CHARACTERIZING PATH-LENGTH MATRICES OF UNROOTED BINARY TREES

Daniele Catanzaro, Raffaele Pesenti, Roberto Ronco

LIDAM Discussion Paper CORE
2024 / 28

UCLouvain

LiDAM
CORE | Louvain Institute of Data Analysis and Modeling in economics and statistics

CORE

Voie du Roman Pays 34, L1.03.01

B-1348 Louvain-la-Neuve

Tel (32 10) 47 43 04

Email: immaq-library@uclouvain.be

https://uclouvain.be/en/research-institutes/lidam/core/core-discussion-papers.html

CHARACTERIZING PATH-LENGTH MATRICES OF UNROOTED BINARY TREES

DANIELE CATANZARO*, RAFFAELE PESENTI†, AND ROBERTO RONCO‡

Abstract. We extend some recent results on the necessary and sufficient conditions that a symmetric integer matrix of order $n \geq 3$ must satisfy to encode the *Path-Length Matrix* (PLM) of a *Unrooted Binary Tree* (UBT) with n leaves. This problem is at the core of the combinatorics of the Balanced Minimum Evolution Problem, a $\mathcal{NP}$-hard problem much studied in the literature on molecular phylogenetics. We show that, for any natural $3 \leq n \leq 11$, a reduced set of known conditions, excluding Buneman' strong four-point conditions, is both necessary and sufficient to characterize PLMs of UBTs. In addition, we present a second and more general characterization based solely on linear conditions derived from the topological properties of UBTs.

Key words. combinatorial optimization, tree realization, balanced minimum evolution, unrooted binary trees, path-length matrices, Kraft conditions, Buneman's four-point conditions.

MSC codes. 90C27, 90C10

1. Introduction.

Given a natural $n \geq 3$, consider a *semimetric* $\boldsymbol{\tau} = \{\tau_{ij}\}$ for the set $[n] = \{1, 2, \dots, n\}$, i.e., a square matrix of order n, whose entries satisfy the following conditions:

$$\text{(1.1)} \qquad \text{(separation)} \qquad \tau_{ii} = 0 \qquad \forall\ i \in [n]$$

$$\text{(1.2)} \qquad \text{(positivity)} \qquad \tau_{ij} > 0 \qquad \forall\ i, j \in [n] : i \neq j$$

$$\text{(1.3)} \qquad \text{(symmetry)} \qquad \tau_{ij} = \tau_{ji} \qquad \forall\ i, j \in [n] : i < j.$$

We say that $\boldsymbol{\tau}$ is a *tree realization* if there exists a pair (T, w), referred to as a *fitting tree*, such that

(i) $T = (V, E)$ is an acyclic connected graph having V as node-set, E as edge-set, and $L = [n] \subset V$ as set of leaves;

(ii) $w = \{w_e,\ e \in E\}$ is a set of strictly positive weights associated with the edges of T such that

$$\text{(1.4)} \qquad \sum_{e \in P_{ij}^T} w_e = \tau_{ij} \qquad \forall\ i, j \in [n] : i < j,$$

where P_{ij}^T is the unique path in T connecting the leaves i and j.

For example, Figure 1 shows a semimetric $\boldsymbol{\tau}$ for the set [5] and a possible tree realization associated with it.

We say that a semimetric $\boldsymbol{\tau}$ for $[n]$ is the realization of an *Unrooted Binary Tree* (UBT) if the fitting-tree (T, w) also satisfies the following conditions:

(iii) T is *unrooted* and *binary*, i.e., its internal nodes have degree 3;

(iv) the weights w_e associated with the edges of T are all equal to 1.

For example, Figure 2 shows a semimetric for the set [5] and its associated UBT realization.

*Center for Operations Research and Econometrics, Université Catholique de Louvain, Voie du Roman Pays 34, 1348, Louvain-la-Neuve, Belgium (daniele.catanzaro@uclouvain.be).

†School of Management, University Ca' Foscari, San Giobbe, Cannaregio 837, I-30121 Venice, Italy (pesenti@unive.it).

‡Institute of Marine Engineering, National Research Council of Italy, Via De Marini 6, I-16149, Genoa, Italy (roberto.ronco@cnr.it).

$$\tau = \begin{pmatrix} 0 & 3 & 3 & 3 & 3 \\ 3 & 0 & 3 & 3 & 3 \\ 3 & 3 & 0 & 3 & 3 \\ 3 & 3 & 3 & 0 & 3 \\ 3 & 3 & 3 & 3 & 0 \end{pmatrix}$$

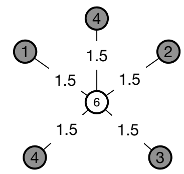

Fig. 1: An example of a semimetric $\boldsymbol{\tau}$ that is a tree realization.

As the edge weights of a fitting UBT (T, w) are all equal to 1, in what follows, we shall omit w for simplicity and denote the fitting UBT as T. We also observe that if $\boldsymbol{\tau}$ is the realization of a UBT T, then its generic entry τ_{ij} constitutes the *path-length* of P_{ij}^T, i.e., the number of edges in P_{ij}^T. For example, by referring to Figure 2, the entry $\tau_{12} = 2$ indicates that the path from leaf 1 to leaf 2 in T consists of two edges. In the light of this observation, in what follows we shall refer to a semimetric $\boldsymbol{\tau}$ that is a realization of a UBT T as the *Path-Length Matrix* (PLM) of T.

The problem of identifying the necessary and sufficient conditions that a semimetric must satisfy to be a tree realization is known in the literature as the *Tree Realization Problem* [1, 2, 8, 10, 11, 12, 13, 15]. A classical result provided by Buneman [2] in 1974 states that

Proposition 1. *A semimetric $\boldsymbol{\tau}$ for $[n]$ has a tree realization if and only if its entries satisfy*

$$\tau_{ij} + \tau_{pq} \leq \max\{\tau_{iq} + \tau_{jp}, \tau_{ip} + \tau_{jq}\} \tag{1.5}$$

together with least one of the following equalities

$$\begin{cases} \tau_{iq} + \tau_{jp} = \tau_{ip} + \tau_{jq} \\ \tau_{iq} + \tau_{jp} = \tau_{ij} + \tau_{pq} \\ \tau_{ip} + \tau_{jq} = \tau_{ij} + \tau_{pq} \end{cases} \tag{1.6}$$

for all $i, j, p, q \in [n]$, at least three of which are distinct. Furthermore, such a realization is unique among the trees whose internal nodes have degree greater than or equal to 3.

Condition (1.5) can be equivalently split into the following two families of inequalities

$$\tau_{ik} + \tau_{kj} \geq \tau_{ij} \qquad \forall \text{ distinct } i, j, k \in [n] \tag{1.7}$$

$$\tau_{ij} + \tau_{pq} \leq \max\{\tau_{iq} + \tau_{jp}, \tau_{ip} + \tau_{jq}\} \qquad \forall \text{ distinct } i, j, p, q \in [n] \tag{1.8}$$

that are usually referred to as the *triangle inequalities* and the *additive* (or *Buneman's four-point*) conditions [1], respectively.

$$\boldsymbol{\tau} = \begin{pmatrix} 0 & 2 & 3 & 4 & 4 \\ 2 & 0 & 3 & 4 & 4 \\ 3 & 3 & 0 & 3 & 3 \\ 4 & 4 & 3 & 0 & 2 \\ 4 & 4 & 3 & 2 & 0 \end{pmatrix}$$

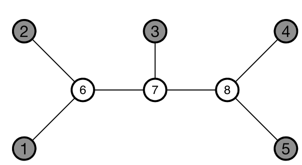

Fig. 2: An example of a semimetric $\boldsymbol{\tau}$ that is the realization of a UBT (edge weights have been omitted).

	CCHF	Ebola	Lassa	COV2	HIV1	HIV2	Nipah
CCHF	0.000	0.005	0.009	0.014	0.011	0.011	0.008
Ebola	0.005	0.000	0.008	0.012	0.016	0.018	0.004
Lassa	0.009	0.008	0.000	0.015	0.033	0.031	0.015
COV2	0.014	0.012	0.015	0.000	0.034	0.041	0.014
HIV1	0.011	0.016	0.033	0.034	0.000	0.003	0.014
HIV2	0.011	0.018	0.031	0.041	0.003	0.000	0.020
Nipah	0.008	0.004	0.015	0.014	0.014	0.020	0.000

	CCHF	Ebola	Lassa	COV2	HIV1	HIV2	Nipah
CCHF	0	3	5	2	6	6	4
Ebola	3	0	4	3	5	5	3
Lassa	5	4	0	5	3	3	3
CoV2	2	3	5	0	6	6	4
HIV1	6	5	3	6	0	2	4
HIV2	6	5	3	6	2	0	4
Nipah	4	3	3	4	4	4	0

Fig. 3: An example of an instance of the BMEP consisting of the input metric $\mathbf{D}$ (shown at the top) and the corresponding optimal solution expressed in terms of path-lengths τ_{ij} between taxa (shown at the bottom). The matrix $\mathbf{D}$ is taken from [3] and its entries encode the *evolutionary distances* between the Crimean-Congo Hemorrhagic Fever (CCHF) orthonairovirus, Ebolavirus, the Lassa mammarenavirus, the Severe Acute Respiratory Syndrome Coronavirus 2 (CoV2), the Human Immunodeficiency Viruses (HIVs) 1 and 2, and the Nipah virus. The optimal solution encoded by the path-length matrix above is shown in Figure 4.

Semimetrics that satisfy (1.7) are referred to as *metrics*. Metrics that additionally satisfy (1.8) are called *additive matrices* [12, 15]. Thus, Proposition 1 can be succinctly restated by saying that $\boldsymbol{\tau}$ is a tree realization if and only if it is additive.

Unfortunately, the additive property proves insufficient in the context of a UBT realization. For example, the matrix shown in Figure 1 is additive, but it is easy to see that there is no fitting UBT for it. Characterizing the mathematical conditions that ensure the existence of a UBT realization for matrices of order n has been a long-standing open problem at the core of the combinatorics of the *Balanced Minimum Evolution Problem* (BMEP), a $\mathcal{NP}$-hard problem defined over the set Θ_n of the PLMs induced by UBTs with n leaves and much studied in the literature on molecular phylogenetics [3]. Given a set $[n]$ of $n \geq 3$ molecular sequences (e.g., DNA, RNA, codon sequences, or whole genomes), commonly referred to as *taxa*, and a semimetric $\mathbf{D} = \{d_{ij}\}$ for $[n]$, whose generic entry d_{ij} encodes a measure of similarity between the pair of distinct taxa $i, j \in [n]$, the BMEP requires to find a UBT T with n leaves whose PLM $\boldsymbol{\tau} = \{\tau_{ij}\}$ minimizes the following objective function

$$\sum_{i\in[n]} \sum_{j\in[n]\setminus\{i\}} d_{ij} 2^{-\tau_{ij}}. \tag{1.9}$$

Figures 3 and 4 show an example of an instance of the BMEP and of the corresponding optimal solution.

The study of the convex hull of the solutions to the BMEP, motivated by the development of exact solution algorithms of practical use, involves the characterization of the set Θ_n. The only characterization of Θ_n known in the literature so far, introduced in [6] and recalled in Section 2 for the sake of completeness, combines strengthened versions of (1.7) and (1.8) with n non-linear equalities, called *Kraft's equalities*. These last conditions capture in closed form both the degree constraint

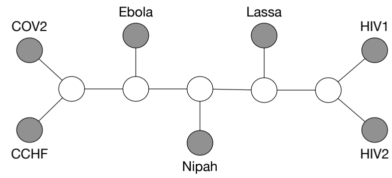

Fig. 4: The unrooted binary tree corresponding to the path-length matrix shown in Figure 3. This tree is provably optimal for the instance of the BMEP encoded by the input distance matrix shown in Figure 3.

and the connectivity property of UBTs. As observed in [6], finding an effective way to impose all of these conditions in a valid *Integer Linear Programming* (ILP) formulation for the BMEP is non-trivial due to the non-linearity of Kraft's equalities and the presence of disjunctive constraints related to (strengthened versions of) (1.8). Thus, a question that naturally arises is whether there might exist alternative characterizations of a UBT realization that involve sets of combinatorial conditions different from those identified in [6]. In this article, we investigate this question. First, we present a characterization for $n \leq 11$ that takes advantage of known combinatorial properties of UBT that do not involve Buneman's four-point conditions (1.8). Secondly, we describe a general characterization for Θ_n based on topological arguments, by providing a set of non-disjunctive conditions that are linear in the space of path-lengths.

The article is organized as follows. In Section 2, we introduce some notation, definitions, and known properties of trees that recur throughout the article. In Section 3, we present the characterization of a UBT realization that proves valid for $n \leq 11$. Finally, in Section 4, we describe a more general characterization of Θ_n.

2. Notation, definitions, and properties of path-length matrices. Given two integers a, b, with $a \leq b$, we denote by $[a, b]$ the subset of integers c such that $a \leq c \leq b$. Consistently with the notation used in Section 1, we use $[b]$ for $[a, b]$ when $a = 1$.

We denote by $\mathcal{U}_n$ the set of UBTs with $n \geq 3$ leaves and, for a fixed UBT $T \in \mathcal{U}_n$, we denote by L is set of leaves. As already mentioned for generic trees, we assume that L is labeled with the elements of $[n]$.

We define a *sequence* as a collection of non-negative and not necessarily different natural numbers sorted in non-decreasing order. In particular, given a UBT $T \in \mathcal{U}_n$ and a leaf $i \in L$, we define the *path-length sequence* $\vec{\boldsymbol{\tau}}_i = \text{sort}([\tau_{ij} : j \in L \setminus \{i\}])$ as the sequence of the path-lengths relative to the $n-1$ (unique) paths in T connecting i to the remaining leaves in L. The path-length sequence $\vec{\boldsymbol{\tau}}_i$ describes T as seen from (or, equivalently, as *rooted in*) i. For example, the path-length sequence $\vec{\boldsymbol{\tau}}_1 = \{2, 3, 4, 4\}$ describes the UBT shown in Figure 2 as rooted in leaf 1. Observe that the relationship between path-length sequences and rooted UBTs is not bijective. For example, Figure 5 shows that there exist three possible rooted UBTs that can be associated with the path-length sequence $\vec{\boldsymbol{\tau}}_1 = \{3, 3, 4, 4, 4, 4\}$. In contrast, it is easy to see that the relationship between the PLMs in Θ_n and the UBTs in $\mathcal{U}_n$ is bijective. Observe also that, by definition of a path-length sequence, the entries of $\vec{\boldsymbol{\tau}}_i$ correspond to the non-diagonal entries of the i-th row of the PLM $\boldsymbol{\tau}$ associated with T.

Let $\vec{\Phi}_i = \{\vec{\boldsymbol{\tau}}_i\}$ denote the set of the path-length sequences associated with the leaf $i \in [n]$. We say that two path-length sequences $\vec{\boldsymbol{\tau}}_i \in \vec{\Phi}_i$ and $\vec{\boldsymbol{\tau}}_j \in \vec{\Phi}_j$ are *congruent* with the same UBT $T \in \mathcal{U}_n$ if T has i and j as leaves. For example, $\vec{\boldsymbol{\tau}}_1 = [3, 3, 4, 4, 4, 4]$

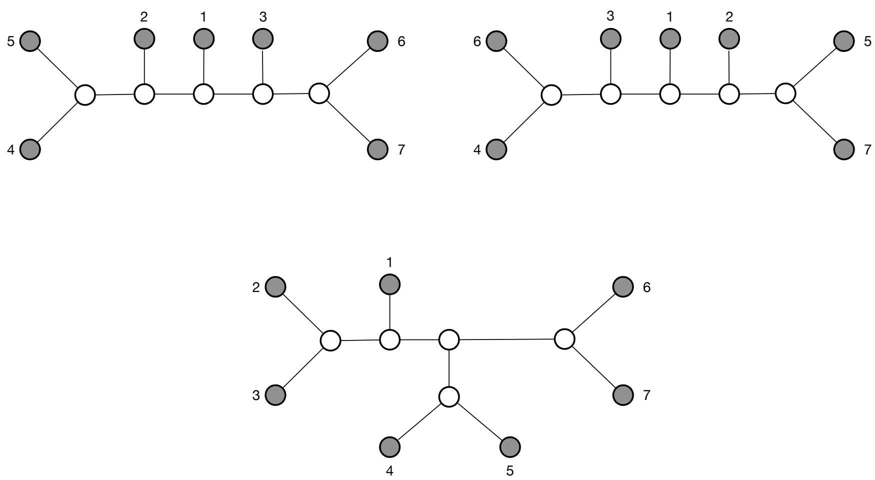

Fig. 5: An example of the three distinct UBTs that can be constructed from the path-length sequence $\vec{\boldsymbol{\tau}}_1 = [3, 3, 4, 4, 4, 4]$.

and $\vec{\boldsymbol{\tau}}_3 = [3, 4, 5, 5, 3, 3]$ are congruent with the UBT shown on the top left of Figure 5. In contrast, $\vec{\boldsymbol{\tau}}_1 = [3, 3, 4, 4, 4, 4]$ and $\vec{\boldsymbol{\tau}}_3 = [4, 4, 2, 5, 6, 6]$ are congruent with no UBT in Θ_n.

The following result characterizes the set $\vec{\Phi}_i$.

PROPOSITION 2. *An integer sequence $\vec{\mathbf{s}} = \{\vec{s}_1, \vec{s}_2, \ldots, \vec{s}_{n-1}\}$ belongs to $\vec{\Phi}_i$ if and only if it satisfies the following condition:*

$$\text{(2.1)} \qquad \textit{(Kraft's equality)} \qquad \sum_{j \in [n-1]} 2^{-\vec{s}_j} = \frac{1}{2}$$

In addition, if $n \geq 3$, (2.1) implies that the entries $\vec{s}_1, \vec{s}_2, \ldots, \vec{s}_{n-1}$ are in $[2, n-1]$.

Proof. The first part of the statement (Kraft's equality) is well-known and proven in [14]. Hence, here we will only focus on proving the second part of the statement, i.e., that the entries $\vec{s}_1, \vec{s}_2, \ldots, \vec{s}_{n-1}$ are in $[2, n-1]$. We start by recalling that the entries of $\vec{\mathbf{s}}$ are sorted in non-decreasing order. Then, it is easy to verify that, for any $n \geq 3$, (2.1) does not hold if $\vec{s}_1 < 2$, since it implies $2^{-\vec{s}_1} > 2^{-1}$. Now suppose, by contradiction, that $\vec{s}_{n-1} > n-1$. For each $k \in [n-2]$, observe that (2.1) implies the following identity:

$$\text{(2.2)} \qquad \sum_{j \in [k]} 2^{-\vec{s}_j} = \frac{1}{2} - \sum_{j \in [k+1, n-1]} 2^{-\vec{s}_j}$$

Since the left-hand side of equation (2.2) is a sum of non-increasing negative powers of 2, it is also a multiple of its smallest term $2^{-\vec{s}_k}$. Similarly, the right-hand side of (2.2) is a multiple of $2^{-\vec{s}_k}$. Since 2^{-1} is a multiple of $2^{-\vec{s}_k}$, then $\sum_{j \in [k+1, n-1]} 2^{-\vec{s}_j}$ must be a multiple of $2^{-\vec{s}_k}$ and therefore

$$\text{(2.3)} \qquad 2^{-\vec{s}_k} \leq \sum_{j \in [k+1, n-1]} 2^{-\vec{s}_j}.$$

Observe that $\sum_{j\in[n-1]} 2^{-\vec{s}_j}$ achieves its maximum when inequality (2.3) holds as an equality, i.e., when

$$\vec{s}_{n-2} = \vec{s}_{n-1} \qquad \text{and} \qquad \vec{s}_{n-(j+1)} = \vec{s}_{n-j} - 1 \quad \forall\, j \in [2, n-2]. \tag{2.4}$$

However, in this case

$$\sum_{j\in[n-1]} 2^{-\vec{s}_j} = 2^{-\vec{s}_1} + \sum_{j\in[2,n-1]} 2^{-\vec{s}_j} = 2^{1-\vec{s}_1} = 2^{1-(\vec{s}_{n-1}-(n-3))} < \frac{1}{2},$$

which contradicts (2.1). Thus, the statement follows. ∎

The following result characterizes the set Θ_n [6].

Proposition 3. *An integer semimetric $\boldsymbol{\tau}$ for $[n]$ is the realization of a UBT $T \in \mathcal{U}_n$ if and only if it satisfies the following conditions*

$$\text{(Kraft's equalities)} \qquad \sum_{j\in[n]\setminus i} 2^{-\tau_{ij}} = \frac{1}{2} \qquad \forall\, i \in [n] \tag{2.5}$$

$$\text{(strong triangle ineq.)} \qquad \tau_{ij} + \tau_{jk} - \tau_{ik} \geq 2 \qquad \forall \text{ distinct } i, j, k \in [n] \tag{2.6}$$

and precisely one of the following Buneman' strong four-point conditions*:*

$$\begin{cases} \tau_{ij} + \tau_{pq} + 2 \leq \tau_{iq} + \tau_{jp} = \tau_{ip} + \tau_{jq} \\ \tau_{iq} + \tau_{jp} + 2 \leq \tau_{ij} + \tau_{pq} = \tau_{ip} + \tau_{jq} \\ \tau_{ip} + \tau_{jq} + 2 \leq \tau_{ij} + \tau_{pq} = \tau_{iq} + \tau_{jp} \end{cases} \quad \forall \text{ distinct } i, j, p, q \in [n]. \tag{2.7}$$

Kraft's equalities (2.5) establish a correspondence between the rows of $\boldsymbol{\tau}$ and n path-length sequences. They imply that each row of a matrix $\boldsymbol{\tau} \in \Theta_n$ contains entries with values in $[2, n-1]$ and, for $n \geq 4$, at most one entry with value 2. They also imply that Θ_3 includes only the following *basic* matrix:

$$\begin{bmatrix} 0 & 2 & 2 \\ 2 & 0 & 2 \\ 2 & 2 & 0 \end{bmatrix}. \tag{2.8}$$

It is worth noting, however, that Kraft's equalities are not sufficient to ensure that the path-lengths associated with the rows of $\boldsymbol{\tau}$ are congruent with a UBT in Θ_n. Conditions (2.6) and (2.7), instead, are stronger versions of (1.7) and (1.8), and take into account the combinatorial properties of UBTs related to the connectivity constraint and the degree constraint on the internal nodes.

We call *caterpillar* any UBT in $\mathcal{U}_n$ whose internal vertices induce a path-graph. We call *cherry* any subgraph of a UBT in $\mathcal{U}_n$ induced by two of its leaves and the internal vertex adjacent to both of them. Then, the lower bound of 2 for the value of an entry of the PLM $\boldsymbol{\tau}$ of a UBT is reached in the presence of a cherry, while the upper bound is reached only if $\boldsymbol{\tau}$ encodes a caterpillar. For example, the UBT in Figure 2 is a caterpillar with $n = 5$ leaves and two cherries. One of these cherries is induced by leaves 1 and 2 and internal vertex 6, while the other is induced by leaves 4 and 5 and internal vertex 8; indeed, $\tau_{12} = \tau_{45} = 2$. The internal vertices of the UBT, i.e. 6, 7, and 8, induce a path graph, and accordingly $\tau_{14} = \tau_{15} = \tau_{24} = \tau_{25} = n - 1 = 4$.

Notation	Legend	Definition
UBT	Unrooted Binary Tree	Section 1
P_{ij}	The unique path in a UBT $T \in \mathcal{U}_n$ between nodes i and j	Section 1
τ_{ij}	The number of edges in (or *path-length* of) P_{ij}, $T \in \mathcal{U}_n$	Section 1
PLM	Path-Length Matrix	Section 1
Θ_n	Set of PLMs induced by the UBTs in $\mathcal{U}_n$	Section 1
$\mathcal{U}_n$	Set of UBTs with $n \geq 3$ leaves	Section 2
$\vec{\boldsymbol{\tau}}_i$	Path-Length Sequence	Section 2
SM	Strong metric matrix	Section 2
SA	Strong additive matrix	Section 2
KSM	Kraft strong metric matrix	Section 2
KSA	Kraft strong additive matrix	Section 2
KSM-UBT	Kraft strong metric matrix that satisfies (3.11)	Section 3

Table 1: Summary of the most important notation used in the article.

For each PLM $\boldsymbol{\tau} \in \Theta_n$ we define a square matrix of order n, denoted by $\mathbf{2}^{-\boldsymbol{\tau}}$, whose generic entry is defined as follows:

$$(\mathbf{2}^{-\boldsymbol{\tau}})_{ij} = \begin{cases} 2^{-\tau_{ij}} & \text{if } i \neq j, \\ 0 & \text{if } i = j. \end{cases} \tag{2.9}$$

Observe that $\mathbf{2}^{-\boldsymbol{\tau}}$ is doubly stochastic up to a factor 2, because Kraft's equalities (2.5) hold on the entries of $\boldsymbol{\tau}$. Thus, up to a factor 2, its rows and columns can be interpreted as *probability density sequences* [4]. In the light of this observation, we shall refer to $\mathbf{2}^{-\boldsymbol{\tau}}$ as the *Path-Length Density Matrix* (PLDM) *associated with* $\boldsymbol{\tau}$. For example, the PLM and the PLDM associated with the UBT shown in Figure 2 read as follows:

$$\boldsymbol{\tau} = \begin{pmatrix} 0 & 2 & 3 & 4 & 4 \\ 2 & 0 & 3 & 4 & 4 \\ 3 & 3 & 0 & 3 & 3 \\ 4 & 4 & 3 & 0 & 2 \\ 4 & 4 & 3 & 2 & 0 \end{pmatrix}; \qquad \mathbf{2}^{-\boldsymbol{\tau}} = \begin{pmatrix} 0 & 1/4 & 1/8 & 1/16 & 1/16 \\ 1/4 & 0 & 1/8 & 1/16 & 1/16 \\ 1/8 & 1/8 & 0 & 1/8 & 1/8 \\ 1/16 & 1/16 & 1/8 & 0 & 1/4 \\ 1/16 & 1/16 & 1/8 & 1/4 & 0 \end{pmatrix}. \tag{2.10}$$

In the remainder of the article we shall refer to

- an integer semimetric that satisfies (2.6) as a *Strong Metric* (SM) matrix;
- a SM matrix that satisfies (2.7) as a *Strong Additive* (SA) matrix;
- a SA matrix that satisfies (2.5) as a *Kraft Strong Additive* (KSA) matrix;
- a SM matrix that satisfies (2.5) as a *Kraft Strong Metric* (KSM) matrix;

For convenience, we summarize in Tables 1 and 2, respectively, the notation and the conditions introduced in these first sections.

3. A characterization of Θ_n for $n \in [3, 11]$. Buneman' strong four-point conditions (2.7) constitute a set of disjunctive conditions. As observed in [6], their inclusion in a valid *Integer Linear Programming* (ILP) formulation for the BMEP proves to be computationally expensive. This fact justifies the search for alternative characterizations of Θ_n that do not involve (2.7). In this section, we show that, for any $n \in [3, 11]$, Buneman' strong four-point conditions (2.7) can be replaced by a single non-linear condition on path-lengths, referred to as the *UBT-manifold*.

Condition	Reference label	Definition
Separation, positivity and symmetry	(1.1)–(1.3)	Section 1
Bounded integrality	(3.2)	Section 2
Strong triangle inequalities	(2.6)	Section 2
Kraft's equalities	(2.5)	Section 2
Buneman's four-point conditions	(2.7)	Section 2
UBT-manifold condition	(3.11)	Section 3
Weak triangle inequalities	(3.24)	Section 3

Table 2: Known necessary conditions that must be satisfied by PLMs of UBTs.

As a first step, in Subsection 3.1 we bridge the algebraic characterization of PLMs with their topological interpretation in order to precisely identify the role played by Buneman' strong four-point conditions. Then, in Subsection 3.2 we provide the characterization.

3.1. A characterization of PLMs based on contractions. Given a generic matrix $\boldsymbol{\tau}$ of order $n \geq 3$, we say its rows i and j are *equivalent* if

$$\tau_{ij} = 2 \wedge \tau_{ip} = \tau_{jp}, \forall\, p \in [n] \setminus \{i, j\}. \tag{3.1}$$

In addition, we denote by $\boldsymbol{\xi}(\boldsymbol{\tau})$ the submatrix of $\boldsymbol{\tau}$ obtained by removing its null rows and columns.

We say that a semimetric matrix $\boldsymbol{\tau}$ with at least two equivalent rows is *contractible* if its entries satisfy the following *bounded integrality* conditions:

$$\tau_{ij} \in [2, n-1], \forall\, i, j \in [n], i \neq j. \tag{3.2}$$

Given a contractible matrix $\boldsymbol{\tau}$, we define a *contraction* as the operation that, given $i, j \in [n], i \neq j$, yields a matrix $\hat{\boldsymbol{\tau}}^{(i,j)}$ obtained from $\boldsymbol{\tau}$ by first setting its i-th row and column to zero and then subtracting one from the non-diagonal entries of its j-th row and j-th column that do not belong to either the i-th row or the i-th column. Thus, the entry (p, q) of $\hat{\boldsymbol{\tau}}^{(i,j)}$ is

$$\hat{\tau}_{pq}^{(i,j)} = \begin{cases} 0 & \text{if } p = i \text{ or } q = i, \\ \tau_{pq} - 1 & \text{if } p \neq q \text{ and } ((p = j \text{ and } q \neq i) \text{ or } (q = j \text{ and } p \neq i)), \\ \tau_{pq} & \text{otherwise.} \end{cases} \tag{3.3}$$

For example, a contraction on

$$\boldsymbol{\tau} = \begin{pmatrix} 0 & 3 & 3 & 2 \\ 3 & 0 & 2 & 3 \\ 3 & 2 & 0 & 3 \\ 2 & 3 & 3 & 0 \end{pmatrix}$$

with respect to rows 2 and 3 results in

$$\hat{\boldsymbol{\tau}}^{(2,3)} = \begin{pmatrix} 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 2 \\ 2 & 0 & 2 & 0 \end{pmatrix} \qquad \boldsymbol{\xi}(\hat{\boldsymbol{\tau}}^{(2,3)}) = \begin{pmatrix} 0 & 2 & 2 \\ 2 & 0 & 2 \\ 2 & 2 & 0 \end{pmatrix}.$$

In the context of PLMs of UBTs, contractions have a precise topological interpretation. Specifically, consider a UBT $T \in \mathcal{U}_n$ and its associated PLM $\boldsymbol{\tau} \in \Theta_n$. If the leaves i and j of T belong to the same cherry, the rows i and j of $\boldsymbol{\tau}$ are equivalent. Then, if we apply a contraction on $\boldsymbol{\tau}$ with respect to the rows i and j, the resulting matrix $\boldsymbol{\xi}(\hat{\boldsymbol{\tau}}^{(i,j)})$ encodes the UBT with $n-1$ leaves induced by all the vertices in T, except for the leaves i and j. As an example, consider the PLM $\boldsymbol{\tau}$ of the UBT shown in Figure 2 and contract $\boldsymbol{\tau}$ with respect to the rows associated with the cherry constituted by leaves 1 and 2. Figure 7 shows the UBT encoded by matrix $\boldsymbol{\xi}(\hat{\boldsymbol{\tau}}^{(1,2)})$, with

$$\hat{\boldsymbol{\tau}}^{(1,2)} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 3 & 3 \\ 0 & 2 & 0 & 3 & 3 \\ 0 & 3 & 3 & 0 & 2 \\ 0 & 3 & 3 & 2 & 0 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\xi}(\hat{\boldsymbol{\tau}}^{(1,2)}) = \begin{pmatrix} 0 & 2 & 3 & 3 \\ 2 & 0 & 3 & 3 \\ 3 & 3 & 0 & 2 \\ 3 & 3 & 2 & 0 \end{pmatrix}. \tag{3.4}$$

Given a matrix $\boldsymbol{\tau}$, we will drop the superscript (i,j) from $\hat{\boldsymbol{\tau}}^{(i,j)}$ for simplicity whenever the context is unambiguous. With a little abuse, we will also say that a matrix $\hat{\boldsymbol{\tau}}$ obtained by possibly repeated contractions has a certain property if this property is enjoyed by its submatrix $\boldsymbol{\xi}(\hat{\boldsymbol{\tau}})$. For example, we say that the above $\hat{\boldsymbol{\tau}}^{(1,2)}$ is a KSM because $\boldsymbol{\xi}(\hat{\boldsymbol{\tau}}^{(1,2)})$ is a KSM.

The following result, introduced in [7], gives the necessary and sufficient conditions in terms of contractions for a matrix $\boldsymbol{\tau}$ to be in Θ_n.

PROPOSITION 4. *Let $\boldsymbol{\tau}$ be an integer square matrix of order $n \geq 3$. Matrix $\boldsymbol{\tau}$ is a PLM if and only if there exist matrices $\boldsymbol{\tau}^s : s \in [0, n-3]$, such that $\boldsymbol{\tau}^0 = \boldsymbol{\tau}$ and, for each $s \in [0, n-3]$, $\boldsymbol{\tau}^s$ is obtained by performing a contraction operation on $\boldsymbol{\tau}^{s-1}$.*

The two following results are a straightforward consequence of Proposition 4:
- if matrix $\boldsymbol{\tau}$ is a PLM, then $\boldsymbol{\tau}^{n-3}$ is the basic matrix (2.8);
- $\boldsymbol{\tau}$ is a PLM if and only if the properties that define a contractible matrix (i.e., of being semimetric, respecting the bounded integrality conditions (3.2) and having at least two equivalent rows) are preserved by subsequent contractions.

Propositions 3 and 4 together imply that if a matrix $\boldsymbol{\tau}$ is a KSA matrix, then:
- $\boldsymbol{\tau}$ is contractible;
- $\boldsymbol{\xi}(\hat{\boldsymbol{\tau}})$ obtained by a contraction on $\boldsymbol{\tau}$ is a KSA matrix and hence contractible.

We now examine how the necessary and sufficient conditions in Proposition 3 contribute to ensuring the contractibility of a KSA matrix. First, we observe that a contraction on a contractible semimetric $\boldsymbol{\tau}$, whose rows satisfy Kraft's equalities, yields a matrix $\hat{\boldsymbol{\tau}}$ that is also a semimetric. In fact, Kraft's equalities impose that the values of the non-diagonal entries of $\boldsymbol{\tau}$ are in $[2, n-1]$ and, by definition of the contraction operation, $\hat{\boldsymbol{\tau}}$ is symmetric, the diagonal entries of $\hat{\boldsymbol{\tau}}$ are zero, while the non-diagonal entries $\hat{\tau}_{ij}$ of $\boldsymbol{\xi}(\hat{\boldsymbol{\tau}})$ are integer values such that $0 < \tau_{ij} - 1 \leq \hat{\tau}_{ij} \leq \tau_{ij}$. The next proposition proves that Kraft's equalities are preserved by a contraction.

PROPOSITION 5. *Let $\boldsymbol{\tau}$ be a contractible matrix whose rows satisfy Kraft's equalities (2.5) and rows i, j, with $i < j$, are equivalent. Then, $\hat{\boldsymbol{\tau}} = \{\hat{\tau}^{(i,j)}\}$ satisfies Kraft's equalities.*

Proof. Observe that j-th row of $\hat{\boldsymbol{\tau}}$ satisfies Kraft's equality, as

$$\sum_{q\in[n]\setminus\{i,j\}} 2^{-\hat{\tau}_{jq}} = 2\sum_{q\in[n]\setminus\{i,j\}} 2^{-\tau_{jq}} = 2\left(\sum_{q\in[n]\setminus\{j\}} 2^{-\tau_{jq}} - 2^{-\tau_{ij}}\right) = \frac{1}{2},$$

where the first and the third equality hold due to $\hat{\tau}_{jq} = \tau_{jq}-1$ and $\tau_{ij} = 2$, respectively. Kraft's equalities also hold for each row $p \in [n] \setminus \{i,j\}$ of $\boldsymbol{\tau}$, as

$$\begin{aligned}\sum_{p\in[n]\setminus\{i,p\}} 2^{-\hat{\tau}_{pq}} &= \sum_{p\in[n]\setminus\{i,j,p\}} 2^{-\tau_{pq}} + 2^{-\hat{\tau}_{pj}} = \sum_{p\in[n]\setminus\{i,j,p\}} 2^{-\tau_{pq}} + 2\cdot 2^{-\tau_{pj}}\\ &= \sum_{p\in[n]\setminus\{i,j,p\}} 2^{-\tau_{pq}} + 2^{-\tau_{pi}} + 2^{-\tau_{pj}} = \sum_{q\in[n]\setminus\{p\}} 2^{-\tau_{pq}} = \frac{1}{2}.\end{aligned}$$

where the first equality holds due to $\hat{\tau}_{pq} = \tau_{pq}$ for each $q \neq j$, the second equality holds due to $\hat{\tau}_{pj} = \tau_{pj} - 1$, and the third equality holds due to $\tau_{pi} = \tau_{pj}$. ◻

An immediate consequence of Proposition 5 is that a contraction on a contractible matrix of order n whose rows satisfy Kraft's equalities yields a matrix whose non-zero entries are in $[2, n-2]$ and that satisfy the bounded integrality conditions (3.2).

Let's consider now the relationship between the strong triangle inequalities (2.6) and the existence of two equivalent rows in a semimetric $\boldsymbol{\tau}$ of order n. Specifically, we show that if $\tau_{ij} = 2$, then the strong triangle inequalities alone are sufficient for rows i and j of $\boldsymbol{\tau}$ to be equivalent. If $\boldsymbol{\tau}$ is a PLM, the claim is consistent with the fact that two leaves of a UBT belonging to the same cherry have the same path-length distance from each of the other leaves. We also show that the strong triangle inequalities are "almost" preserved by contraction.

PROPOSITION 6. *Let $\boldsymbol{\tau}$ denote a SM matrix of order $n \geq 3$ and let i and j be a pair of distinct integers in $[n]$ such that $\tau_{ij} = 2$. Then,*

$$\tau_{ik} = \tau_{jk} = \tau_{ki} = \tau_{kj} \quad \forall\ k \in [n] \setminus \{i,j\}. \tag{3.5}$$

In addition, the matrix $\hat{\boldsymbol{\tau}} = \{\hat{\tau}^{(i,j)}\}$ obtained from $\boldsymbol{\tau}$ by a contraction satisfies the following strong triangle inequalities:

$$\hat{\tau}_{jp} + \hat{\tau}_{pq} - \hat{\tau}_{jq} \geq 2 \quad \forall\ j,p,q \in [n] \setminus \{i\},\ \ j,p,q \text{ distinct} \tag{3.6}$$

$$\hat{\tau}_{pr} + \hat{\tau}_{pq} - \hat{\tau}_{qr} \geq 2 \quad \forall\ p,q,r \in [n] \setminus \{i,j\},\ \ p,q,r \text{ distinct.} \tag{3.7}$$

Proof. As (2.6) holds for $\boldsymbol{\tau}$, we have that

$$\tau_{ij} + \tau_{jk} \geq 2 + \tau_{ik} \qquad \forall\ k \in [n] \setminus \{i,j\} \tag{3.8}$$

$$\tau_{ij} + \tau_{ik} \geq 2 + \tau_{jk} \qquad \forall\ k \in [n] \setminus \{i,j\}. \tag{3.9}$$

By combining (3.8) and (3.9) with $\tau_{ij} = 2$, we have that $\tau_{ik} = \tau_{jk}$ for all $k \in [n]\setminus\{i,j\}$. The other equalities in (3.5) are a direct consequence of the symmetry of $\boldsymbol{\tau}$.

As regards the final part of the statement, we observe first that for a fixed distinct triplet of indices $j,p,q \in [n] \setminus \{i\}$, the following relationships hold:

$$\begin{aligned}\tau_{jp} + \tau_{pq} - \tau_{jq} \geq 2 \quad &\Leftrightarrow \quad (\tau_{jp} - 1) + \tau_{pq} - (\tau_{jq} - 1) \geq 2 \quad \Leftrightarrow\\ &\Leftrightarrow \quad \hat{\tau}_{jp} + \hat{\tau}_{pq} - \hat{\tau}_{jq} \geq 2.\end{aligned}$$

In particular, the first inequality holds as $\boldsymbol{\tau}$ is a SM matrix. The second inequality holds by definition of a contraction on rows i, j, where $\hat{\tau}_{jk} = \tau_{jk} - 1$ for all $k \in [n] \setminus \{i, j\}$. Finally, observe that that for a fixed distinct triplet of indices $p, q, r \in [n] \setminus \{i, j\}$,

$$\tau_{pr} + \tau_{pq} - \tau_{qr} \geq 2 \quad \Leftrightarrow \quad \hat{\tau}_{pr} + \hat{\tau}_{pq} - \hat{\tau}_{qr} \geq 2.$$

As the second inequality holds by definition of contraction on rows i, j, we have $\hat{\tau}_{hk} = \tau_{hk}$ for all distinct $h, k \in [n] \setminus \{i, j\}$. ◻

Proposition 6 highlights, at the same time, the strengths and the limitations of the strong triangle inequalities. Specifically, the strong triangle inequalities of type

$$\hat{\tau}_{jp} + \hat{\tau}_{jq} - \hat{\tau}_{pq} \geq 2 \quad \forall\ j, p, q \in [n] \setminus \{i\},\ j, p, q \text{ distinct} \tag{3.10}$$

may not propagate through a contraction on a SM matrix (as the arguments in Proposition 6 cannot be used). In addition, Kraft's equalities and strong triangle inequalities are not sufficient to impose the presence of an entry with value 2 in a semimetric, as is the case for the following KSM matrix

$$\begin{pmatrix} 0 & 3 & 3 & 3 & 3 \\ 3 & 0 & 3 & 3 & 3 \\ 3 & 3 & 0 & 3 & 3 \\ 3 & 3 & 3 & 0 & 3 \\ 3 & 3 & 3 & 3 & 0 \end{pmatrix}$$

We can prove the following proposition on Buneman' strong four-point conditions with arguments similar to the ones used for Proposition 6.

PROPOSITION 7. *Let $\boldsymbol{\tau}$ be a contractible matrix satisfying Buneman' strong four-point conditions* (2.7) *and whose rows i, j, with $i < j$, are equivalent. Then, $\hat{\boldsymbol{\tau}} = \{\hat{\tau}^{(i,j)}\}$ is a semimetric matrix that still satisfies Buneman' strong four-point conditions.*

Proposition 7, together with the above major limitations relative to the strong triangle inequalities, allow us to claim that Buneman' strong four-point conditions are necessary to impose the existence of an entry with value 2 in a SMA matrix $\boldsymbol{\tau}$. In addition, the satisfaction of Buneman' strong four-point conditions implies also the propagation of the strong triangle inequalities of type (3.10) through contractions. Specifically, consider matrix $\hat{\boldsymbol{\tau}}^{(i,j)}$ obtained by a contraction on a SMA matrix $\boldsymbol{\tau}$. Without loss of generality, suppose that Buneman' strpmg four-point condition $\tau_{ip} + \tau_{jq} \geq \tau_{ij} + \tau_{pq} + 2$ holds for each $p, q \in [n] \setminus \{i, j\}$. Then this inequality, together with the equivalence of rows i and j of $\boldsymbol{\tau}$, implies $\tau_{jp} + \tau_{jq} \geq \tau_{pq} + 4$, and thus $\hat{\tau}_{jp} + \hat{\tau}_{jq} \geq \hat{\tau}_{pq} + 2$.

3.2. The UBT-manifold condition. So far, we have established that the satisfaction of Buneman' strong four-point conditions is sufficient to guarantee that in a given a matrix $\boldsymbol{\tau}$ there exists an entry whose value is 2. In addition, Buneman's conditions ensure that the strong triangle inequalities are preserved through contractions. We now show that, for any $n \in [3, 11]$, the condition

$$\sum_{i \in [n]} \sum_{j \in [n] \setminus i} \tau_{ij} 2^{-\tau_{ij}} = 2n - 3, \tag{3.11}$$

referred to as the *UBT-manifold condition* [5], can be exploited to provide a characterization of Θ_n that does not rely on Buneman' strong four-point conditions. This condition binds the path-lengths τ_{ij} to their own *path-lengths densities* $2^{-\tau_{ij}}$. As observed in [5], the UBT-manifold states that all UBTs with n leaves have $2n-3$ edges. Hence, the vertices of the convex hull of the solutions to the BMEP (i.e., all the PLMs $\boldsymbol{\tau} \in \Theta_n$) lie on this manifold. For brevity, we will use the acronym KSM-UBT to denote a KSM matrix that also satisfies the UBT-manifold.

In order to provide the alternative characterization of Θ_n we first show that if a matrix $\boldsymbol{\tau}$ of order $n \in [3, 11]$ satisfies Kraft's equalities and the UBT-manifold condition, then at least one of its entries has value 2. We start from the following preliminary result.

PROPOSITION 8. *Given a positive integer $n \geq 3$ and an integer $i \in [n]$, consider a path-length sequence $\vec{\boldsymbol{\tau}}_i = \{\vec{\tau}_{i1}, \vec{\tau}_{i2}, \ldots, \vec{\tau}_{i,n-1}\} \in \vec{\Phi}_i$. If $\vec{\tau}_{i,n-1}$ is equal to $n-1$, then the remaining entries of $\vec{\boldsymbol{\tau}}_i$ satisfy*

$$\vec{\tau}_{i,n-2} = \vec{\tau}_{i,n-1} = n-1, \quad \vec{\tau}_{ik} = \vec{\tau}_{i,k+1} - 1 = k+1 \qquad \forall\, k \in [1, n-3]. \tag{3.12}$$

Proof. Proposition 2 implies that $\vec{\boldsymbol{\tau}}_i$ satisfies Kraft's equality. Furthermore, the arguments of its proof state that $\sum_{j \in [n-1]} 2^{-\vec{\tau}_{ij}}$ reaches its maximum value $2^{1-(\vec{\tau}_{i,n-1}-(n-3))}$ if and only if (2.4), which is equivalent to (3.12), holds. Since we assume $\vec{\tau}_{i,n-1} = n-1$, then $2^{1-(\vec{\tau}_{i,n-1}-(n-3))} = 2^{-1}$ as required by Kraft's equality and thus (3.12) must hold. ∎

A direct consequence of Proposition 8 is that every $\vec{\boldsymbol{\tau}}_i \in \vec{\Phi}_i$ that has an entry equal to $n-1$ also has an entry equal to 2.

PROPOSITION 9. *Let $\boldsymbol{\tau}$ denote a matrix of order $n \in [3, 11]$ that satisfies Kraft's equalities and the UBT-manifold condition. Then, at least one of its entries must be equal to 2.*

Proof. Suppose, by contradiction, that all the entries of $\boldsymbol{\tau}$ are such that $\tau_{ij} \in [3, n-1]$ for all distinct $i, j \in [n]$. Consider the generic i-th row of $\boldsymbol{\tau}$ and its associated path-length sequence $\vec{\boldsymbol{\tau}}_i$ and observe that $\vec{\boldsymbol{\tau}}_i$ cannot include an entry with value equal to $n-1$ otherwise it would also include an entry with value equal to 2 by Proposition 8, in contradiction with the assumption $\tau_{ij} \in [3, n-1]$. Now, consider the problem of finding a path-length sequence $\vec{\theta}^*_i$ that constitutes an optimal solution to the following problem:

$$L_i(\tau) = \min \left\{ \sum_{j \in [n-1]} \vec{\tau}_{ij} 2^{-\vec{\theta}_{ij}} : \vec{\theta}_i \in \vec{\Phi}_i,\ \vec{\theta}_{ij} \in [3, n-2],\ \forall\, j \in [n] \setminus \{i\} \right\}. \tag{3.13}$$

We claim that the optimal solution to L_i is

$$\vec{\theta}^*_i = \{3, 3, 3, 4, 5, \ldots, n-3, n-2, n-2\},$$

that encodes a UBT U^*_i consisting of a caterpillar rooted in i (see Figure 6). Indeed, suppose, by contradiction, that there is an optimal solution $\vec{\theta}'_i \in \vec{\Phi}_i$, $\vec{\theta}'_i \neq \vec{\theta}^*_i$. Then, there is a pair $p, q \in [n] \setminus \{i\}, p < q$, such that $\vec{\theta}'_{ip} \geq \vec{\theta}'_{iq}$. Moreover, since $\vec{\theta}'_i$ is optimal,

$$\vec{\tau}_{ip} 2^{-\vec{\theta}'_{ip}} + \vec{\tau}_{iq} 2^{-\vec{\theta}'_{iq}} \leq \vec{\tau}_{ip} 2^{-\vec{\theta}'_{iq}} + \vec{\tau}_{iq} 2^{-\vec{\theta}'_{ip}},$$

which implies

$$(\vec{\tau}_{ip} - \vec{\tau}_{iq})(2^{-\vec{\theta}'_{ip}} - 2^{-\vec{\theta}'_{iq}}) \leq 0, \tag{3.14}$$

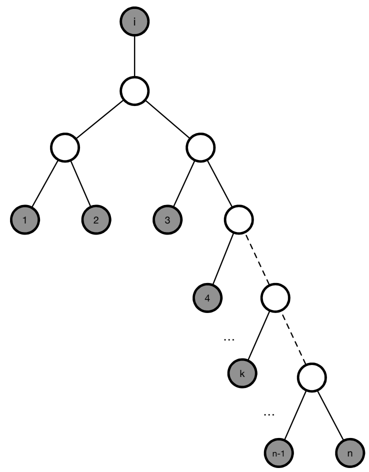

Fig. 6: A graphical representation of the optimal solution to (3.13).

As $\vec{\tau}_{ip} \leq \vec{\tau}_{iq}$ and $2^{-\vec{\theta}'_{ip}} \leq 2^{-\vec{\theta}'_{iq}}$ (since $\vec{\theta}'_{ip} \geq \vec{\theta}'_{iq}$), inequality (3.14) only holds to equality if $\vec{\theta}'_i = \vec{\theta}^*_i$, which contradicts $\vec{\theta}'_i \neq \vec{\theta}^*_i$. Then, we can write (3.13) as

$$L_i(\tau) = \vec{\tau}_{i,3}2^{-3} + \sum_{j\in[2,n-2]} \vec{\tau}_{ij}2^{-j} + \vec{\tau}_{i,n-1}2^{-(n-2)}. \tag{3.15}$$

Observe that $\vec{\theta}^*_i$ satisfies Kraft's equalities and that $x^* = \vec{\theta}^*_i$ is also the solution to the problem

$$\min\left\{\sum_{j\in[n-1]} x_j 2^{-\vec{\theta}^*_{ij}} : x \in \mathbb{R}^{n-1}, \sum_{j\in[n-1]} 2^{-x_j} = \frac{1}{2}\right\}$$

as it can be immediately verified by applying the Karush–Kuhn–Tucker conditions. Then, the following lower bound holds for (3.15):

$$L_i(\tau) \geq \sum_{j\in[n-1]} \vec{\theta}^*_{ij}2^{-\vec{\theta}^*_{ij}} \geq 3\cdot\frac{3}{8} + \sum_{j\in[3,n-2]} j2^{-j} + (n-2)2^{-(n-2)} \quad \forall\, \tau \in \Theta_n, i \in [n].$$

Since $\sum_{j\in[n]\setminus i} \tau_{ij}2^{-\tau_{ij}} = \sum_{j\in[n-1]} \vec{\tau}_{ij}2^{-\vec{\tau}_{ij}} \geq L_i(\tau)$, we have that

$$\sum_{i\in[n]}\sum_{j\in[n]\setminus i} \tau_{ij}2^{-\tau_{ij}} \geq \sum_{i\in[n]} L_i(\tau) \quad \geq n\left(3\cdot\frac{3}{8} + \sum_{j\in[3,n-2]} j2^{-j} + (n-2)2^{-(n-2)}\right). \tag{3.16}$$

We recall that *Gabriel's staircase* power series $\sum_{j=1}^{\infty} j2^{-j}$ converges to 2 and the partial sum of its first k terms is given by

$$G_k = \sum_{j\in[k]} j2^{-j} = 2 - (k+2)2^{-k}.$$

Hence,

$$\sum_{j\in[3,n-2]} j2^{-j} = G_{n-2} - G_3 = \frac{5}{8} - n2^{-(n-2)},$$

which implies

$$\sum_{i\in[n]}\sum_{j\in[n]\setminus i} \tau_{ij}2^{-\tau_{ij}} \geq \sum_{i\in[n]} L_i(\tau) \geq n\left(\frac{7}{4} - 2^{-(n-3)}\right) > 2n-3, \tag{3.17}$$

by contradicting the hypothesis that $\boldsymbol{\tau}$ satisfies the UBT-manifold. Thus, the statement follows. ∎

The next proposition presents some useful properties of KSM matrices.

Proposition 10. *Let $\boldsymbol{\tau}$ denote a KSM matrix of order $n \geq 3$ and let i and j be a pair of distinct integers in $[n]$ such that $\tau_{ij} = 2$. Also assume that for some $p \in [n] \setminus \{i,j\}$, $\tau_{ip} = n-1$. Then, there exists an integer $q \in [n] \setminus \{i,j,p\}$ such that*

$$\tau_{ip} = \tau_{iq} = \tau_{jp} = \tau_{jq} = n-1 \tag{3.18}$$

and that

$$\tau_{pq} = 2. \tag{3.19}$$

Furthermore, each non-diagonal entry of $\boldsymbol{\tau}$ different from (r,s), with $r,s \in \{i,j,p,q\}$, has value strictly less than $n-1$.

Proof. Consider the i-th row of $\boldsymbol{\tau}$ and first observe that, as $\tau_{ij} = 2$, by Proposition 8 there exists an index $q \in [n] \setminus \{i\}$ such that $\tau_{iq} = n-1$. Moreover, by Proposition 6, we have that $\tau_{ip} = \tau_{jp} = n-1$ and $\tau_{iq} = \tau_{jq} = n-1$. Hence, (3.18) holds true. Also note that Proposition 8, applied to row p of the matrix $\boldsymbol{\tau}$, guarantees the existence of an index $r \in [n]$ such that $\tau_{rp} = 2$, as $\tau_{ip} = n-1$. The strong triangle inequality

$$\tau_{ir} + \tau_{rp} = \tau_{ir} + 2 \geq 2 + \tau_{ip} = 2 + n - 1,$$

together with

$$\tau_{ir} \in [2, n-1],$$

implies $\tau_{ir} = n-1$. Proposition 8, applied again to row i of the matrix $\boldsymbol{\tau}$, imposes that there can be only two entries of row i equal to $n-1$. Thus, the indices r and q must be equal, hence $\tau_{pr} = \tau_{pq} = 2$, and (3.19) holds true.

We prove now the last statement of the proposition. To this end, consider the i, j, p, and q-th rows of matrix $\boldsymbol{\tau}$. Then, by Proposition 8, in each of such rows, there is one and only one entry with value equal to 2 and exactly two entries equal to $n-1$. Hence, $\tau_{rs} \in [3, n-2]$, for all distinct $r, s \in [n] \setminus \{i,j,p,q\}$. Finally, consider a row $u \in [n] \setminus \{i,j,p,q\}$ of matrix $\boldsymbol{\tau}$. As $\tau_{ij} = 2$, by Proposition 6 we have that $\tau_{ui} = \tau_{uj}$. Moreover, by Proposition 8, both τ_{ui} and τ_{uj} are in $[3, n-2]$. Now observe that $\tau_{uv} \in [3, n-2]$ holds true for any $v \in [n] \setminus \{i,j,u\}$. Indeed, if $\tau_{uv} = n-1$ held true then we had $\tau_{uv} \neq \tau_{ui}$, as $\tau_{ui} \in [3, n-2]$ for $i \in [n] \setminus \{j,u\}$. However, by Proposition 8, the u-th row of $\boldsymbol{\tau}$ cannot include two identical entries with value different from $n-1$. Thus, the statement follows. ∎

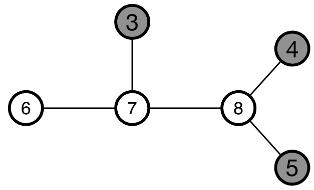

Fig. 7: The UBT corresponding to the PLM $\boldsymbol{\xi}(\hat{\boldsymbol{\tau}}^{(1,2)})$ given in (3.4).

If $\boldsymbol{\tau}$ is a PLM, Proposition 10 equivalently states that caterpillars, which are the only UBTs in $\mathcal{U}_n$ having two leaves whose path-length distance is equal to $n-1$, have only two cherries.

Proposition 9 guarantees the existence of a non-diagonal entry $\tau_{ij} = 2$ for some $i, j \in [n]$ in any semimetric matrix $\boldsymbol{\tau}$ of order $n \in [3, 11]$ satisfying Kraft's equalities and the UBT-manifold condition. The subsequent proposition ensures that if $\boldsymbol{\tau}$ also satisfies the strong triangle inequalities, a contraction with respect to this specific entry yields a semimetric matrix that satisfies Kraft's equalities and the UBT-manifold condition.

Proposition 11. *Let $\boldsymbol{\tau}$ denote a KSM matrix of order $n \geq 3$ and let $i, j \in [n]$ such that $i < j$ and $\tau_{ij} = 2$. Then, $\hat{\boldsymbol{\tau}} = \{\hat{\tau}^{(i,j)}\}$ is also a semimetric matrix that satisfies Kraft's equalities and the UBT-manifold condition.*

Proof. Observe first that $\hat{\boldsymbol{\tau}}$ is semimetric because $\hat{\tau}_{rs} = \tau_{rs}$ for each $r, s \in [n] \setminus \{i, j\}$, the diagonal elements $\hat{\tau}_{jj} = \tau_{jj}$ are zero, and finally $\hat{\tau}_{jr} = \tau_{jr} - 1 = \hat{\tau}_{rj} = \tau_{rj} - 1 \geq 2$ for each $r \setminus \{i, j\}$, as $\boldsymbol{\tau}$ is symmetric and $\tau_{jr} \geq 3$ since $\tau_{jr} \neq \tau_{ij}$. Moreover, $\hat{\tau}_{r,s} \in [2, n-2]$ for all distinct $r, s \in [n]$. Indeed, such entries are integers greater than or equal to 2 by the previous argument. In addition, due to Proposition 10, the entries of $\boldsymbol{\tau}$ that are possibly equal to $n-1$ belong either to its i-th or j-th row. Hence, $\hat{\tau}_{rs} \leq n-2$.

Observe that j-th row of $\hat{\boldsymbol{\tau}}$ satisfies Kraft's equality, as

$$
\begin{aligned}
\sum_{q \in [n] \setminus \{i,j\}} 2^{-\hat{\tau}_{jq}} &= 2 \sum_{q \in [n] \setminus \{i,j\}} 2^{-\tau_{jq}} \\
&= 2 \left(\sum_{q \in [n] \setminus \{j\}} 2^{-\tau_{jq}} - 2^{-\tau_{ij}} \right) \\
&= 2 \left(\frac{1}{2} - \frac{1}{4} \right) = \frac{1}{2},
\end{aligned}
$$

where the first and the third equality hold due to $\hat{\tau}_{jq} = \tau_{jq} - 1$ and $\tau_{ij} = 2$, respectively. Kraft's equalities also hold for each row $p \in [n] \setminus \{i, j\}$ of $\boldsymbol{\tau}$, as

$$
\begin{aligned}
\sum_{p \in [n] \setminus \{i,p\}} 2^{-\hat{\tau}_{pq}} &= \sum_{p \in [n] \setminus \{i,j,p\}} 2^{-\tau_{pq}} + 2^{-\hat{\tau}_{pj}} \\
&= \sum_{p \in [n] \setminus \{i,j,p\}} 2^{-\tau_{pq}} + 2 \cdot 2^{-\tau_{pj}} \\
&= \sum_{p \in [n] \setminus \{i,j,p\}} 2^{-\tau_{pq}} + 2^{-\tau_{pi}} + 2^{-\tau_{pj}} = \sum_{q \in [n] \setminus \{p\}} 2^{-\tau_{pq}} = \frac{1}{2}.
\end{aligned}
$$

where the first equality holds due to $\hat{\tau}_{pq} = \tau_{pq}$ for each $q \neq j$, the second equality holds due to $\hat{\tau}_{pj} = \tau_{pj} - 1$, and the third equality holds due to $\tau_{pi} = \tau_{pj}$. An immediate consequence of the fact that Kraft's equalities hold on $\hat{\boldsymbol{\tau}}$ is that its non-zero entries are greater than or equal to 2.

Now, observe that the entries of $\hat{\boldsymbol{\tau}}$ satisfy the UBT-manifold (3.11), as

$$\begin{aligned}
\sum_{p\in[n]\setminus\{i\}} \sum_{q\in[n]\setminus\{i,p\}} \hat{\tau}_{pq} 2^{-\hat{\tau}_{pq}} &= \sum_{q\in[n]\setminus\{i,j\}} \hat{\tau}_{jq} 2^{-\hat{\tau}_{jq}} + \sum_{p\in[n]\setminus\{i,j\}} \hat{\tau}_{pj} 2^{-\hat{\tau}_{pj}} + \\
&\quad + \sum_{p\in[n]\setminus\{i,j\}} \sum_{q\in[n]\setminus\{i,j,p\}} \hat{\tau}_{pq} 2^{-\hat{\tau}_{pq}} = \\
&= \sum_{q\in[n]\setminus\{i,j\}} (\tau_{jq} - 1) 2^{-(\tau_{jq}-1)} + \\
&\quad + \sum_{p\in[n]\setminus\{i,j\}} (\tau_{pj} - 1) 2^{-(\tau_{pj}-1)} + \\
&\quad + \sum_{p\in[n]\setminus\{i,j\}} \sum_{q\in[n]\setminus\{i,j,p\}} \tau_{pq} 2^{-\tau_{pq}} = \\
&= 2 \sum_{q\in[n]\setminus\{i,j\}} \tau_{jq} 2^{-\tau_{jq}} + 2 \sum_{p\in[n]\setminus\{i,j\}} \tau_{pj} 2^{-\tau_{pj}} + \\
&\quad + \sum_{p\in[n]\setminus\{i,j\}} \sum_{q\in[n]\setminus\{i,j,p\}} \tau_{pq} 2^{-\tau_{pq}} + \\
&\quad - 4 \sum_{q\in[n]\setminus\{i,j\}} 2^{-\tau_{jq}} = \\
&= \sum_{q\in[n]\setminus\{i\}} \tau_{iq} 2^{-\tau_{iq}} + \sum_{q\in[n]\setminus\{j\}} \tau_{jq} 2^{-\tau_{jq}} + \\
&\quad + \sum_{p\in[n]\setminus\{i,j\}} \sum_{q\in[n]\setminus\{i,j,p\}} \tau_{pq} 2^{-\tau_{pq}} + \\
&\quad + \sum_{p\in[n]\setminus\{i\}} \tau_{pi} 2^{-\tau_{pi}} + \sum_{p\in[n]\setminus\{j\}} \tau_{pj} 2^{-\tau_{pj}} + \\
&\quad - 4 \sum_{q\in[n]\setminus\{i,j\}} 2^{-\tau_{jq}} + 4(\tau_{ij} 2^{-\tau_{ij}} - \tau_{ij} 2^{-\tau_{ij}}) \\
&= \sum_{p\in[n]} \sum_{q\in[n]\setminus\{p\}} \tau_{pq} 2^{-\tau_{pq}} - 4 \sum_{q\in[n]\setminus\{i\}} 2^{-\tau_{jq}} \\
&= 2n - 3 - 2 = 2(n-1) - 3,
\end{aligned}$$

where the second equality holds due to $\hat{\tau}_{pj} = \tau_{pj} - 1$ and $\hat{\tau}_{pq} = \tau_{pq}$, for all $p \in [n] \setminus \{i\}$ and $q \in [n] \setminus \{i, j\}$; the fourth equality holds due to $\tau_{pi} = \tau_{pj}$ for all $p \in [n] \setminus \{i, j\}$; and the fifth equality holds due to $\boldsymbol{\tau}$ being symmetric and satisfying Kraft's equalities. ∎

Given a matrix $\boldsymbol{\tau}$ satisfying the hypothesis of Proposition 11, a natural question to ask is whether $\hat{\boldsymbol{\tau}} = \{\hat{\tau}^{(i,j)}\}$ also satisfies the strong triangle inequalities. To this end, we first observe that for a fixed distinct triplet of indices $j, p, q \in [n] \setminus \{i\}$, the following holds:

$$\begin{aligned}
\tau_{jp} + \tau_{pq} - \tau_{jq} \geq 2 \quad &\Leftrightarrow \quad (\tau_{jp} - 1) + \tau_{pq} - (\tau_{jq} - 1) \geq 2 \quad \Leftrightarrow \\
&\Leftrightarrow \quad \hat{\tau}_{jp} + \hat{\tau}_{pq} - \hat{\tau}_{jq} \geq 2. \qquad (3.20)
\end{aligned}$$

The same argument cannot be used to show that

$$\hat{\tau}_{jp} + \hat{\tau}_{jq} - \hat{\tau}_{pq} \geq 2. \tag{3.21}$$

Then assume by contradiction that, for some $p, q \in [n] \setminus \{i, j\}$ and $\alpha < 2$, we have

$$\hat{\tau}_{jp} + \hat{\tau}_{jq} - \hat{\tau}_{pq} = \alpha \quad \Leftrightarrow \quad \hat{\tau}_{pq} = \hat{\tau}_{jp} + \hat{\tau}_{jq} - \alpha. \tag{3.22}$$

Note that the above condition can only hold if $\hat{\tau}_{pq} > 2$, otherwise it is immediate to see that $\hat{\tau}_{jp} + \hat{\tau}_{jq} \geq \hat{\tau}_{pq} + 2$ as $\hat{\tau}_{jp}$ and $\hat{\tau}_{jq}$ are both greater than or equal to 2.

First, consider the case $\hat{\tau}_{jp} = 2$. From (3.21) and (3.22), we get:

$$\begin{aligned} \hat{\tau}_{pq} = \hat{\tau}_{jq} + (2 - \alpha) \quad &\Leftrightarrow \quad 2^{-\hat{\tau}_{pq}} = 2^{-\hat{\tau}_{jq}} 2^{\alpha - 2} \\ \hat{\tau}_{pk} \geq \hat{\tau}_{jk} \quad &\Leftrightarrow \quad 2^{-\hat{\tau}_{pk}} \leq 2^{-\hat{\tau}_{jk}} \quad \forall\, k \in [n] \setminus \{i, j, p, q\}. \end{aligned} \tag{3.23}$$

As $2^{-\hat{\tau}_{pj}} = 2^{-\hat{\tau}_{jp}}$ holds and $\alpha < 2$ implies $2^{\alpha-2} < 1$, the above conditions contradict Kraft's equalities, which require

$$\sum_{r \in [n] \setminus \{i,p\}} 2^{-\hat{\tau}_{pr}} = \sum_{r \in [n] \setminus \{i,j\}} 2^{-\hat{\tau}_{jr}} = \frac{1}{2}.$$

Then, the strong triangle inequalities (3.10) holds for all $q \in [n] \setminus \{i, j, p\}$.

We now prove that if $\hat{\tau}_{jp} = 2$, then $\hat{\tau}_{jr} + \hat{\tau}_{js} - \hat{\tau}_{rs} \geq 2$ for all $r, s \in [n] \setminus \{i, j, p\}$. Specifically, as (3.10) holds for all $q \in [n] \setminus \{i, j, p\}$, we have

$$\begin{aligned} \hat{\tau}_{jr} \geq \hat{\tau}_{pr} + 2 - \hat{\tau}_{jp} \quad &\Leftrightarrow \quad \hat{\tau}_{jr} \geq \hat{\tau}_{pr} \\ \hat{\tau}_{js} \geq \hat{\tau}_{ps} + 2 - \hat{\tau}_{jp} \quad &\Leftrightarrow \quad \hat{\tau}_{js} \geq \hat{\tau}_{ps}. \end{aligned}$$

Since we have already shown that strong triangle inequalities hold for all the entries of $\hat{\boldsymbol{\tau}}$ that do not belong to the j-th row or column, we have

$$\hat{\tau}_{jr} + \hat{\tau}_{js} \geq \hat{\tau}_{pr} + \hat{\tau}_{ps} \geq \hat{\tau}_{rs} + 2 \quad \Rightarrow \quad \hat{\tau}_{jr} + \hat{\tau}_{js} - \hat{\tau}_{rs} \geq 2.$$

In the case $\hat{\tau}_{jp} > 2$, we claim that $\hat{\tau}_{jp} + \hat{\tau}_{jq} - \hat{\tau}_{pq} \geq 2$ holds, provided that $n \leq 11$. However, proving this claim algebraically is currently an open problem, as all the known necessary conditions enjoyed by KSMs yield weak upper bounds on n. Providing a purely algebraic proof of Proposition 10 warrants further analysis of the combinatorics of Θ_n aimed at uncovering key polyhedral properties.

In the light of the consideration above, we will follow a different to prove that, for any $n \in [3, 11]$, Kraft's equalities, the UBT-manifold condition and triangle inequalities are not only necessary but also sufficient to characterize Θ_n. To this end, we first introduce some notation and simple results.

Given a matrix $\boldsymbol{\tau}$, we denote the i-th row of $\boldsymbol{\tau}$ by $\boldsymbol{\tau}_i$. Moreover, given a path-length sequence $\vec{\mathbf{s}}$, we say that $\boldsymbol{\tau}_i$ has *type* $\vec{\mathbf{s}}$ if its non-zero entries match the elements of $\vec{\mathbf{s}}$. Finally, we refer to the following conditions

$$\begin{aligned} \tau_{jk} &\geq \tau_{ik} + (n-3)(2 - \tau_{ij}) \\ \tau_{ik} &\geq \tau_{jk} + (n-3)(2 - \tau_{ij}) \end{aligned} \tag{3.24}$$

as the *Weak Triangle Inequalities* (WTI). Similarly, we refer to the *Strong Triangle Inequalities* as STI. Interestingly, Proposition 4 still holds even if the hypothesis is weakened, requiring only that $\boldsymbol{\tau}$ satisfies the WTI.

The total number a_n of path-length sequences of length $n-1$ is provided by the sequence A002572 in [9], with

$$a_3 = 1, a_4 = 1, a_5 = 2, a_6 = 3, a_7 = 5, a_8 = 9, a_9 = 16, a_{10} = 28, a_{11} = 50.$$

Similarly, the total number m_n of different UBT topologies for $n \geq 3$ is provided by the sequence A000672 in [9], with

$$m_3 = 1, m_4 = 1, m_5 = 1, m_6 = 2, m_7 = 2, m_8 = 4, m_9 = 6, m_{10} = 11, m_{11} = 18.$$

Hereafter, given a matrix $\boldsymbol{\tau}$ of order $n \geq 3$, we call *families* of matrices the sets of symmetric matrices that are isomorphic under symmetric permutations of rows and columns and we describe the families by *canonical representatives* with the understanding that all the other matrices of the families can be obtained from them by symmetric permutations. We consider *combinations* of n, possibly repeated, path-length sequences and we associate each combination with the matrices whose row types have a bijective correspondence with the path-sequences in the combination. We say that a combination $\mathbf{c}$ satisfies the UBT-manifold condition if

$$\sum_{\vec{\boldsymbol{\tau}} \in \mathbf{c}} \sum_{\vec{\tau}_i \in \vec{\boldsymbol{\tau}}} \vec{\tau}_i 2^{-\vec{\tau}_i} = \frac{1}{2}$$

Furthermore, we say that a combination $\mathbf{c}$ satisfies the symmetry conditions (and, possibly, either the WTI or STI) if we can associate with it a symmetric matrix $\boldsymbol{\tau}$ with a null diagonal (which also satisfies either the WTI or STI). Note that Kraft's equalities imply the bounded integrality of the elements of the sequences, and thus the satisfaction of the symmetry conditions together with the condition of the null diagonal entries implies that $\boldsymbol{\tau}$ is semimetric.

In the next proposition, we use the following argument to prove the sufficiency of our conditions. We first generate all the path-length sequences that satisfy Kraft's equalities. Then, we derive all the combinations of n of these path-length sequences that satisfy the considered symmetry and the UBT-manifold conditions. Finally, we show that the number of families of matrices whose row types have a bijective correspondence with the path-sequences in the obtained combinations and that also satisfy, when required, the WTI or the STI conditions, coincides with the number of families of matrices having a UBT realization. This proves the statement, as the considered conditions are also necessary, ensuring that the obtained families cannot differ from those of the matrices having a UBT realization.

PROPOSITION 12. *Consider a matrix $\boldsymbol{\tau}$ of order $n \in [3, 11]$. Then, $\boldsymbol{\tau}$ has a UBT realization if:*

- *$n = 3$ and $\boldsymbol{\tau}$ satisfies Kraft's equalities;*
- *$n = 4$ and $\boldsymbol{\tau}$ satisfies Kraft's equalities and Symmetry;*
- *$n = 5$ and $\boldsymbol{\tau}$ satisfies Kraft's equalities, Symmetry, and the UBT-manifold condition;*
- *$n \in [6, 7]$ and $\boldsymbol{\tau}$ satisfies Kraft's equalities, Symmetry, the UBT-manifold condition, and the WTI.*
- *$n \in [8, 11]$ and $\boldsymbol{\tau}$ satisfies Kraft's equalities, Symmetry, the UBT-manifold condition, and the STI.*

Proof. For ease of reading, we refer the reader to Appendix A as the proof involves the consideration of several distinct cases. ∎

Note that an immediate consequence of Proposition 12 is that the STI are preserved by contractions when $n \leq 11$. On the other hand, satisfying integrality, symmetry, Kraft's equalities, the UBT manifold condition, and the STI does not generally ensure the existence of a UBT realization for $n \geq 12$. In fact, the following matrix:

$$\begin{bmatrix} 0 & 2 & 4 & 4 & 5 & 5 & 6 & 6 & 7 & 7 & 7 & 7 \\ 2 & 0 & 4 & 4 & 5 & 5 & 6 & 6 & 7 & 7 & 7 & 7 \\ 4 & 4 & 0 & 2 & 6 & 6 & 5 & 5 & 7 & 7 & 7 & 7 \\ 4 & 4 & 2 & 0 & 6 & 6 & 5 & 5 & 7 & 7 & 7 & 7 \\ 5 & 5 & 6 & 6 & 0 & 2 & 5 & 5 & 5 & 5 & 6 & 6 \\ 5 & 5 & 6 & 6 & 2 & 0 & 5 & 5 & 5 & 5 & 6 & 6 \\ 6 & 6 & 5 & 5 & 5 & 5 & 0 & 2 & 6 & 6 & 5 & 5 \\ 6 & 6 & 5 & 5 & 5 & 5 & 2 & 0 & 6 & 6 & 5 & 5 \\ 7 & 7 & 7 & 7 & 5 & 5 & 6 & 6 & 0 & 2 & 4 & 4 \\ 7 & 7 & 7 & 7 & 5 & 5 & 6 & 6 & 2 & 0 & 4 & 4 \\ 7 & 7 & 7 & 7 & 6 & 6 & 5 & 5 & 4 & 4 & 0 & 2 \\ 7 & 7 & 7 & 7 & 6 & 6 & 5 & 5 & 4 & 4 & 2 & 0 \end{bmatrix} \tag{3.25}$$

satisfies the conditions above, but does not have a UBT realization.

4. A general characterization of Θ_n.

In this section, we present an algebraic characterization of Θ_n that is linear in the entries of the $\boldsymbol{\tau}$ matrix and does not involve disjunctive conditions. This characterization constitutes the algebraic counterpart of Proposition 4 and builds upon the following two observations.

First, consider a matrix $\boldsymbol{\tau}^0$ of order $n \geq 3$. Proposition 4 ensures that if $\boldsymbol{\tau}^0$ has a UBT realization, then there exist matrices $\boldsymbol{\tau}^1, \boldsymbol{\tau}^2, \dots, \boldsymbol{\tau}^{n-3}$ such that $\boldsymbol{\tau}^s$ is obtained through exactly one contraction on $\boldsymbol{\tau}^{s-1}$ for each $s \in [1, n-3]$. Given an integer $r \in [n-3]$, if $\boldsymbol{\tau}^r$ is obtained through a contraction on matrix $\boldsymbol{\tau}^{r-1}$ with respect to entry (i, j), we denote row i as *collapsed* in matrices $\boldsymbol{\tau}^r, \boldsymbol{\tau}^{r+1}, \dots, \boldsymbol{\tau}^{n-3}$. Observe that there is no $s \in [r+1, n-3]$ such that $\boldsymbol{\tau}^s$ is obtained from $\boldsymbol{\tau}^{s-1}$ through a contraction with respect to some entry (i, p) where $p \in [n] \setminus \{i, j\}$, as row i cannot be equivalent to any other row of matrix $\boldsymbol{\tau}^s$. In other words, each row (and column) is zeroed at most once in a contraction in matrices $\boldsymbol{\tau}^1, \boldsymbol{\tau}^2, \dots, \boldsymbol{\tau}^{n-3}$. Hence, Proposition 4 can be equivalently stated by not requiring to zero out collapsed rows. This observation will prove particularly useful in the remainder of the section.

Secondly, observe also that if $\boldsymbol{\tau}^0$ has a UBT realization, then $\boldsymbol{\tau}^s$ has a UBT realization for each $s \in [1, n-3]$, and, in particular, $\boldsymbol{\tau}^{n-3}$ has a UBT realization consisting of a star with 3 leaves. In this sense, there is no loss of generality in assuming that such a UBT realization is the star induced by the leaves labeled $n-2$, $n-1$, and n.

Hereafter, we provide the characterization of Θ_n. For brevity, when we state that matrix $\boldsymbol{\tau}^s$, $s \in [0, n-3]$, enjoys a particular property, such as the existence of a UBT realization, we specifically refer to the submatrix of $\boldsymbol{\tau}^s$ identified by its non-collapsed rows and columns. Given an integer $n \geq 3$, let z_{ij}^s, y_i^s, u_j^s, $\forall\, s \in [n-3], i \in [n], j \in [n]$, be binary decision variables, and consider the polytope $\mathcal{P}_n$ defined by the following conditions:

$$z^s_{i,j} \in \{0,1\} \quad \forall\, i,j \in [n], s \in [n-3] \tag{4.1}$$

$$z^s_{n-2,n-1} = z^s_{n-2,n} = z^s_{n-1,n} = 0 \quad \forall\, s \in [n-3] \tag{4.2}$$

$$\sum_{s=1}^{n-3} \sum_{j=i+1}^{n} z^s_{i,j} = 1 \quad \forall\, i \in [n-3] \tag{4.3}$$

$$\sum_{i=1}^{n-3} \sum_{j=i+1}^{n} z^s_{i,j} = 1 \quad \forall\, s \in [n-3] \tag{4.4}$$

$$z^s_{i,j} = 0 \quad \forall\, i \in [n], j \in [i], s \in [n-3] \tag{4.5}$$

$$y^s_i = \sum_{j=i+1}^{n} z^s_{i,j} \quad \forall\, i \in [n], s \in [n-3] \tag{4.6}$$

$$u^s_j = \sum_{i=1}^{j-1} z^s_{i,j} \quad \forall\, j \in [n], s \in [n-3] \tag{4.7}$$

$$\sum_{r=s+1}^{n} u^r_i \le (n-s)(1-y^s_i) \quad \forall\, i \in [n], s \in [n-3] \tag{4.8}$$

$$\tau^s_{i,j} = \tau^s_{j,i} \quad \forall\, i \in [n], j \in [i+1,n], s \in [n-3] \tag{4.9}$$

$$\tau^{n-3}_{n-2,n-1} = \tau^{n-3}_{n-2,n} = \tau^{n-3}_{n-1,n} = 2 \tag{4.10}$$

$$\tau^s_{i,i} = 0 \quad \forall\, s \in [n-3], i \in [n] \tag{4.11}$$

$$2z^s_{i,j} \le \tau^{s-1}_{i,j} \le (n-s) + (2+s-n)z^s_{i,j} \quad \forall\, i \in [n], j \in [i+1,n], s \in [n-3] \tag{4.12}$$

$$\tau^s_{i,k} - \tau^s_{j,k} \ge (3-n)\left(\tau^s_{i,j} - 2 + \sum_{r=1}^{s} y^r_i + y^r_j + y^r_k\right) \quad \forall\, s,i,j,k, i<j, i \ne j \ne k \tag{4.13}$$

$$\tau^s_{i,k} - \tau^s_{j,k} \le (n-3)\left(\tau^s_{i,j} - 2 + \sum_{r=1}^{s} y^r_i + y^r_j + y^r_k\right) \quad \forall\, s,i,j,k, i<j, i \ne j \ne k. \tag{4.14}$$

$$\tau^s_{i,j} \le \tau^{s-1}_{i,j} \quad \forall\, s \in [1,n-3], i,j \in [1,n] : i \ne j \tag{4.15}$$

$$\tau^s_{i,j} \ge \tau^{s-1}_{i,j} - u^s_i - u^s_j - (n-1)\sum_{r=1}^{s} y^r_i + y^r_j \quad \forall\, s \in [1,n-3], i<j \tag{4.16}$$

$$\tau^s_{i,j} \le \tau^{s-1}_{i,j} - u^s_i - u^s_j + \sum_{r \le s}(y^r_i + y^r_j) \quad \forall\, s \in [1,n-3], i<j. \tag{4.17}$$

We claim that $\boldsymbol{\tau}^0$ has a UBT realization if and only if conditions (4.1)–(4.17) are satisfied by some $z^s_{ij} \in \{0,1\}$ and $y^s_i, u^s_j, \tau^s_{i,j} \in \mathbb{R}$, $\forall\ s \in [n-3], i \in [n], j \in [n]$. We prove this claim by showing that $\mathcal{P}_n$ is a valid formulation for Θ_n. We start by observing that in each integral point of $\mathcal{P}_n$, $z^s_{i,j}$, y^s_i and u^s_j are binary for each $i,j \in [n]$ and $s \in [n-3]$; indeed, $y^s_i = 1$ if $z^s_{i,p} = 1$ for some $p \in [i+1,n]$ and $y^s_i = 0$ otherwise (condition (4.6)), and similarly $u^s_j = 1$ if $z^s_{q,j} = 1$ for some $q \in [j-1]$ and $u^s_j = 0$ otherwise (condition (4.7)).

PROPERTY 1. *In each integral point of $\mathcal{P}_n$, for each $s \in [n-3]$, $\boldsymbol{\tau}^s$ is a semimetric obtained from matrix $\boldsymbol{\tau}^{s-1}$ through exactly one contraction with respect to an entry (p,q) such that $\tau^{s-1}_{p,q} = \tau^{s-1}_{q,p} = 2$, and $\tau^{s-1}_{p,k} = \tau^{s-1}_{q,k}$ $\forall\ k \in [1,n] \setminus \{p,q\}$.*

Proof. We claim that such an entry (p,q) exists. Observe that there exists an entry (p',q'), where $p' < q'$ (condition (4.5)), such that $z^s_{\hat{p},\hat{q}} = 1$ for each $s \in [1,n-3]$ (condition (4.4)), and thus $\tau^{s-1}_{p',q'} = 2$ (condition (4.12)). Moreover, both rows p' and q' of $\boldsymbol{\tau}^{s-1}$ are not collapsed, and so conditions (4.14) and (4.14) are sufficient to ensure that such rows are equal, except for their p'-th and q'-th elements. Conditions (4.15)–(4.17) entail that $\tau^s_{i,q'} = \tau^{s-1}_{i,q'} - 1$ for each $i \in [n] \setminus \{p',q'\}$ such that i is not a collapsed row in matrices $\boldsymbol{\tau}^r$ $\forall$ $r \in [s]$. Observing that the submatrix obtained from $\boldsymbol{\tau}^{n-3}$ by considering its rows (and columns) $n-2$, $n-1$ and n is equal to matrix (2.8) (conditions (4.2), (4.10), and (4.11)), and each row $i \in [1, n-3]$ collapses exactly once in matrices $\boldsymbol{\tau}^s$, $s \in [n-3]$ (condition (4.3)), proves the claim. We conclude the proof by noting that matrix $\boldsymbol{\tau}^s$ is also semimetric, as it is symmetric (condition (4.9)) and it has a zero diagonal (condition (4.11)) for each $s \in [n-3]$.

PROPOSITION 13. *Given an integer $n \geq 3$, formulation $\mathcal{P}_n$ is valid for Θ_n.*

Proof. The statement directly follows from Property 1 and Proposition 4. If a matrix $\boldsymbol{\tau}$ of order $n \geq 3$ has a UBT realization, then there exists an integral point in $\mathcal{P}_n$ such that $\boldsymbol{\tau}^0 = \boldsymbol{\tau}$. Indeed, assume by contradiction that no such integral point exists. Then, Property 1 would imply that $\boldsymbol{\tau}^{n-3}$ has no UBT realization, contradicting Proposition 4. Conversely, in each integral point of $\mathcal{P}_n$, the submatrix obtained from $\boldsymbol{\tau}^{n-3}$ by considering its rows (and columns) $n-2$, $n-1$ and n is equal to (2.8) and thus has a UBT realization. Hence, matrix $\boldsymbol{\tau}^0$ has a UBT realization by Proposition 4. $\square$

Final remarks. Identifying the conditions beyond Buneman's ones that ensure both the preservation of the STI through contractions and the existence of an entry with a value of 2 in a matrix of order $n \geq 12$ is critical for advancing on the study of the polyhedral combinatorics of the BMEP. As of this writing, this challenge remains one of the most significant obstacles to developing ILP formulations for the BMEP that can scale to handle hundreds of taxa in practical computational time.

Acknowledgments. The first author acknowledges support from the Belgian National Fund for Scientific Research (FNRS) via the grant FNRS PDR 40007831, and the Fondation Louvain via the grant "COALESCENS". We dedicate this work to the memory of Douglass Stott Parker, who passed away prematurely and whose foundational work [14] has inspired us throughout all these years.

References.

[1] P. BUNEMAN, *The recovery of trees from measure of dissimilarities*, in Archaeological and Historical Science, F. R. Hodson, D. G. Kendall, and P. Tautu, eds., Edinburgh University Press, Edinburgh, UK, 1971, pp. 387–395.

[2] P. BUNEMAN, *A note on the metric properties of trees*, Journal of combinatorial theory, 17 (1974), pp. 48–50.

[3] D. CATANZARO, M. FROHN, O. GASCUEL, AND R. PESENTI, *A tutorial on the balanced minimum evolution problem*, European Journal of Operational Research, 300 (2022), pp. 1–19, https://doi.org/10.1016/j.ejor.2021.08.004, https://doi.org/10.1016/j.ejor.2021.08.004.

[4] D. CATANZARO, M. FROHN, AND R. PESENTI, *An information theory perspective on the balanced minimum evolution problem*, Operations Research Letters, 48 (2020), pp. 362–367.

[5] D. CATANZARO, M. LABBÉ, R. PESENTI, AND J. J. SALAZAR-GONZÁLES, *The balanced minimum evolution problem*, INFORMS Journal on Computing, 24 (2012), pp. 276–294.

[6] D. CATANZARO, R. PESENTI, A. SAPUCAIA, AND L. A. WOLSEY, *Optimizing over path-length matrices of unrooted binary trees*, tech. report, CORE, Université Catholique de Louvain, Belgium. Submitted to Mathematical Programming, 2023. Available at https://dial.uclouvain.be/pr/boreal/object/boreal%3A277008/datastream/PDF_01/view.

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

[8] S. L. Hakimi and S. S. Yau, *Distance matrix of a graph and its realizability*, Quarterly of Applied Mathematics, 22 (1964), pp. 305–317.

[9] OEIS Foundation Inc., *The On-Line Encyclopedia of Integer Sequences*, 2024. Published electronically at http://oeis.org.

[10] A. N. Patrinos and S. L. Hakimi, *The distance matrix of a graph and its tree realization*, Quaterly of Applied Mathematics, 30 (1972), pp. 255–269.

[11] J. M. S. S. Pereira, *A note on the tree realizability of a distance matrix*, Journal of Combinatorial Theory, 6 (1969), pp. 303–310.

[12] C. Semple and M. A. Steel, *Phylogenetics*, Oxford University Press, New York, NY, 2003.

[13] M. Steel, *Phylogeny: Discrete and random processes in evolution*, SIAM, Philadelphia, PA, 2016.

[14] D. Stott-Parker and P. Ram, *The construction of Huffman codes is a submodular ("convex") optimization problem over a lattice of binary trees*, SIAM Journal on Computing, 28 (1996), pp. 1875–1905.

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

Appendix A. Proof of Proposition 12 for $n \in [8, 11]$.

We prove the statement of the proposition by considering each case separately. Throughout the proof, we implicitly understand the condition that imposes the diagonal entries of the matrices to be zero. Moreover, for each of the considered cases, we recall the number m_n of families of semimetric matrices $\boldsymbol{\tau}$ of order n having UBT realizations that we expect to be implied by the conditions in the proposition hypotheses. We express a combination of p path-length sequences as a collection $\{k_i : \mathbf{s}_i,\ \forall\, i \in [p]\}$, where $\mathbf{s}_i$ is the i-th path-length sequence and k_i is the number of times it appears in the combination. For example, $\{4 : \{2,3,4,4\}, 1 : \{3,3,3,3\}\}$ denotes a combination that includes four path-length sequences $\{2,3,4,4\}$ and one given by $\{3,3,3,3\}$.

Case $n = 3$ ($m_3 = 1$). The only path-length sequence that satisfies Kraft's equality is $\vec{\boldsymbol{s}} = \{2,2\}$. The only matrix of order $n = 3$ with rows of type $\vec{\boldsymbol{s}}$, i.e., that can be associated to the combination $\{3 : \{2,2\}\}$, is the semimetric (2.8) with a UBT realization.

Case $n = 4$ ($m_4 = 1$). The only path-length sequence that satisfies Kraft's equality is $\vec{\boldsymbol{s}} = \{2,3,3\}$. Then, we consider only the combinations $\mathbf{c} = \{4 : \{3,2,2\}\}$. The only family of matrices satisfying the symmetry conditions that can be associated with the combination $\mathbf{c}$ is the family of semimetric matrices with UBT realizations whose canonical representative is

$$\boldsymbol{\tau} = \begin{bmatrix} 0 & 2 & 3 & 3 \\ 2 & 0 & 3 & 3 \\ 3 & 3 & 0 & 2 \\ 3 & 3 & 2 & 0 \end{bmatrix}.$$

Case $n = 5$ ($m_5 = 1$). The path-length sequences that satisfy Kraft's equality are $\{2,3,4,4\}$ and $\{3,3,3,3\}$. The only combination of these path-length sequences that satisfies the UBT-manifold condition is $\mathbf{c} = \{4 : \{2,3,4,4\}, 1 : \{3,3,3,3\}\}$. Finally, the only family of matrices satisfying the symmetry conditions that can be associated with the combination $\mathbf{c}$ is the family of semimetric matrices with UBT realizations whose canonical representative is

$$\boldsymbol{\tau} = \begin{bmatrix} 0 & 2 & 3 & 4 & 4 \\ 2 & 0 & 3 & 4 & 4 \\ 3 & 3 & 0 & 3 & 3 \\ 4 & 4 & 3 & 0 & 2 \\ 4 & 5 & 3 & 4 & 0 \end{bmatrix}. \tag{A.1}$$

Indeed, without loss of generality (w.l.o.g.), suppose that the first row of $\boldsymbol{\tau}$ is $\boldsymbol{\tau}_1 = [0, 2, 3, 4, 4]$. Then, $\boldsymbol{\tau}_2 = [2, 0, 3, 4, 4]$ as the first column of $\boldsymbol{\tau}$, by symmetry, is equal to $[0, 2, 3, 4, 4]^{\mathrm{T}}$. Also, $\boldsymbol{\tau}_3 = [3, 3, 0, 3, 3]$, as it is impossible for both $\boldsymbol{\tau}_2$ and $\boldsymbol{\tau}_3$ to contain two entries whose value is 4; otherwise, the fourth and fifth columns (and rows) of $\boldsymbol{\tau}$ would contain three entries whose value is 4. Thus, $\boldsymbol{\tau}_4 = [4, 4, 3, 0, 2]$ and $\boldsymbol{\tau}_5 = [4, 4, 3, 2, 0]$. Finally, we can generate all the semimetric matrices of the family whose canonical representative is (A.1) by symmetric permutations of rows and columns.

Case $n = 6$ ($m_6 = 2$). The path-length sequences that satisfy Kraft's equality are $\{2, 3, 4, 5, 5\}$, $\{2, 4, 4, 4, 4\}$, and $\{3, 3, 3, 4, 4\}$. The only combinations of these path-length sequences that satisfy the UBT-manifold condition are $\mathbf{c}_1 = \{6 : \{2, 4, 4, 4, 4\}\}$ and $\mathbf{c}_2 = \{4 : \{2, 3, 4, 5, 5\}, 2 : \{3, 3, 3, 4, 4\}\}$. Consider the combination $\mathbf{c}_1$. The only family of matrices satisfying the symmetry conditions and the WTI that can be associated with the combination $\mathbf{c}_1$ is the family of semimetric matrices with UBT realizations whose canonical representative is

$$\boldsymbol{\tau} = \begin{bmatrix} 0 & 2 & 4 & 4 & 4 & 4 \\ 2 & 0 & 4 & 4 & 4 & 4 \\ 4 & 4 & 0 & 2 & 4 & 4 \\ 4 & 4 & 2 & 0 & 4 & 4 \\ 4 & 4 & 4 & 4 & 0 & 2 \\ 4 & 4 & 4 & 4 & 2 & 0 \end{bmatrix}.$$

Indeed, w.l.o.g., suppose that the first row of $\boldsymbol{\tau}$ is $\boldsymbol{\tau}_1 = [0, 2, 4, 4, 4, 4]$. Then, $\boldsymbol{\tau}_2 = [2, 0, 4, 4, 4, 4]$. Again w.l.o.g., let $\boldsymbol{\tau}_3 = [4, 4, 0, 2, 4, 4]$ and $\boldsymbol{\tau}_4 = [4, 4, 4, 0, 2, 4]$. Note that the value of the fourth and the fifth entry of $\boldsymbol{\tau}_3$ and $\boldsymbol{\tau}_4$, respectively, is equal to 2 due to the WTI. As four entries of columns 5 and 6 have value 4, then $\boldsymbol{\tau}_5 = [4, 4, 4, 4, 0, 2]$ and $\boldsymbol{\tau}_6 = [4, 4, 4, 4, 2, 0]$.

Consider the combination $\mathbf{c}_2$. The only family of matrices satisfying the symmetry condition and the WTI that can be associated with the combination $\mathbf{c}_2$ is the family of semimetric matrices with UBT realizations whose canonical representative is

$$\boldsymbol{\tau} = \begin{bmatrix} 0 & 2 & 3 & 4 & 5 & 5 \\ 2 & 0 & 3 & 4 & 5 & 5 \\ 3 & 3 & 0 & 3 & 4 & 4 \\ 4 & 4 & 3 & 0 & 3 & 3 \\ 5 & 5 & 4 & 3 & 0 & 2 \\ 5 & 5 & 4 & 3 & 2 & 0 \end{bmatrix}.$$

W.l.o.g., suppose that the first row of $\boldsymbol{\tau}$ is $\boldsymbol{\tau}_1 = [0, 2, 3, 4, 5, 5]$; this implies that $\boldsymbol{\tau}_2 = [2, 0, 3, 4, 5, 5]$, and in turn they both entail $\boldsymbol{\tau}_4 = [4, 4, 3, 0, 3, 3]$ and impose columns 5 and 6 to have type $\{2, 3, 4, 5, 5\}$. Thus, $\boldsymbol{\tau}_5 = [5, 5, 4, 3, 0, 2]$ and $\boldsymbol{\tau}_6 = [5, 5, 4, 3, 2, 0]$ due to the WTI, and consequently $\boldsymbol{\tau}_3 = [3, 3, 0, 3, 4, 4]$.

Case $n = 7$ ($m_7 = 2$). The path-length sequences that satisfy Kraft's equality are:

$$\{2,3,4,5,6,6\}, \{2,3,5,5,5,5\}, \{2,4,4,4,5,5\}, \{3,3,3,4,5,5\}, \{3,3,4,4,4,4\}.$$

The only combinations that satisfy the UBT-manifold condition are

$$\mathbf{c}_1 = \{2 : \{2,3,4,5,6,6\}, 2 : \{2,3,5,5,5,5\}, 3 : \{3,3,3,4,5,5\}\}$$
$$\mathbf{c}_2 = \{4 : \{2,3,4,5,6,6\}, 2 : \{3,3,3,4,5,5\}, 1 : \{3,3,4,4,4,4\}\}$$
$$\mathbf{c}_3 = \{2 : \{2,3,5,5,5,5\}, 4 : \{2,4,4,4,5,5\}, 1 : \{3,3,4,4,4,4\}\}.$$

Observe that $n = 7$ is the smallest matrix order for which the UBT-manifold condition is not sufficient to ensure that each combination corresponds to a family of PLMs. Specifically, no matrix $\boldsymbol{\tau}$ satisfying the symmetry conditions and the WTI can be associated with combination $\mathbf{c}_1$. In fact, suppose, w.l.o.g., that a matrix $\boldsymbol{\tau}$ exists and its first row is $\boldsymbol{\tau}_1 = [0,2,3,5,5,5,5]$. This implies $\boldsymbol{\tau}_2 = [2,0,3,5,5,5,5]$ by the WTI. Then, the symmetry conditions require that the columns $1,2,4,5,6,7$ contain at least two entries whose value is 5, which is impossible.

Consider the combination $\mathbf{c}_2$. The only family of matrices satisfying the symmetry conditions and the WTI that can be associated with the combination $\mathbf{c}_2$ is the family of semimetric matrices with UBT realizations whose canonical representative is

$$\boldsymbol{\tau} = \begin{bmatrix} 0 & 2 & 3 & 4 & 5 & 6 & 6 \\ 2 & 0 & 3 & 4 & 5 & 6 & 6 \\ 3 & 3 & 0 & 3 & 4 & 5 & 5 \\ 4 & 4 & 3 & 0 & 3 & 4 & 4 \\ 5 & 5 & 4 & 3 & 0 & 3 & 3 \\ 6 & 6 & 5 & 4 & 3 & 0 & 2 \\ 6 & 6 & 5 & 4 & 3 & 2 & 0 \end{bmatrix}.$$

Indeed, suppose, w.l.o.g., that the first row of $\boldsymbol{\tau}$ is $\boldsymbol{\tau}_1 = [0,2,3,4,5,6,6]$. This implies that $\boldsymbol{\tau}_2 = [2,0,3,4,5,6,6]$ by the WTI, and that $\boldsymbol{\tau}_6$ and $\boldsymbol{\tau}_7$ are equivalent rows of type $\{2,3,4,5,6,6\}$, since the symmetry conditions impose $\tau_{61} = \tau_{71} = 6$. The symmetry conditions also require that $\tau_{41} = \tau_{42} = 4$ and thus $\boldsymbol{\tau}_4$ has type $\{3,3,4,4,4,4\}$. Consequently $\boldsymbol{\tau}_3$ and $\boldsymbol{\tau}_5$ have type $\{3,3,3,4,5,5\}$ and it is easy to verify that the only semimetric whose rows are of the given types is the canonical representative above.

Consider the combination $\mathbf{c}_3$. The only family of matrices satisfying the symmetry conditions and the WTI that can be associated with the combination $\mathbf{c}_3$ is the family of semimetric matrices with UBT realizations whose canonical representative is

$$\boldsymbol{\tau} = \begin{bmatrix} 0 & 2 & 3 & 5 & 5 & 5 & 5 \\ 2 & 0 & 3 & 5 & 5 & 5 & 5 \\ 3 & 3 & 0 & 4 & 4 & 4 & 4 \\ 5 & 5 & 4 & 0 & 2 & 4 & 4 \\ 5 & 5 & 4 & 2 & 0 & 4 & 4 \\ 5 & 5 & 4 & 4 & 4 & 0 & 2 \\ 5 & 5 & 4 & 4 & 4 & 2 & 0 \end{bmatrix}.$$

Indeed, suppose, w.l.o.g., that the first row of $\boldsymbol{\tau}$ is $\boldsymbol{\tau}_1 = [0,2,3,5,5,5,5]$, which implies $\boldsymbol{\tau}_2 = [2,0,3,5,5,5,5]$. Then, $\boldsymbol{\tau}_3 = [3,3,0,4,4,4,4]$ and $\tau_{i3} = 4$ for each $i \in \{4,5,6,7\}$.

The other values of the entries of rows $\boldsymbol{\tau}_i$, for $i \in \{4,5,6,7\}$, are finally determined by the symmetry conditions and the WTI.

Case $n = 8$ ($m_8 = 4$). The path-length sequences that satisfy Kraft's equality are:

$$\begin{gathered}\{2,3,4,5,6,7,7\}, \{2,3,4,6,6,6,6\}, \{2,3,5,5,5,6,6\},\\ \{2,4,4,4,5,6,6\}, \{2,4,4,5,5,5,5\}, \{3,3,3,4,5,6,6\},\\ \{3,3,3,5,5,5,5\}, \{3,3,4,4,4,5,5\}, \{3,4,4,4,4,4,4\}.\end{gathered}$$

The 28 combinations of these path-length sequences that satisfy the UBT-manifold condition are reported in Table 3.

There is no matrix $\boldsymbol{\tau}$ satisfying the symmetry conditions and the WTI that can be associated with combinations $\mathbf{c}_1, \ldots, \mathbf{c}_4, \mathbf{c}_6, \mathbf{c}_8, \mathbf{c}_9, \mathbf{c}_{12}, \ldots, \mathbf{c}_{27}$. In particular, all the combinations above, with the exception of $\mathbf{c}_{12}$, do not satisfy the *simplified* symmetry conditions and the WTI. In such cases, the existence of a matrix $\boldsymbol{\tau}$ corresponding to such combinations is immediately precluded by the values of the entries of one path-length sequence, or two of them if the existence of an entry equal to 2 make the WTI imply that $\boldsymbol{\tau}$ should have two equivalent rows. For example, consider $\mathbf{c}_3$. A row having type $\{3,4,4,4,4,4,4\}$ implies the existence of eight columns with at least one entry equal to 4, but there are only six row types in $\mathbf{c}_3$ satisfying such a property. Consider $\mathbf{c}_1$. A row having type $\{2,4,4,5,5,5,5\}$ implies, by the symmetry conditions and the WTI, the existence of six columns with at least two entries equal to 5, but there are only two row types in $\mathbf{c}_1$ satisfying such a property.

Combination $\mathbf{c}_{12}$ does not satisfy the symmetry conditions and the WTI. Assume that a matrix $\boldsymbol{\tau}$ corresponding to $\mathbf{c}_{12}$ exists and suppose that, w.l.o.g., the first row of $\boldsymbol{\tau}$ is $\boldsymbol{\tau}_1 = [0,2,3,4,6,6,6,6]$, which implies $\boldsymbol{\tau}_2 = [2,0,3,4,6,6,6,6]$ by the WTI. Symmetry conditions entail that rows $\boldsymbol{\tau}_i$, for $i \in \{5,6,7,8\}$, contain at least two entries equal to 6, hence they must have type $\{2,4,4,4,5,6,6\}$. Thus, rows $\boldsymbol{\tau}_3$ and $\boldsymbol{\tau}_4$ must have type $\{3,3,4,4,4,5,5\}$. This fact implies the existence of an entry $\tau_{4i} = 3$ for some $i \in \{5,6,7,8\}$. This fact contradicts the type $\{2,4,4,4,5,6,6\}$ of the columns symmetric to rows $\boldsymbol{\tau}_i$, for $i \in \{5,6,7,8\}$.

The combination $\mathbf{c}_5$ does not satisfy the symmetry conditions and the STI. Indeed, the following matrix:

$$\boldsymbol{\tau} = \begin{bmatrix} 0 & 2 & 3 & 4 & 5 & 6 & 7 & 7 \\ 2 & 0 & 3 & 4 & 5 & 6 & 7 & 7 \\ 3 & 3 & 0 & 4 & 3 & 5 & 6 & 6 \\ 4 & 4 & 4 & 0 & 3 & 4 & 4 & 4 \\ 5 & 5 & 3 & 3 & 0 & 3 & 5 & 5 \\ 6 & 6 & 5 & 4 & 3 & 0 & 3 & 3 \\ 7 & 7 & 6 & 4 & 5 & 3 & 0 & 2 \\ 7 & 7 & 6 & 4 & 5 & 3 & 2 & 0 \end{bmatrix}$$

is the only matrix whose row types correspond to $\mathbf{c}_5$. However, $\boldsymbol{\tau}$ satisfies Kraft's equalities, the symmetry conditions, the UBT-manifold condition and the WTI, but not the STI, as $\tau_{13} + \tau_{35} = \tau_{15} + 1$.

The only family of matrices that satisfies the symmetry conditions and the WTI that can be associated with the combinations $\mathbf{c}_7$, $\mathbf{c}_{10}$, $\mathbf{c}_{11}$, and $\mathbf{c}_{28}$ are, respectively, the families of semimetric matrices with UBT realizations, whose canonical representatives are

$$
\begin{array}{cc}
\begin{bmatrix}
0 & 3 & 4 & 4 & 4 & 4 & 4 & 4\\
3 & 0 & 3 & 3 & 5 & 5 & 5 & 5\\
4 & 3 & 0 & 2 & 6 & 6 & 6 & 6\\
4 & 3 & 2 & 0 & 6 & 6 & 6 & 6\\
4 & 5 & 6 & 6 & 0 & 2 & 4 & 4\\
4 & 5 & 6 & 6 & 2 & 0 & 4 & 4\\
4 & 5 & 6 & 6 & 4 & 4 & 0 & 2\\
4 & 5 & 6 & 6 & 4 & 4 & 2 & 0
\end{bmatrix}
&
\begin{bmatrix}
0 & 2 & 3 & 4 & 5 & 6 & 7 & 7\\
2 & 0 & 3 & 4 & 5 & 6 & 7 & 7\\
3 & 3 & 0 & 3 & 4 & 5 & 6 & 6\\
4 & 4 & 3 & 0 & 3 & 4 & 5 & 5\\
5 & 5 & 4 & 3 & 0 & 3 & 4 & 4\\
6 & 6 & 5 & 4 & 3 & 0 & 3 & 3\\
7 & 7 & 6 & 5 & 4 & 3 & 0 & 2\\
7 & 7 & 6 & 5 & 4 & 3 & 2 & 0
\end{bmatrix}
\\
\begin{bmatrix}
0 & 2 & 4 & 4 & 5 & 5 & 5 & 5\\
2 & 0 & 4 & 4 & 5 & 5 & 5 & 5\\
4 & 4 & 0 & 4 & 3 & 3 & 5 & 5\\
4 & 4 & 4 & 0 & 5 & 5 & 3 & 3\\
5 & 5 & 3 & 5 & 0 & 2 & 6 & 6\\
5 & 5 & 3 & 5 & 2 & 0 & 6 & 6\\
5 & 5 & 5 & 3 & 6 & 6 & 0 & 2\\
5 & 5 & 5 & 3 & 6 & 6 & 2 & 0
\end{bmatrix}
&
\begin{bmatrix}
0 & 2 & 4 & 4 & 5 & 5 & 5 & 5\\
2 & 0 & 4 & 4 & 5 & 5 & 5 & 5\\
4 & 4 & 0 & 2 & 5 & 5 & 5 & 5\\
4 & 4 & 2 & 0 & 5 & 5 & 5 & 5\\
5 & 5 & 5 & 5 & 0 & 2 & 4 & 4\\
5 & 5 & 5 & 5 & 2 & 0 & 4 & 4\\
5 & 5 & 5 & 5 & 4 & 4 & 0 & 2\\
5 & 5 & 5 & 5 & 4 & 4 & 2 & 0
\end{bmatrix}
\end{array}. \tag{A.2}
$$

Indeed, consider a matrix $\boldsymbol{\tau}$, whose rows have types in combination $\mathbf{c}_7$, and suppose, w.l.o.g., that the first row of $\boldsymbol{\tau}$ is $\boldsymbol{\tau}_1 = [0,3,4,4,4,4,4,4]$, which implies $\boldsymbol{\tau}_2 = [3,0,3,3,5,5,5,5]$, as there is a single row type that does not include any 4. Hence, $\boldsymbol{\tau}_3$ and $\boldsymbol{\tau}_4$ must have type $\{2,3,4,6,6,6,6\}$, and so $\boldsymbol{\tau}_3 = [4,3,0,2,6,6,6,6]$ and $\boldsymbol{\tau}_4 = [4,3,2,0,6,6,6,6]$. The other values of the rows $\boldsymbol{\tau}_i$ for $i \in \{5,6,7,8\}$ are immediately defined by the symmetry and STI conditions. Hence, the only feasible matrices $\boldsymbol{\tau}$ for $\mathbf{c}_7$ are those in the family of the matrix in the left upper corner of (A.2).

Consider a matrix $\boldsymbol{\tau}$ whose rows have the types in combination $\mathbf{c}_{10}$, and suppose, w.l.o.g., that the first row of $\boldsymbol{\tau}$ is $\boldsymbol{\tau}_1 = [0,2,3,4,5,6,7,7]$, which implies $\boldsymbol{\tau}_2 = [2,0,3,4,5,6,7,7]$ by the STI. Then, $\boldsymbol{\tau}_7$ and $\boldsymbol{\tau}_8$ must have type $\{2,3,4,5,6,7,7\}$, $\boldsymbol{\tau}_4$ and $\boldsymbol{\tau}_5$ must have type $\{3,3,4,4,4,5,5\}$, and τ_6 must have type $\{3,3,3,4,5,6,6\}$. Hence, $\boldsymbol{\tau}_3$ can only have type $\{3,3,3,4,5,6,6\}$. Accordingly, the values of the entries of $\boldsymbol{\tau}$ are immediately defined by the symmetry conditions and the STI. Hence, the only feasible matrices $\boldsymbol{\tau}$ for $\mathbf{c}_{10}$ are those in the family of the matrix in the right upper corner of (A.2).

Consider a matrix $\boldsymbol{\tau}$ whose rows have types in combination $\mathbf{c}_{11}$ and suppose, w.l.o.g., that the first row of $\boldsymbol{\tau}$ is $\boldsymbol{\tau}_1 = [0,2,4,4,5,5,5,5]$, which implies $\boldsymbol{\tau}_2 = [2,0,4,4,5,5,5,5]$ by the STI. Hence, $\boldsymbol{\tau}_3$ and $\boldsymbol{\tau}_4$ must have type $\{3,3,4,4,4,5,5\}$, and then, w.l.o.g., $\boldsymbol{\tau}_3 = [4,4,0,4,3,3,5,5]$ and $\boldsymbol{\tau}_4 = [4,4,4,0,5,5,3,3]$. The other values of the rows $\boldsymbol{\tau}_i$ for $i \in \{5,6,7,8\}$ are immediately defined by the symmetry conditions and the WTI. Hence, the only feasible matrices $\boldsymbol{\tau}$ for $\mathbf{c}_{11}$ are those in the family of the matrix in the left lower corner of (A.2).

Consider a matrix $\boldsymbol{\tau}$ whose rows have types in combination $\mathbf{c}_{28}$ and suppose, w.l.o.g., that the first row of $\boldsymbol{\tau}$ is $\boldsymbol{\tau}_1 = [0,2,4,4,5,5,5,5]$, which implies that $\boldsymbol{\tau}_2 = [2,0,4,4,5,5,5,5]$ by the STI. Then, the other values of the rows $\boldsymbol{\tau}_i$, for $i \in [3,8]$, are immediately defined by the symmetry conditions and the WTI. Hence, the only feasible matrices $\boldsymbol{\tau}$ for $\mathbf{c}_{28}$ are those in the family of the matrix in the right lower corner of (A.2).

The proof for $n \in [9,11]$ is code-assisted. The code, available at https://github.com/tmp-maker/Characterizing-Path-Length-Matrices-of-Unrooted-Binary-Trees, was used to automatically verify whether the combination of the path-length sequences that we generate satisfies the conditions in the statement of the proposition. In contrast, the verification that any combination satisfying all the conditions can be associated with a unique family of semimetric matrices with a UBT realization was carried out by using arguments similar to those applied in the case $n = 8$, which will not be reported in the following for the sake of brevity.

Case $n = 9$ ($m_9 = 6$). There are 469 path-length sequences that meet the UBT-manifold condition, but only 25 satisfy both the simplified symmetry conditions and the WTI. Of these, only 7 satisfy the standard symmetry conditions and the WTI, and only 6 also satisfy the STI. These 6 combinations correspond to 6 distinct families of semimetric matrices with UBT realizations, one family for each combination.

Case $n = 10$ ($m_{10} = 11$). There are 18,433 path-length sequences that meet the UBT-manifold condition, but only 586 satisfy both the simplified symmetry conditions and the WTI. From among these, only 34 satisfy the standard symmetry conditions and the WTI, and only 11 also satisfy the STI. These 11 combinations correspond to 11 distinct families of semimetric matrices with UBT realizations, one family for each combination.

Case $n = 11$ ($m_{11} = 18$). There are 2,158,260 path-length sequences that meet the UBT-manifold condition, but only 54,734 satisfy both the simplified symmetry conditions and the WTI. From among these, only 868 satisfy the standard symmetry conditions and the WTI, and only 18 also satisfy the STI. These 18 combinations correspond to 18 families of semimetric matrices with UBT realizations, one family for each combination.

	$\mathbf{c}_1$	$\mathbf{c}_2$	$\mathbf{c}_3$	$\mathbf{c}_4$	$\mathbf{c}_5$	$\mathbf{c}_6$	$\underline{\mathbf{c}_7}$	$\mathbf{c}_8$	$\mathbf{c}_9$	$\underline{\mathbf{c}_{10}}$	$\underline{\mathbf{c}_{11}}$	$\mathbf{c}_{12}$	$\mathbf{c}_{13}$	$\mathbf{c}_{14}$	$\mathbf{c}_{15}$	$\mathbf{c}_{16}$	$\mathbf{c}_{17}$	$\mathbf{c}_{18}$	$\mathbf{c}_{19}$	$\mathbf{c}_{20}$	$\mathbf{c}_{21}$	$\mathbf{c}_{22}$	$\mathbf{c}_{23}$	$\mathbf{c}_{24}$	$\mathbf{c}_{25}$	$\mathbf{c}_{26}$	$\mathbf{c}_{27}$	$\underline{\mathbf{c}_{28}}$
{2,3,4,5,6,7,7}					4			2	2	4			4	2				2	2		2							
{2,3,4,6,6,6,6}	4	2	2				2	2				2		2	4	2				4		2	2					
{2,3,5,5,5,6,6}		2	2	4		4					4						2	2			2	2		2	4	2		
{2,4,4,4,5,6,6}		2		2			4		2			4				2	4							2			6	
{2,4,4,5,5,5,5}	2		2			2			2		2					2			4				4	2		4		8
{3,3,3,4,5,6,6}					2			3	1	2				1	2			3	1		1	2			4	2	2	
{3,3,3,5,5,5,5}					1	1	1						3	2	1	1	1			4	3	2	2	2				
{3,3,4,4,4,5,5}			1	1						2	2	2	1	1	1	1	1	1	1									
{3,4,4,4,4,4,4}	2	2	1	1	1	1	1	1	1																			

Table 3: Path-length sequences for $n = 8$ (first column) and their combinations that satisfy the manifold and the WTI conditions. The combinations that correspond to semimetric matrices having UBT realizations are underlined.