_{12} + \tau_{34} < 3 + 3 = \tau_{13} + \tau_{24} = \tau_{14} + \tau_{23}$. This configuration coincides with the UBT constructed by the tree provided by Algorithm 1. Assume now that the statement holds true for $n = r > 4$; then, we prove it for $n = r + 1$. For each $i \in \Gamma$, define

$$L^i(T) = \sum_{\ell \in \Gamma} \sum_{\ell' \in \Gamma \setminus \{i\}} d_{\ell\ell'} 2^{-\tau_{\ell\ell'}}.$$

Also observe that minimizing the BMEP on the instance $\mathbf{D}^i$ is equivalent to minimizing $L^i(T)$ over all UBTs on $r + 1$ items. In addition, in the additive case we also have

$$\sum_{i \in \Gamma} L^i(T) = (r-1)L_{\mathbf{D}}(T).$$

Then, the proof of the statement for $n = r + 1$ can be obtained by using the same rationale used in the proof of Proposition 3.2. □

Combining the Algorithm 1 with Propositions 3.1, 2.3, and 2.3 leads to the following result:

Corollary 3.4. *Let T denote any UBT in Θ_n and $\mathbf{D}$ denote any dissimilarity matrix such that $d_{ij} = \alpha\tau_{ij} + k$, for some strictly positive constants α and k. Then, T is optimal for $\mathbf{D}$ and the optimal value is $L_{\mathbf{D}}(T) = \alpha(2n-3) + k \cdot n/2$.*

This result implies that if $\mathbf{D}$ is derived from an affine transformation of a tree, then the BMEP can be solved efficiently, and the optimal value is trivial.

Finally, we prove that for an additive instance of the BMEP, the optimal solution to the problem and the additive tree computed by Waterman et al. [23] iterative algorithm are identical, up to the edge weights.

Proposition 3.5. *Let $\mathbf{D}$ denote an additive instance of the BMEP and let $\hat{T}$ be the additive tree computed by Waterman et al. [23] iterative algorithm when considering the input $\mathbf{D}$. Then, $T^*_{\mathbf{D}} = \hat{T}$.*

Proof. We proceed by induction on n. For $n = 3$, the solution is unique.

Consider $n = 4$. If $d_{12} + d_{34} < d_{13} + d_{24} = d_{14} + d_{23}$, the optimal solution of the BMEP (and the one returned by Algorithm 1) is $\tau_{12} = \tau_{34} = 2$, $\tau_{13} = \tau_{24} = \tau_{14} = \tau_{23} = 3$. Let u, v be the two internal vertices of $\hat{T}$. Since $d_{ij} = w(P^{\hat{T}}_{ij})$ and the four–point condition holds with strict inequality, we have $w(uv) > 0$. Assume $T^*_{\mathbf{D}} \neq \hat{T}$. Then $\hat{\tau}_{13} + \hat{\tau}_{24} < \hat{\tau}_{12} + \hat{\tau}_{34}$, where $\hat{\tau}_{ij}$ denotes the path-length between leaves i and j in $\hat{T}$. By additivity, $d_{13} + d_{24} = w(1u) + w(u3) + w(2v) + w(4v)$, and $d_{12} + d_{34} = w(1u) + w(uv) + w(2v) + w(3u) + w(uv) + w(4v)$. Since $w(uv) > 0$, it follows that $d_{13} + d_{24} < d_{12} + d_{34}$, contradicting the assumption.

If instead $d_{12} + d_{34} = d_{13} + d_{24} = d_{14} + d_{23}$, then $w(uv) = 0$, hence every $T \in \mathcal{T}$ is additive. By Remark 1, each such T is also an optimal solution of the BMEP.

Now, suppose that the statement holds for $n = r > 4$, with r integer. We show that it also holds for $n = r + 1$.

Let x be the new element to be added to the additive tree associated with the matrix $\mathbf{D}^x$ (note that $\mathbf{D}^x$ is additive by construction). Assume that $T^*_{\mathbf{D}} \neq \hat{T}$, and consider any leaf $y \neq x$. Let y' and y'' denote the vertices adjacent to y in $T^*_{\mathbf{D}}$ and $\hat{T}$, respectively. Then the two UBTs T' and T'', obtained by removing y and contracting the two edges incident to y' and y'', respectively, are distinct, since x is positioned differently in $T^*_{\mathbf{D}}$ and in $\hat{T}$.

Note that also $\mathbf{D}^y$ is additive. Hence, induction applies to $\mathbf{D}^y$. Assume further that the edge weights of T'' coincide with those of $\hat{T}$, except that the edge created by contraction in T'' has weight equal to the sum of the two contracted edges of $\hat{T}$.

By Proposition 3.3, the optimal solution for $\mathbf{D}^y$ is $T^*_{\mathbf{D}^y}$, which coincides with T' by construction. Moreover, since $d_{ij} = w(P^{\hat{T}}_{ij})$ for all $i, j \in \Gamma \setminus \{y\}$ and $w(P^{T''}_{ij}) = w(P^{\hat{T}}_{ij})$ by construction, it follows that T'' is the additive tree associated with $\mathbf{D}^y$. By the induction hypothesis, $T' = T''$, contradicting the assumption that $T^*_{\mathbf{D}} \neq \hat{T}$. □

4. Upper and lower bounds on the optimal value of the BMEP

Consider a dissimilarity matrix $\mathbf{D}$, and let L^-_i and L^+_i denote the costs of the least and most expensive trees nearly rooted in i (i.e., all d_{pq} with either $p \neq i$ or $q \neq i$ are considered to be zero), respectively. Then, the following result holds:

Proposition 4.1. *The objective function of the BMEP satisfies*

$$d^- \cdot n/2 \leq \sum_{i\in\Gamma} L^-_i \leq L_{\mathbf{D}}(T) \leq \sum_{i\in\Gamma} L^+_i \leq d^+ \cdot n/2$$

for any $T \in \Theta_n$.

Proof. We first observe that, by definition of L^-_i and Kraft's equality, $d^-/2 = d^- \cdot (1/4 + 1/8 + \ldots + 2 \cdot 1/2^{n-1}) \leq L^-_i$ holds true for every $i \in \Gamma$. Moreover, by Proposition 3.2, we have that $L^-_i \leq \sum_j d_{ij} 2^{-\tau_{ij}}$, hence $\sum_i L^-_i \leq \sum_i \sum_j d_{ij} 2^{-\tau_{ij}} = L_{\mathbf{D}}(T)$ holds true for every UBT T in Γ. Similarly, we have that $d^+/2 \geq L^+_i \geq \sum_j d_{ij} 2^{-\tau_{ij}}$. By summing with respect to i, the statement follows. □

Note that the inequalities in Proposition 4.1 are tight, as both the equality $d^- \cdot n/2 = \sum_i L^-_i$ and $d^+ \cdot n/2 = \sum_i L^+_i$ hold if and only if $d^- = d^+$ (and the solution of this case is claimed in Proposition 3.1). In addition, Proposition 4.1 implies that the error between any feasible solution and the optimal one grows linearly with respect to the size of the instance. In particular, if the difference d^- and d^+ is negligible, the quality of any feasible solution can be regarded as consistently good.

5. On the approximation of practical instances of the BMEP

Given any dissimilarity matrix $\mathbf{D} = \{d_{ij}\}$, with $d_{ij} > 0$, for all distinct $i, j \in \Gamma$, define the following two differential quantities:

$$\delta = \max_{i,j,k\in\Gamma} (d_{ik} - d_{ij} - d_{jk})$$

and

$$\Delta = \frac{2}{n} \sum_{s\in\Gamma} (L^+_s - 2L^-_s).$$

In addition, define

$$\varphi = \frac{\delta}{2\sum_{i\in\Gamma} L^-_i}$$

and

$$\rho = \frac{\sum_{i\in\Gamma} L^+_i}{\sum_{i\in\Gamma} L^-_i}.$$

Then, the following proposition holds:

Proposition 5.1. *The BMEP admits a polynomial-time approximation algorithm with an approximation ratio of at most*

$$\begin{cases} 2+\varphi\, n, & \text{if } \delta < \Delta, \\ \rho, & \text{otherwise.} \end{cases}$$

PROOF. If $\delta \geq \Delta$, the statement trivially follows from Proposition 4.1. We then consider the case in which $\delta < \Delta$. Observe that in this case $2+\varphi\, n < \rho$ by definition of δ and Δ, i.e., the approximation ratio $2+\varphi\, n$ is not trivial to obtain. Define a new instance of the BMEP $\mathbf{D}' = \{d'_{ij}\}$ as:

$$d'_{ij} := d_{ij} + \delta, \quad \text{for all distinct } i, j \in \Gamma$$

and $d'_{ij} := 0$ otherwise. By definition of δ, $\mathbf{D}'$ is metric. Indeed, the following relationship holds

$$\begin{aligned} d'_{ij} + d'_{jk} &= (d_{ij}+\delta) + (d_{jk}+\delta) = \\ &= d_{ij} + d_{jk} + 2\delta \geq \\ &d_{ik} + \delta \geq d_{ik} + \delta = d'_{ik} \end{aligned}$$

for all distinct $i, j, k \in \Gamma$. Now, we prove that

$$T^*_{\mathbf{D}} = T^*_{\mathbf{D}'}$$

Indeed, with respect to $\mathbf{D}'$ the following chain of equalities holds for any UBT $T \in \Theta_n$:

$$\begin{aligned} L(T) &= \sum_{i\in\Gamma}\sum_{j\in\Gamma\setminus\{i\}} d'_{ij} 2^{-\tau_{ij}} = \\ &= \sum_{i\in\Gamma}\sum_{j\in\Gamma\setminus\{i\}} (d_{ij}+\delta) 2^{-\tau_{ij}} \\ &= \sum_{i\in\Gamma}\sum_{j\in\Gamma\setminus\{i\}} d_{ij} 2^{-\tau_{ij}} + \delta \sum_{i\in\Gamma}\sum_{j\in\Gamma\setminus\{i\}} 2^{-\tau_{ij}} \\ &= \sum_{i\in\Gamma}\sum_{j\in\Gamma\setminus\{i\}} d_{ij} 2^{-\tau_{ij}} + \delta \sum_{i\in\Gamma}\left(\sum_{j\in\Gamma} 2^{-\tau_{ij}}\right) \\ &= \sum_{i\in\Gamma}\sum_{j\in\Gamma\setminus\{i\}} d_{ij} 2^{-\tau_{ij}} + \frac{n}{2}\delta \end{aligned} \tag{9}$$

where the last equality follows from Proposition 3.1. Hence, $T^*_{\mathbf{D}}$ is also optimal for $\mathbf{D}'$ and has value

$$L_{\mathbf{D}'}(T^*_{\mathbf{D}}) = L_{\mathbf{D}}(T^*_{\mathbf{D}}) + \frac{n}{2}\delta.$$

Now, because $\mathbf{D}'$ is metric, by Proposition 2.1, it is possible to compute in polynomial time a UBT $T' \in \Theta_n$ for $\mathbf{D}'$ such that $L_{\mathbf{D}'}(T') \leq 2 \cdot L_{\mathbf{D}'}(T^*_{\mathbf{D}})$, i.e.,

$$L_{\mathbf{D}'}(T') \leq 2\left(L_{\mathbf{D}}(T^*_{\mathbf{D}}) + \frac{n}{2}\delta\right). \tag{10}$$

By Proposition 4.1, we get

$$L_{\mathbf{D}}(T^*_{\mathbf{D}}) \geq \sum_{i\in\Gamma} L_i^- \tag{11}$$

and by combining (10) and (11), we get

$$\frac{L_{\mathbf{D}}(T')}{L_{\mathbf{D}}(T^*_{\mathbf{D}})} \leq 2 + \frac{\frac{n}{2}\delta}{L_{\mathbf{D}}(T^*_{\mathbf{D}})} \leq 2 + \frac{\frac{n}{2}\delta}{\sum_{i\in\Gamma} L_i^-} = 2+\varphi\, n.$$

As $\mathbf{D}'$ can be constructed in $\mathcal{O}(n^2)$ and T' can be computed in $\mathcal{O}(n^3)$ [14], the statement follows. □

Proposition 5.1 establishes that the BMEP admits an efficient approximation whenever the dissimilarity matrix $\mathbf{D}$ has strictly positive off-diagonal entries and a bounded ratio ρ. This assumption is particularly meaningful in practice, since dissimilarity matrices of this kind naturally arise from biological data. Indeed, standard models of molecular evolution [5, 13, 20] typically yield strictly positive dissimilarities between distinct molecular sequences. Moreover, the numerical values of d_{ij} are inherently subject to finite precision (stemming from both data acquisition and floating-point representation), which effectively bounds ρ by a polynomial function of the input size. Furthermore, when ρ grows exponentially, if the corresponding set of items/taxa can be partitioned to expose a block-structured form of ρ, a good-quality solution can still be obtained by leveraging Proposition 3.2.

To conclude this section, we investigate the problem of providing a meaningful measure of algorithmic quality for the BMEP in contexts where no constant-factor approximation is possible. To this end, we recall that for a minimization problem, the *differential approximation ratio* is defined as

$$\frac{c_{\text{apx}} - OPT}{c_{\text{worst}} - OPT},$$

where c_{apx}, c_{worst}, and OPT denote, respectively, the value of the solution found by an approximation algorithm, the maximum value achievable in the set of feasible solutions to the problem, and the value of the optimal solution. Interestingly, although Proposition 5.1 does not yield tight bounds for Fiorini and Joret [14] approximation algorithm, the following general elementary result still holds:

Proposition 5.2. *The differential approximation ratio of the BMEP is the same of the metric BMEP.*

PROOF. Let $\mathbf{D}$ denote any instance of the BMEP. Consider the matrix $\mathbf{D}' = \{d'_{ij}\}$, defined by $d'_{ij} = d_{ij} + d^+$ for every $i \neq j$. It is easy to see that $\mathbf{D}'$ is a metric instance of the BMEP. By Proposition 3.1, $L_{\mathbf{D}'}(T) = L_{\mathbf{D}}(T) + d^+ \cdot n/2$, for every $T \in \Theta_n$. Thus, if T' is the UBT obtained from Fiorini and Joret [14] approximation algorithm and $T'' \in \operatorname{argmax}_{T\in\Theta_n}\{L_{\mathbf{D}}(T)\}$, then we have

$$\frac{L_{\mathbf{D}'}(T') - L_{\mathbf{D}'}(T^*_{\mathbf{D}})}{L_{\mathbf{D}'}(T'') - L_{\mathbf{D}'}(T^*_{\mathbf{D}})} = \frac{L_{\mathbf{D}}(T') - L_{\mathbf{D}}(T^*_{\mathbf{D}})}{L_{\mathbf{D}}(T'') - L_{\mathbf{D}}(T^*_{\mathbf{D}})}.$$

□

The last proposition leaves open the possibility to provide better upper bounds for Fiorini and Joret [14] approximation algorithm whenever the ratio ρ is bounded.

A simple example. Consider the matrix

$$\mathbf{D} = \begin{bmatrix} 0.000 & 5.691 & 3.872 & 4.268 & 0.163 & 5.610 \\ 5.691 & 0.000 & 6.401 & 0.051 & 2.812 & 2.721 \\ 3.872 & 6.401 & 0.000 & 4.200 & 3.098 & 2.244 \\ 4.268 & 0.051 & 4.200 & 0.000 & 2.796 & 3.396 \\ 0.163 & 2.812 & 3.098 & 2.796 & 0.000 & 4.502 \\ 5.610 & 2.721 & 2.244 & 3.396 & 4.502 & 0.000 \end{bmatrix}.$$

A direct computation shows that $2 + \varphi\, n = 3.226$, with $\delta = 2.717$, while $\rho = 4.08$. In particular, the optimum (computed with a branch and cut algorithm) is 7.39, while the cost of the solution obtained by the Fiorini-Joret algorithm is 8.343, and hence, the actual ratio is 1.129. Note that the differential approximation ratio is 0.1906, where we used $\max_{T \in \Theta_n} L_{\mathbf{D}}(T) = 12.39$ to compute it.

Acknowledgments

The first and the third authors acknowledge support from the Belgian National Fund for Scientific Research (FNRS) via the grant FNRS PDR 40007831. This first author also acknowledges support from the Fondation Louvain, via the grant "COALESCENS".

References

[1] R. Aringhieri, D. Catanzaro, and M. Di Summa. Optimal solutions for the balanced minimum evolution problem. *Computers and Operations Research*, 38:1845–1854, 2011.

[2] P. Buneman. The recovery of trees from measure of dissimilarities. In F. R. Hodson, D. G. Kendall, and P. Tautu, editors, *Archaeological and Historical Science*, pages 387–395. Edinburgh University Press, Edinburgh, UK, 1971.

[3] D. Catanzaro. The minimum evolution problem: Overview and classification. *Networks*, 53, 2009.

[4] D. Catanzaro. Estimating phylogenies from molecular data. In R. Bruni, editor, *Mathematical approaches to polymer sequence analysis and related problems*, pages 149–176. Springer, NY, 2011.

[5] D. Catanzaro, R. Pesenti, and M. Milinkovitch. A non-linear optimization procedure to estimate distances and instantaneous substitution rate matrices under the GTR model. *Bioinformatics*, 22(6):708–715, 2006.

[6] D. Catanzaro, M. Labbé, R. Pesenti, and J. J. Salazar-Gonzáles. The balanced minimum evolution problem. *INFORMS Journal on Computing*, 24(2):276–294, 2012.

[7] D. Catanzaro, M. Frohn, and R. Pesenti. An information theory perspective on the balanced minimum evolution problem. *Operations Research Letters*, 48(3):362–367, 2020.

[8] D. Catanzaro, R. Pesenti, and L. A. Wolsey. On the Balanced Minimum Evolution polytope. *Discrete Optimization*, 36:1–33, 2020.

[9] D. Catanzaro, M. Frohn, O Gascuel, and R. Pesenti. A tutorial on the balanced minimum evolution problem. *European Journal of Operational Research*, 300(1):1–19, 2022.

[10] D. Catanzaro, R. Pesenti, A. Sapucaia, and L. Wolsey. Optimizing over path-length matrices of unrooted binary trees. *Mathematical Programming*, pages 1–53, 2025.

[11] M. A. Cueto and F. A. Matsen. Polyhedral geometry of phylogenetic rogue taxa. *Bulletin of Mathematical Biology*, 73(6):1202–1226, 2011.

[12] R. Desper and O. Gascuel. Theoretical foundations of the balanced minimum evolution method of phylogenetic inference and its relationship to the weighted least-squares tree fitting. *Molecular Biology and Evolution*, 21(3):587–598, December 2004.

[13] J. Felsenstein. *Inferring phylogenies*. Sinauer Associates, Sunderland, MA, 2004.

[14] S. Fiorini and G. Joret. Approximating the balanced minimum evolution problem. *Operations Research Letters*, 40(1):31–35, January 2012.

[15] S. Forcey, L. Keefe, and W. Sands. Facets of the balanced minimal evolution polytope. *Journal of Mathematical Biology*, 73(2):447–468, December 2015.

[16] S. Forcey, L. Keefe, and W. Sands. Split-facets for balanced minimal evolution polytopes and the permutoassociahedron. *Bulletin of Mathematical Biology, in press*, 79:975–994, March 2017.

[17] O. Gascuel. *Mathematics of evolution and phylogeny*. Oxford University Press, New York, NY, 2005.

[18] O. Gascuel and M. Steel. Neighbor-joining revealed. *Molecular Biology and Evolution*, 23(11):1997–2000, 2006.

[19] D. C. Haws, T. L. Hodge, and R. Yoshida. Optimality of the neighbor joining algorithm and faces of the balanced minimum evolution polytope. *Bulletin of Mathematical Biology*, 73:2627–2648, March 2011.

[20] R. D. M. Page and E. C. Holmes. *Molecular Evolution: A Phylogenetic Approach*. Blackwell Science, Oxford, UK, 1998.

[21] F. Pardi. *Algorithms on Phylogenetic Trees*. PhD thesis, University of Cambridge, UK, 2009.

[22] C. Semple and M. A. Steel. Cyclic permutations and evolutionary trees. *Advances in Applied Mathematics*, 32(4):669–680, May 2004.

[23] M. S. Waterman, T. F. Smith, M. Singh, and W. A. Beyer. Additive evolutionary trees. *Journal of Theoretical Biology*, 64(2):199–213, 1977.