# Characterizing Path-Length Matrices of Unrooted Binary Trees

Daniele Catanzaro[1], Raffaele Pesenti[2], and Roberto Ronco[3]

[1]Center for Operations Research and Econometrics, Université Catholique de Louvain, Louvain-la-Neuve, Belgium, ✉ daniele.catanzaro@uclouvain.be
[2]Department of Management, Ca' Foscari University of Venice, Venice, Italy, ✉ pesenti@unive.it
[3]Institute of Marine Engineering (INM), National Research Council of Italy (CNR), Genoa, Italy, ✉ roberto.ronco@cnr.it

Given a set $[n] = \{1, \ldots, n\}$, an *unrooted binary tree* (UBT) is a tree having internal vertices of degree 3. UBTs play an important role in distance-based phylogenetics and in related optimization problems. A convenient encoding of a UBT $T$ consists of its associated *path-length matrix* (PLM) $\boldsymbol{\tau}$, where $\tau_{ii} = 0$ for each $i \in [n]$ and $\tau_{ij}$ equals the number of edges on the unique path between leaves $i$ and $j$ in $T$ for each distinct $i, j \in [n]$.

Let $\Theta_n$ denote the set of PLMs associated with all UBTs on $n$ labeled leaves. A key motivation to search for characterizations of $\Theta_n$ is the *Balanced Minimum Evolution Problem* (BMEP) [2]. Given a symmetric dissimilarity matrix $\mathbf{D} = [d_{ij}]_{i,j\in[n]}$, the BMEP requires finding a UBT associated with a PLM $\boldsymbol{\tau} \in \Theta_n$ that minimizes:

$$\sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} d_{ij}\, 2^{-\tau_{ij}}.$$

The BMEP is NP-hard, and the dependence on $2^{-\tau_{ij}}$ makes its objective function nonlinear. Hence, integer linear programming (ILP) approaches rely on extended formulations of $\Theta_n$.

A classical structural condition for PLMs of UBTs is given by *Kraft's equalities*:

$$\sum_{\substack{j=1 \\ j\neq i}}^{n} 2^{-\tau_{ij}} = \frac{1}{2}. \tag{1}$$

A recent work has derived a characterization of $\Theta_n$ for any $n \geq 3$ by using Kraft's equalities with *strengthened* versions of *triangle inequalities*:

$$\tau_{ij} + \tau_{jk} - \tau_{ik} \geq 2 \qquad \text{for all distinct } i, j, k \in [n]. \tag{2}$$

and *Buneman's four-point conditions* [4, 1]:

$$\begin{cases} \tau_{ij} + \tau_{pq} + 2 \leq \tau_{ip} + \tau_{jq} = \tau_{iq} + \tau_{jp} \\ \tau_{ip} + \tau_{jq} + 2 \leq \tau_{ij} + \tau_{pq} = \tau_{iq} + \tau_{jp} \\ \tau_{iq} + \tau_{jp} + 2 \leq \tau_{ij} + \tau_{pq} = \tau_{ip} + \tau_{jq} \end{cases} \qquad \text{for all distinct } i, j, p, q \in [n]$$

Such conditions are disjunctive, leading to computationally demanding ILP formulations. In this work, we provide an alternative route to characterize $\Theta_n$ for $n \leq 11$ by replacing the Buneman' strong four-point condition with a single nonlinear equality, called the *UBT-manifold condition*:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \tau_{ij}\, 2^{-\tau_{ij}} = 2n - 3. \tag{4}$$

This condition expresses a global invariant of UBTs, by coupling each path-length $\tau_{ij}$ with its density $2^{-\tau_{ij}}$ and relating their sum over each $i, j \in [n]$ to the number of edges of a UBT, given by $2n - 3$.

Specifically, we show that, given a symmetric integer-valued matrix $\boldsymbol{\tau}$ with null diagonal of order $n$ such that $3 \leq n \leq 11$, conditions (1), (2) and (4) are necessary and sufficient for $\boldsymbol{\tau}$ to be in $\Theta_n$ [3]. We also show that these conditions can be weakened for $3 \leq n \leq 8$. For $n > 11$, these conditions are

generally not sufficient. However, we verified computationally that, for $n = 12$, sufficiency is achieved by also including the *circular-order inequalities*:

$$\sum_{j\in[n-2]} \tau_{i_j,i_{j+1}} + \tau_{i_{n-1},i_1} \geq 4n - 8. \tag{5}$$

where $[i_1, i_2, \ldots, i_{n-1}]$ is a sequence of distinct elements in $[n]$. Determining the precise role of conditions (5) is still open.

The characterization based on the UBT-manifold condition leads to a compact non-disjunctive ILP formulation for the BMEP through a suitable linearization for $n \leq 11$. Conditions (1) and (4) can be linearized, e.g., with the use of binary variables $x_{ijl}$ for each $i, j, l \in [n]$, where $x_{ijl} = 1$ if and only if $\tau_{ij} = l$ [3]. To evaluate the computational effectiveness of the formulation, we compared it with an ILP formulation based on Buneman' strong four-point conditions on the benchmark instances in the literature on the BMEP. The computational results show that the formulation based on the UBT-manifold condition tightens the root-node relaxation substantially, translating into smaller branch-and-bound trees and faster solution times. These improvements largely persist even when the formulation based on the UBT-manifold condition is compared with a non-disjunctive formulation based on Buneman' strong four-point conditions that retains only those four-point constraints that are active at optimality.

# References

[1] P. Buneman. A note on the metric properties of trees. *Journal of Combinatorial Theory, Series B*, 17:48–50, 1974.

[2] Daniele Catanzaro, Martine Labbé, Raffaele Pesenti, and Juan-José Salazar-González. The balanced minimum evolution problem. *INFORMS Journal on Computing*, 24:276–294, 2012.

[3] Daniele Catanzaro, Raffaele Pesenti, and Roberto Ronco. Characterizing path-length matrices of unrooted binary trees. LIDAM Discussion Paper CORE 2024/28, UCLouvain, Center for Operations Research and Econometrics (CORE), 2024.

[4] Daniele Catanzaro, Raffaele Pesenti, A. Sapucaia, and L. Wolsey. Optimizing over path-length matrices of unrooted binary trees. *Mathematical Programming*, 2025.