Tequio *4*(11), 2021: 27-40
ISSN: 2594-0546

On some polytopes in phylogenetics
Politopos en filogenética

Linard Hoessly[1]

Fecha de recepción: 6 de noviembre de 2020
Fecha de aceptación: 23 de diciembre de 2020

Resumen - Presentamos las nociones matemáticas utilizadas en filogenética y tres clases de politopos de la filogenética. El Tight span y el politopo de Lipschitz se asocian a espacios métricos finitos y pueden conectarse a incrustaciones que conservan la distancia, mientras el politopo de evolución mínima balanceada (BME) se asocia con números naturales.

Palabras clave: Filogenética, politopo, espacio métrico finito, politopo fundamental, tight span.

Abstract - We introduce mathematical notions used in phylogenetics and three sorts of phylogenetics polytopes. The Tight span and the Lipschitz polytope are both associated to finite metric spaces and can be connected to distance-preserving embeddings, while the balanced minimum evolution (BME) polytope is associated to natural numbers.

Keywords: Phylogenetics, Polytope, Finite metric space, Fundamental polytope, Tight span.

1. Introduction

Phylogenetics studies the methods and the practice of identifying evolutionary relationships among biological species. Finding such relationships is a current focus of research, and is usually performed via phylogenetic inference based on mathematical models of evolution (Semple & Steel, 2003; Steel, 2016), which are represented as phylogenetic trees or networks (Huson, Rupp, & Scornavacca, 2010). Usually, genetic material is transferred from parents to offspring, resulting in tree-like representations. However, different biological species can transfer genetic information between otherwise unrelated organisms. Horizontal gene transfer e.g. is a mechanism where genetic material from one species is moved to another one which is relevant in how bacteria acquire antibiotic resistance. This suggests the possibility that corresponding parts of the evolutionary history might not be tree-like, and such relationships are often represented via phylogenetic networks. There are different approaches to phylogenetic reconstruction. We briefly introduce and elaborate on distance-based and likelihood-based methods. Distance-based techniques first compute a pairwise distance-like function between the taxa to construct a phylogenetic tree T (or structure) that best represents the distances obtained, usually via some optimality criterion. Distance-based methods are popular as they tend to be fast. Concerning likelihood-based methods there are two main paradigms: maximum likelihood (ML) and bayesian methods. In both, evolution is described through probabilistic model of sequence evolution, enabling in principle computations of likelihoods of observing the data given the model and its parameters. While these methods are assumed to be more correct from a foundational level, corresponding computations can be slow.

[1] Department of Mathematical Sciences, University of Copenhagen, Copenhagen, Denmark; email: hoessly@math.ku.dk ORCID-ID: 0000-0002-2745-2141

Many fascinating objects originated from methods and structures used to understand the evolutionary history of species (Dress, Huber, Koolen, Moulton, & Spillner, 2011; Semple & Steel, 2003). We will first introduce objects from discrete mathematics and then focus on three polytopes which can be related to objects of interest in phylogenetics. Our treatment does not aim to be comprehensive in its scope, as these are fairly developed fields. In our exposition and treatment we mostly focus on phylogenetic trees, tree-like metric spaces and corresponding polytopes.

2. Introduction to some discrete objects

We introduce notions related to the combinatorics of phylogenetics, i.e. graphs in ξ 2.1, polytopes in ξ 2.2, finite metric spaces and splits in ξ 2.3 and phylogenetic trees in ξ 2.4.

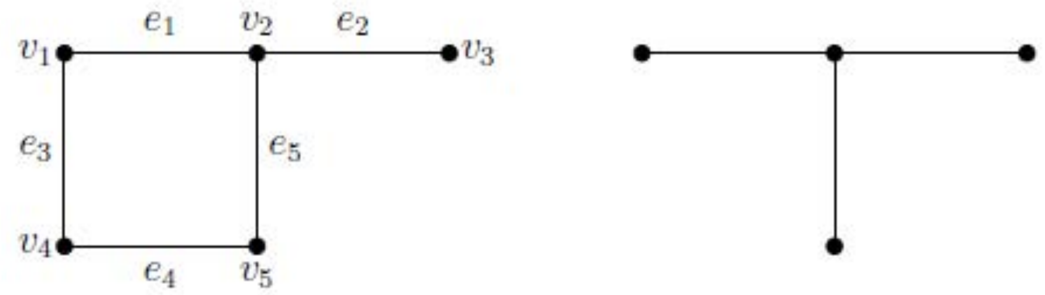
2.1 Graphs and trees

In phylogenetics, graphs and trees are used to represent evolutionary relationships between species. We will focus on undirected graphs,[2] since this is our main setting of interest.

A finite undirected graph $G = (V, E)$ consists of vertices $V = V(G)$ and edges $E \subseteq V^2$, written as $E = E(G)$. A *path* is a sequence $e_0, e_1, \cdots, e_n$ of edges which join a sequence of distinct vertices. Graphs in which two arbitrary vertices are connected by exactly one path are called *trees*, as an example consider Figure 1. A graph G is *connected* if there is a path between any two vertices. The *degree* $deg(v)$ of a vertex $v \in V$ is the number of edges incident to v.

Figure 1.

Only the right graph is a tree.

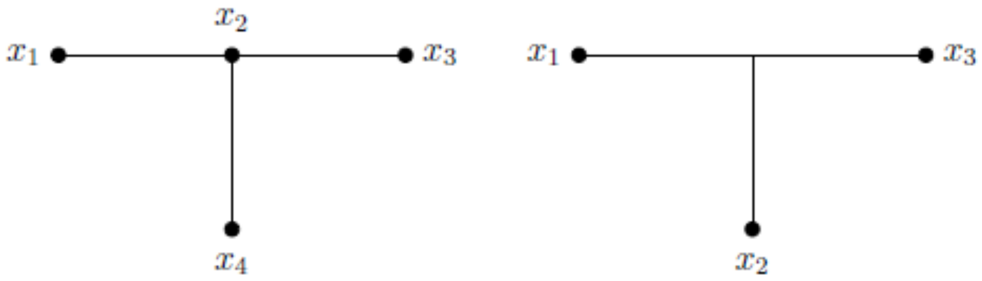


Lemma 2.1 ((Bollobas, 1998, equation (1), p.4 and ξ 1.2)) Let $G = (V, E)$ be a connected graph. Then $\sum_{v \in V} d\,eg(v) = 2|E|$, and furthermore $|E| = |V| - 1$ if and only if G is a tree.

Next we introduce a particular notion of a tree used in phylogenetics. Let X be a set. An X-tree is a labelling of some of the vertices of a tree T, where every leaf of T is labelled. We denote an X-tree by (T, ϕ), where $\phi: X \to V(T)$ is the labelling map. When all internal vertices are of degree 3, we call it binary X-tree.

[2] I.e. graphs where the edges are not directed.

Figure 2.
Two different binary X-trees on the tree T of figure 1.



2.2 Polytopes

Polytopes are seemingly simple geometric objects with flat sides. They appear as convex hulls of a finite set of points in Euclidean space (like, e.g., the plane $\mathbb{R}^2$ or 3-dimensional space $\mathbb{R}^3$), and exhibit a rich variety of combinatorial structures (Ziegler, 1995). The convex hull of a set of points $\{a_1, \cdots, a_m\} \subset \mathbb{R}^n$ is defined as

$$\mathrm{conv}\{a_1, \cdots, a_m\} := \{x \in \mathbb{R}^n \mid x = \sum_{i=1}^{m} \lambda_i \, a_i, \sum_{i=1}^{m} \lambda_i = 1, \lambda_i \geq 0\}$$

A *polytope* is a convex hull of a finite set of points. Well-known examples include two-dimensional polytopes that are convex polygons like the square (cf. Figure 3). The *dimension* of a polytope P is the dimension of the smallest Euclidean space which could contain it. As an example, the square of figure [fig_octa] has dimension two. A *face* of a polytope P is any intersection of the polytope with a half-space such that none of the interior points of the polytope lie on the boundary of the half-space. Any face of a polytope is a polytope itself. Some faces have a special name, faces of dimension $0{,}1$ and $\dim(P) - 1$ are called *vertices, edges and facets*. Moreover, the faces of polytopes can be ordered by inclusion, giving the poset of faces. A rougher invariant are its *face numbers* $f_0^P, \dots, f_{\dim(P)}^P$, which are defined as

$$f_i^P = \#\{i - \text{dimensional faces of } P\}.$$

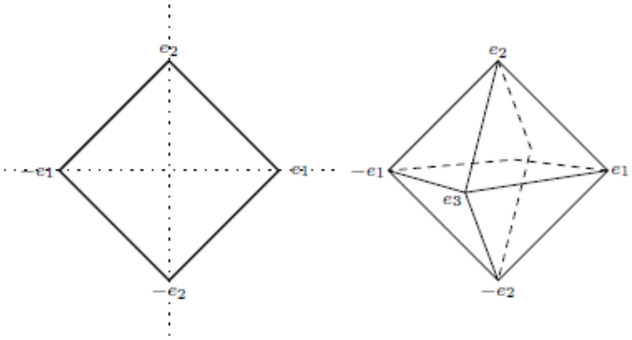
Putting all the face numbers together gives a convenient way of writing them as the so-called f-vector $(f_0^P, \dots, f_{m-1}^P)$, where $m = \dim(P)$. Note that convex polytopes may equivalently be defined as an intersection of a finite number of half-spaces, corresponding to the so-called *hyperplane description*, see, e.g., (Ziegler, 1995, §2.4).

Example 2.2. Consider the *d-crosspolytope*, which is defined as

$$\beta_d := \mathrm{conv}\{e_1, -e_1, \cdots, e_d, -e_d\} \subseteq \mathbb{R}^d,$$

where e_1 is the unit vector with entry one in the first coordinate and zeros otherwise, i.e., e_i is the vector with the only nonzero entry one in the i-th[3] coordinate.

Figure 3.
A square (β_2) and an octahedron (β_3) with f-vectors (4; 4) and (6; 8; 8).



Another interesting class of polytopes are zonotopes, which are Minkowski sums of lines. Their combinatorial structure connects to hyperplane arrangements, tilings or oriented matroids (Ziegler, 1995, ξ 7). As an example consider the square of figure 3 as the sum of the lines $[e_1, e_2]$, $[e_1, -e_2]$.

2.3 Finite metric spaces and splits

Let X be a set. A metric (or distance function) on X is a symmetric function $d: X \times X \to \mathbb{R}_{\geq 0}$ such that
(1) For all $x, y \in X$, $d(x, y) = 0$ implies $x = y$.
(2) For all $x, y, z \in X$, $d(x, z) \leq d(x, y) + d(y, z)$ ("triangle inequality").
If condition (1) is dropped, then d is called a pseudometric. In the following we will focus on *finite metric spaces* with $|X| < \infty$.

Example 2.3 (Metric spaces from weighted graphs) A *weighting* of a graph G is any function $w: E(G) \to \mathbb{R}_{>0}$, and the pair (G, w) is called a *weighted graph*. Set

$$d_w(v, v') := \min\{w(e_1) + \cdots + w(e_k) \mid v, e_1, v_1 \ldots, e_k, v' \text{ is a path joining } v \text{ with } v'\}$$

such that the pair $(V(G), d_w)$ is a metric space.

If (G, w) represents (X, d) it is called a *graph realisation* of the metric space. Note that any finite metric space has a graph realisation from the complete graph[4] by setting the weight of the edge $e_{i,j}$ between i, j to $d(i, j)$. Next, we introduce metric spaces coming from X-trees.

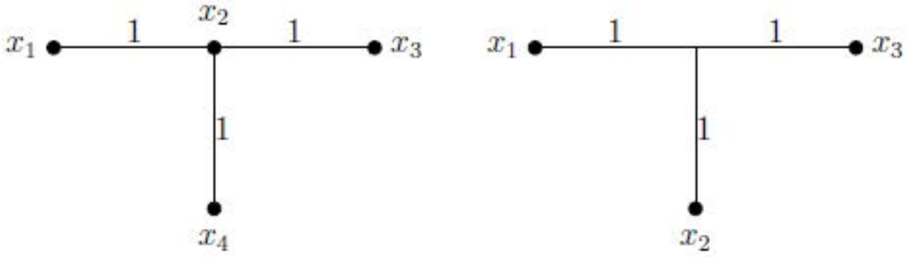
[3] I.e. i stands for any of the elements i∈{1,⋯,d}.
[4] The complete graph on a set of vertices is the graph where any two vertices are connected to each other through an edge.

Definition 2.4 (Tree-like metrics) *A (pseudo)metric d on a set X is called a* tree-like (pseudo)metric *if there exists an X-tree (T, ϕ) and a weighting w of T such that for all $x, y \in X$*

$$d(x, y) = d_w(\phi(x), \phi(y)).$$

The pseudometric d is a metric if and only if ϕ is injective.

Figure 4.
Two *X*-trees with edge weight one for each edge.



Next we consider splits. Let X be a finite set.

- A *split* of X is a bipartition of X, i.e., a pair of disjoint subsets $A, B \subseteq X$ such that the union[5] $A \cup B = X$, which is written as $A|B$.
- Two splits $A|B$ and $C|D$ are *compatible* if at least one of the four intersections[6] $A \cap C, A \cap D, B \cap C, B \cap D$ is empty.
- A *system of splits* on X is just a set of splits of X; the system is called [compatible]*compatible* if its elements are pairwise compatible.

There are more general definitions for split systems, e.g. weakly compatible or circular splits (Semple & Steel, 2003, x 3.8 or x 7.4). Next we consider weightings on splits.

Definition 2.5 A *weighted split system is a pair $(\mathcal{S}, \alpha)$ where $\mathcal{S}$ is a system of splits on X and $\alpha \in (\mathbb{R}_{\geq 0})^{\mathcal{S}}$ is any weighting. Any such weighted split system defines a nonnegative function $d_\alpha: X \times X \to \mathbb{R}$* via $d_\alpha(x, y) = \sum_{\sigma \in \mathcal{S}} \alpha_\sigma \, \delta_\sigma(x, y)$ *where δ_σ is defined for $\sigma = A|B$ as*

$$\delta_\sigma(i, j) = \begin{cases} 0 & i, j \in A \text{ or } i, j \in B \\ 1 & \text{otherwise.} \end{cases}$$

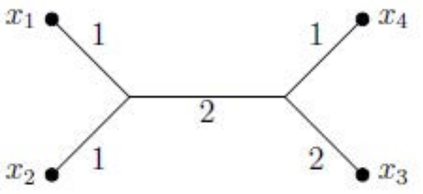
The functions of the form d_α are called *split-decomposable (pseudo)metrics* associated to $\mathcal{S}$, where (X, d_α) is a pseudometric space. A *positively weighted* split system is one where $\alpha_\sigma > 0$ for all $\sigma \in \mathcal{S}$.
For metric spaces from weighted trees we have the following.

[5] The union of two sets A,B which is denoted as A∪B is the set containing all the elements that are either in A or in B.
[6] The intersection of two sets A,C which is denoted A∩C is the set of all elements that are both in A and in C.

Theorem 2.6 ((Semple & Steel, 2003, Theorems 3.1.4, 7.1.8, 7.3.2)) *Let (X, d) be a pseudometric space. The following are equivalent:*
(i) d is a tree-like pseudo-metric on X (in the sense of Definition [df:tm2]).
(ii) d is a split-decomposable pseudometric associated to a positively weighted system of compatible splits. Moreover, this system is unique.
Under the equivalence of (I) with (II), splits in the decomposition of the metric correspond bijectively[7] to edges in the tree.

Example 2.7 *Consider the metric on $X = \{x_1, x_2, x_3, x_4\}$ given as follows*

$d(x_i, x_j)$	x_1	x_2	x_3	x_4
x_1	0	2	5	4
x_2	2	0	5	4
x_3	5	5	0	3
x_4	4	4	3	0



The metric is tree-like, where the underlying tree can be illustrated in the sense of Definition [df:tm2] as above. With Theorem [tree], the corresponding splits can be read off the graph leading to the decomposition of the distance as

-$x_1|x_2, x_3, x_4, \quad x_2|x_1, x_3, x_4, \quad x_3|x_1, x_2, x_4, \quad x_4|x_1, x_2, x_3, \quad x_1, x_2|x_3, x_4$
- $d(\cdot,\cdot) = \delta_{x_1|x_2,x_3,x_4} + \delta_{x_2|x_1,x_3,x_4} + \delta_{x_3|x_1,x_2,x_4} + \delta_{x_4|x_1,x_2,x_3} + 2 \cdot \delta_{x_1,x_2|x_3,x_4}$

Remark 2.8 *For a finite metric space (X, d), it is often of interest to obtain a decomposition of the metric into a sum of more elementary parts. One possible family of functions are the δ_σ from splits of Definition 2.5.*

Remark 2.9 *There is a more general theory for decompositions into weighted split systems. (Bandelt & Dress, 1992, Theorem 2) says that any metric (X, d) can be uniquely decomposed into $d = d_0 + \sum_{\sigma\in\mathcal{S}} \alpha_\sigma \, \delta_\sigma$, where d_0 is split prime and $\mathcal{S}$ is a (unique) weakly compatible system of splits.[8] Furthermore if in this decomposition $d_0 = 0$, then the metric is called totally split decomposable.*

2.4 Phylogenetic trees

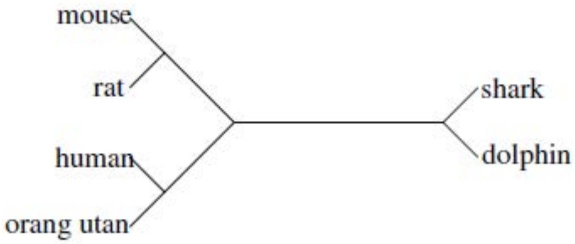
Phylogenetic trees describe evolutionary relationships, and we will mostly focus on undirected phylogenetic trees. However, both directed versions and networks are also used in phylogenetics, see, e.g., (Huson *et al.*, 2010).

[7] I.e. in a one-to-one relationship.
[8] In (Bandelt & Dress, 1992), a split prime metric is such that it is not further decomposable with respect to split metrics.

A *phylogenetic tree* is an X-tree where only the leaves are labelled and all internal vertices have a degree of at least 3. As in the case of X-trees, a phylogenetic tree is binary if all internal vertices have degree three. As a more concrete example of a binary phylogenetic tree consider

Figure 6.

A binary phylogenetic tree.



For $n = |X| \geq 3$ denote by $\mathcal{T}_n$ the set of all binary X-trees with n leaves. If the context is clear we will also simply say binary tree for binary phylogenetic X-trees.

Given an X-tree, there is an associated system of splits on X obtained by considering the two connected components obtained by the removal of e in T for each edge $e \in E(T)$. Denote the so-obtained set of splits by $\sum(T)$. For an example we refer to example [ex_dist]. On the other hand, by Theorem [tree], we get that for $\sum$ a system of splits, there is an X-tree T such that $\sum = \sum(T)$ if and only if the system of splits is compatible.

More general split systems are employed for generalizations of unrooted phylogenetic trees, where graphs in such split networks are not necessarily trees, and one or more edges in the graph are used to represent a split (Dress *et al.,* 2011; Huson *et al.,* 2010).

3. Polytopes in phylogenetics

Both the Tight span and the Lipschitz polytope are associated to a (finite) metric space and relate to a distance-preserving embedding in a bigger space. The minimum evolution polytope on the other hand is associated to natural numbers $n \in \mathbb{N}_{\geq 3}$. In the following, we aim to introduce and motivate the main objects. However, the topics are mature research directions and we restrict to a non-exhaustive treatment.

3.1 Tight span

Isbell studied the tight span in his investigation of injectivity for metric spaces (Isbell, 1964). In phylogenetics, it appeared in relation to reconstruction of phylogenetic trees from finite metric spaces (Dress, 1984). Representations of distances of phylogenetic trees can be seen as a connected one-dimensional polytope. Distances between vertices correspond to the sum of the edge lengths of the shortest paths. Hence it is natural to ask whether we can embed a given finite metric space distance-preserving into a low-dimensional compact polytope. One such possibility is the so-called Tight span.

The Tight span often helped to establish properties of classes of metrics, particularly in relation to decompositions that are of interest in phylogenetics. Furthermore, the 1-skeleton[9] of the Tight span is a graph realisations of the metric (Dress, 1984). For more on the motivation and connection of the study of Tight span

[9] I.e. the one dimensional faces.

to phylogenetics we refer to, e.g., (Dress *et al.,* 2011) or (Huson *et al.,* 2010), and for a concrete algorithmic application to, e.g., First we consider an unbounded polytope $U_{(X,d)} := \{z \in \mathbb{R}^X \mid z_i + z_j \geq d(i,j) \ \forall i,j \in X\}$.

Definition 3.1 The *Tight span* of (X,d) is given by the minimal points of $U_{(X,d)}$, which are defined as $T_{(X,d)} := \{z \in U_{(X,d)} \mid y \in U_{(X,d)} \text{ and } y \leq z \text{ implies } z = y\}$.

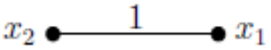
Note that the Tight span $T_{(X,d)}$ corresponds to the bounded faces of the polyhedron $U_{(X,d)}$. The Tight span is a polytopal complex that is associated to any finite metric space (X,d), whose structure often catches features of (X,d). The Kuratowski embedding is a map $f_{(X,d)}: X \to \mathbb{R}^X$ that sends elements of $X = \{x_1, \cdots, x_n\}$ to its Tight span, while preserving their pairwise distance. It is defined as

$$\begin{aligned} f_{(X,d)}: \quad & X \to \mathbb{R}^X \\ & x_i \mapsto f_{(X,d)}(x_i) := (d(x_i, x_j))_{j \in X} \end{aligned}$$

We have the following.

Lemma 3.2 The function $f_{(X,d)}: (X,d) \to (T_{(X,d)}, ||\cdot||_\infty)$ (where $T_{(X,d)} \subseteq \mathbb{R}^X$) is an isometric map into the tight span $T_{(X,d)}$, where for $z \in \mathbb{R}^X$, $||z||_\infty := \max_{x_i \in X}\{|z_i|\}$.

Example 3.3 *Consider the metric on $X = \{x_1, x_2\}$ with $d(x_1, x_2) = 1$. As a tree-like metric, it can be illustrated as*

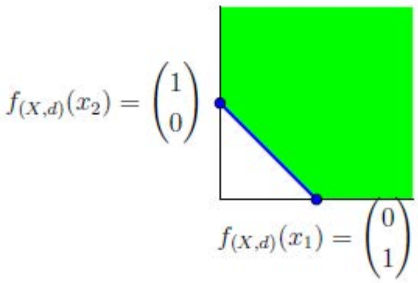


The upper map sends x_1 to $\begin{pmatrix} d(x_1,x_1) \\ d(x_1,x_2) \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, x_2 to $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$, hence

$$||f_{(X,d)}(x_1) - f_{(X,d)}(x_2)||_\infty = ||\begin{pmatrix} 1 \\ 0 \end{pmatrix} - \begin{pmatrix} 0 \\ 1 \end{pmatrix}||_\infty = 1 = d(x_1, x_2),$$

and the Tight span looks as follows.

Figure 7.
The Tight-span as the blue line.


$f_{(X,d)}(x_2) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$
$f_{(X,d)}(x_1) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$

Example 3.3 generalises as follows to tree-like metric spaces, which can be classified via their Tight span

Theorem 3.4 *(Dress, 1984, Theorem 8)* The metric space (X,d) is tree-like if and only if the tight span $T_{(X,d)}$ is an $\mathbb{R}$-tree.[10]

This has been generalised to show that for (X,d) a finite metric that is totally decomposable, the Buneman graph D representing the split decomposition is contained in the 1-skeleton of the tight span $T_{(X,d)}$ (cf., i.e., (Huson *et al.,* 2010, x 5.12)). Injectivity of metrics corresponds to some factorisation property, where Isbell showed that $(T_{(X,d)}, ||\cdot||_\infty)$ is injective and that every metric can be isometrically embedded into this metric space (Isbell, 1964).

3.2 Lipschitz polytope

Studying fundamental polytopes was proposed by Vershik (Vershik, 2015) as an approach to a combinatorial classification of metric spaces. It also relates to an isometric embedding of the metric space, however, through optimal transport. As in the case of the Tight span it can be expected that properties of metric spaces can be connected to properties of the fundamental polytope.

The polar dual of the fundamental polytope consists of the real-valued functions with Lipschitz constant at most 1, called Lipschitz polytope. As polar duality preserves all combinatorial data, it is enough to classify the combinatorial structure of Lipschitz polytopes. We will mostly focus on Lipschitz polytopes in the following.

The structure of fundamental polytopes of tree-like metric spaces were studied via associated hyperplane arrangements and corresponding decompositions of the matroid in (Delucchi & Hoessly, 2020), enabling explicit formulas for face numbers of tree-like finite metric spaces. Values of the f-vectors as well as concrete values for f-vectors for "generic"[11] metrics were given in (Gordon & Petrov, 2017). For more on connections, terminology, history and further context around fundamental polytopes we refer to, e.g., (Ostrovska & Ostrovskii, 2019, ξ 1.6) or (Delucchi & Hoessly, 2020), where we further remark that direct applications to phylogenetics are still outstanding.

Definition 3.5 *The* Lipschitz polytope *of (X,d) is given as an intersection of halfspaces by*

$$LIP(X,d) := \left\{x \in \mathbb{R}^X \mid \textstyle\sum_i x_i = 0,\ x_i - x_j \leq d(i,j)\ \forall i,j \in X\right\}. \tag{1}$$

Next we concentrate on the case of tree-like metric spaces and their Lipschitz polytopes as in (Delucchi & Hoessly, 2020). Let X be a finite set and consider a split $\sigma = A|B$ of X, where $|X| = n$. To σ we associate the

$$S_\sigma := \operatorname{conv}\left\{\frac{|B|}{n}\cdot \mathbb{1}_A - \frac{|A|}{n}\cdot \mathbb{1}_B, \frac{|A|}{n}\cdot \mathbb{1}_B - \frac{|B|}{n}\cdot \mathbb{1}_A\right\} \subseteq \mathbb{R}^X$$

line segment (one-dimensional polytope)
:::
where Accordingly, associated to a split system $\mathcal{S}$ we define the zonotope defined by the Minkowski sum $Z(\mathcal{S}) := \sum_{\sigma\in\mathcal{S}} S_\sigma$.

[10] An R-tree (also called real trees) in some R^n corresponds to the points of a graph-theoretical embedding of a tree.
[11] In (Gordon & Petrov, 2017), finite metric spaces are called generic if the triangle inequality is always strict and the fundamental polytope is simplicial.

Then the form of Lipschitz polytopes of finite tree-like spaces can be given as follows.

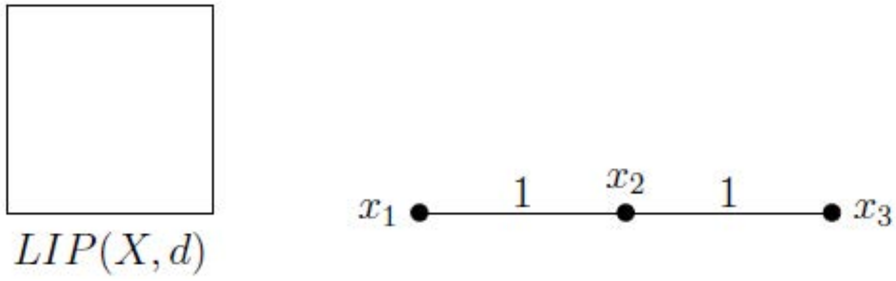
Theorem 3.6 *(Delucchi & Hoessly, 2020, Theorem 3.1) Let (X, d) be a tree-like pseudometric space. Then, $LIP(X, d) = \sum_{\sigma \in \mathcal{S}} \alpha_\sigma \, S_\sigma$ where $(\mathcal{S}, \alpha)$ is the unique weighted system of compatible splits of X such that $d = d_\alpha$ (cf. Theorem 2.6).*

$$\mathbb{1}_A := \sum_{x \in A} \mathbb{1}_x.$$

We next go through an example.

Example 3.7 (Points in $\mathbb{R}^1$) *Distances defined by a set of n points in $\mathbb{R}^1$ come from a metric graph in a line. The associated set of splits from the split-metric are compatible, as such distances are tree-like. Consider the following metric on*
$X = \{x_1, x_2, x_3\}$.

Figure 8.
Lipschitz polytope as a square and graph realisation.



3.3 Minimal evolution polytope

The minimal evolution polytope (BME polytope) originates from the distance based approach to phylogenetic reconstruction. We will first give an intuitive description and then give the definition.

Assume we are given a distance function on the set of taxa, and we are looking for a corresponding phylogenetic representation. Assuming tree-likeness, we look for the best distance from a tree in order to represent the data at hand. One such method is the Balanced Minimum Evolution (BME) principle, that builds on a tree length calculation from (Pauplin, 2000) where the total tree length for phylogenetic trees can be computed via pairwise distances and the number of edges between the leaves. This is in contrast to simply summing all edge lengths in the tree.

Assume we are looking for a tree-like phylogenetic representation while only knowing distances obtained from data. Then, if the distance is from a tree, the correct tree topology minimises the total tree length. Applying this minimisation procedure is the BME method.

The tree with minimal tree length can be found by computing the tree lengths over all possible phylogenetic trees, or equivalently by minimizing over the BME polytope, which allows to reformulate the BME problem as a linear programming problem[12] (Haws, Hodge, & Yoshida, 2011).

While the BME method is just a heuristic, the Neighbor joining method[13] was shown to be a greedy version for the BME method (Gascuel & Steel, 2006).

The combinatorial structure of the BME polytope is of interest for the application in algorithms and as a basic mathematical object in phylogenetics. Some properties of the structure of the BME polytope are in (Eickmeyer, Huggins, Pachter, & Yoshida, 2008; Haws *et al.,* 2011), which were extended to the study of facets in (Forcey, Keefe, & Sands, 2016), whereas a direct algorithmic application is, e.g., in (Lefort, Desper, & Gascuel, 2015).

We represent distance functions $d: X \times X \to \mathbb{R}$ by a vector $D \in \mathbb{R}^{\binom{n}{2}}$, where we index entries of any $v \in \mathbb{R}^{\binom{n}{2}}$ by $\{i,j\} \subset X$ via lexicographic order, so we write v as $v = (v_{1,2}, v_{1,3}, \cdots, v_{n-1,n})$. For every labelled binary tree T on n vertices we consider the associated vector $w^T \in \mathbb{R}^{\binom{n}{2}}$ defined by the entries $w^T_{i,j} := 2^{n-l-2}$ where l is the number of interior nodes of the shortest path between i, j in T. Note that these vectors w^T depend only on the tree topology.

The balanced tree length estimation $l(T)$ of Pauplin is given by

$$l(T) := \sum_{i,j;i<j} w^T_{i,j}\, d(i,j).$$

Note that this is just the dot-product of the vector w^T and and the pairwise distances D, i.e. $l(T) = w^T \cdot D$. The BME principle aims at finding the tree T that minimises the above balanced tree length estimation. In (Haws *et al.,* 2011) they showed that minimising over all trees in $\mathcal{T}_n$ is equivalent to minimising over the convex hull of all the vectors w^T, where $T \in \mathcal{T}_n$.

BME polytopes are associated to natural numbers $n \in \mathbb{N}_{\geq 3}$, and not to distances (i.e. metric spaces) as in the case of the polytopes of ξ 3.1 and ξ 3.2. We define the BME(n) polytope as follows.

Definition 3.8 *The BME(n) polytope* $\mathcal{P}_n$ *for* $n \geq 3$ *is defined as*

$$\mathcal{P}_n := \mathrm{conv}\{w^T \mid T \in \mathcal{T}_n\}.$$

As an example consider

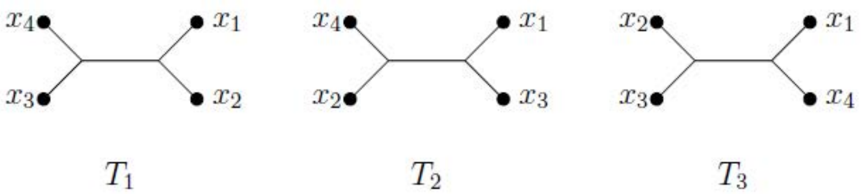
Example 3.9 (*Eickmeyer et al., 2008*) *Consider the case* $n = 4$. *Then first we look at* $\mathcal{T}_4$, *which consists of the following binary trees:*

[12] Linear programming or LP is a method to find a maximum (or a minimum) of a linear objective function over a feasible region given by a convex polytope.

[13] A popular distance-based reconstruction method.

Figure 9.

The binary trees on X = fx1; x2; x3; x4g

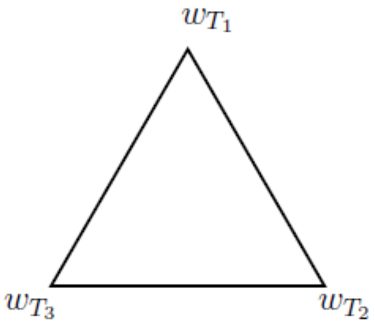


The corresponding vectors $w_{T_i} \in \mathbb{R}^6$ *with coordinates lexicographic order have the form* $(w_{12}, w_{13}, w_{14}, w_{23}, w_{24}, w_{34})$ *and are given by*

$$w_{T_1} = (\frac{1}{2}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{2}), \quad w_{T_1} = (\frac{1}{4}, \frac{1}{2}, \frac{1}{4}, \frac{1}{4}, \frac{1}{2}, \frac{1}{4}), \quad w_{T_1} = (\frac{1}{4}, \frac{1}{4}, \frac{1}{2}, \frac{1}{2}, \frac{1}{4}, \frac{1}{4}).$$

Figure 10.

The BME(4) polytope is given by a triangle in $\mathbb{R}^6$



The BME(n) polytope $\mathcal{P}_n \subseteq \mathbb{R}^{\binom{n}{2}}$ has dimension $\binom{n}{2} - n$, as there are exactly n linear independent[14] equations obeyed by $\mathcal{P}_n$ (Eickmeyer *et al.,* 2008). Furthermore it has $(2n-5)!!$ vertices, which is $|\mathcal{T}_n|$, the cardinality of the set $\mathcal{T}_n$ (see, e.g., (Semple & Steel, 2003)). Some known results are summarized in the following table.

n	$dim(\mathcal{P}_n)$	# of vertices of $\mathcal{P}_n$	# of facets of $\mathcal{P}_n$
3	0	1	0
4	2	3	3
5	5	15	52
6	9	105	90262
n	$\binom{n}{2} - n$	$(2n-5)!!$	?

[14] A set of vectors is linearly independent if none of the vectors in the set can be defined as a linear combination of the others.

It is interesting to note that each NNI-move[15] on $\mathcal{T}_n$ corresponds to an edge of $\mathcal{P}_n$ (Haws *et al.,* 2011). Furthermore, partial results on facet inequalities exist, i.e., e.g. some facets from cherries (Forcey *et al.,* 2016) were characterised.

4. Conclusion and Outlook

We introduced notions from phylogenetics and mathematics that mostly relate to the distance-based approach to phylogenetic reconstruction. The three polytopes are associated to either distances or the number of species. While we focussed on tree-like metrics where tight span and fundamental polytope are well-understood, for more general classes of metrics we still have limited knowledge about their structure. The situation for BME polytopes is similar, where only small examples have complete characterisations. It will be interesting to see in what ways structural properties of the introduced objects relate to each other and to other notions from phylogenetics and mathematics in the future.

References

Bandelt, H.-J., & Dress, A. W. M. (1992). A canonical decomposition theory for metrics on a finite *set. Adv. Math.*, 92(1), 47–105.

Bollobas, B. (1998). *Modern graph theory* (corrected ed.). Heidelberg: Springer.

Delucchi, E., & Hoessly, L. (2020). Fundamental polytopes of metric trees via parallel connections of matroids. *European Journal of Combinatorics*, 87, 103098.

Dress, A. (1984). Trees, tight extensions of metric spaces, and the cohomological dimension of certain groups: A note on combinatorial properties of metric spaces. *Advances in Mathematics*, 53, 321-402.

Dress, A., Huber, K. T., Koolen, J., Moulton, V., & Spillner, A. (2011). *Basic phylogenetic combinatorics*. Cambridge University Press.

Eickmeyer, K., Huggins, P., Pachter, L., & Yoshida, R. (2008). On the optimality of the neighbor-joining algorithm. *Algorithms for Molecular Biology*, 3(1), 5.

Forcey, S., Keefe, L., & Sands, W. (2016). Facets of the balanced minimal evolution polytope. *Journal of Mathematical Biology*, *73*(2), 447–468.

Gascuel, O., & Steel, M. (2006). Neighbor-Joining Revealed. *Molecular Biology and Evolution*, 23(11), 1997-2000. Retrieved from https://doi.org/10.1093/molbev/msl072

Gordon, J., & Petrov, F. (2017). Combinatorics of the Lipschitz polytope. *Arnold Math. J.*, 3(2), 205–218.

Haws, D. C., Hodge, T. L., & Yoshida, R. (2011). Optimality of the neighbor joining algorithm and faces of the balanced minimum evolution polytope. *Bulletin of Mathematical Biology*, 73(11), 2627–2648.

Huson, D. H., Rupp, R., & Scornavacca, C. (2010). *Phylogenetic networks: Concepts, algorithms and applications*. Cambridge University Press.

Isbell, J. R. (1964). Six theorems about injective metric spaces. *Commentarii Mathematici Helvetici*, 39(1), 65–76.

Lefort, V., Desper, R., & Gascuel, O. (2015). FastME 2.0: A Comprehensive, Accurate, and Fast Distance-Based Phylogeny Inference Program. *Molecular Biology and Evolution*, 32(10), 2798-2800.

[15] A nearest-neighbour interchange (NNI) move on a phylogenetic tree rearranges the tree. Such moves are e.g. used in algorithms in order to optimise over the set of trees.

Ostrovska, S., & Ostrovskii, M. (2019). Generalized transportation cost spaces. *arXiv*.

Pauplin, Y. (2000). Direct calculation of a tree length using a distance matrix. *Journal of Molecular Evolution*, 51(1), 41–47.

Semple, C., & Steel, M. (2003). *Phylogenetics*. Oxford University Press.

Steel, M. (2016). Phylogeny: *Discrete and random processes in evolution*. Society for Industrial and Applied Mathematics.

Vershik, A. M. (2015). Classification of finite metric spaces and combinatorics of convex polytopes. *Arnold Math*. J., 1(1), 75–81.

Ziegler, G. M. (1995). *Lectures on polytopes* (Vol. 152). Springer-Verlag, New York.

"Maíz MON 810"
Papel de algodón
30 x 13 x 7 cm
2014