The Dissertation Committee for Yingying Wu
certifies that this is the approved version of the following dissertation:

Comparison Theorems of Phylogenetic Spaces and Algebraic Fans

Committee:

ANDREW J. BLUMBERG, Supervisor

MICHAEL A. MANDELL

SEAN M. KEEL

RACHEL A. WARD

ANDREW M. NEITZKE

DAVID D. BEN-ZVI

Comparison Theorems of Phylogenetic Spaces and Algebraic Fans

by

Yingying Wu

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

August 2019

Dedicated to Shing-Tung Yau,
Clifford H. Taubes,
and Andrew J. Blumberg.

Acknowledgments

I am greatly indebted to my advisor Professor Andrew J. Blumberg for his patience bringing up someone with an evidently weaker mathematical background and lack of mathematical maturity. Mathematics is a formidable journey; I could not possibly have gotten through my graduate studies without the support and nurturing of Professor Blumberg.

I am deeply grateful to my committee, Professor Michael A. Mandell, Professor Sean M. Keel, Professor Rachel A. Ward, Professor Andrew M. Neitzke, and Professor David D. Ben-Zvi, for their generous guidance, kindness, and undeserved encouragement. I owe a huge debt of gratitude for the trust and support of Professor Shing-Tung Yau, Professor Clifford H. Taubes, Professor Karen Uhlenbeck, and Professor Magdalena Czubak.

I would also like to extend my deepest gratitude to Professor Yinbang Lin for countless discussions on projective variety and for his candid opinions on my thesis. I am very grateful to Professor Susan Holmes for discussions on phylogenetic trees and causal relations on mutations. I am very thankful to Professor Satyan L. Devadoss for discussions on phylogenetic networks and inspiration on Christianity. I am very grateful to Professor Bin Chen for discussions on genomics and therapeutics discovery.

I am deeply grateful to Professor Kefeng Liu for many discussions on the moduli space, as well as to Professor Stephen Yau, Professor Dawei Chen, Professor Siu-Cheong Lau, Professor Chiu-Chu Melissa Liu, Professor Mu-Tao Wang, Professor Bong Lian, Professor Shiu-Yuen Cheng, Professor Si Li, Professor Jie Zhou, Professor Zhengwei Liu, Professor Yu-Shen Lin, Professor Yih Sung, Professor Adam J. Jacob, Professor Yi Li (SEU), Professor Shengmao Zhu, and Professor Anand Patel for discussions on algebraic geometry. I am beholden to Professor Lizhen Ji and Professor An Huang for discussions on tropical geometry.

I am deeply grateful to Professor Kui Ren, Professor Richard Yen-Hsi Tsai, and Professor Hao Wu for discussions on numerical analysis. I also owe my gratitude to Professor Andrew C. Yao, Professor Adam Wierman, Professor Yiqun Liu, Professor Shaoping Ma, Professor Min Zhang (Tsinghua), Professor David Xianfeng Gu, Professor Wei Zeng, Professor Jiayu Zhou, Professor Longbo Huang, Professor Yi Li (NTU), Professor Chaohui Wang, Professor Zhiyuan Liu, Professor Shi-Min Hu, Professor Jie Tang, Professor Jian Li, Professor Chenye Wu, and Professor Jialin Zhang for discussions on computer science. I am also grateful to Professor Marie-Amelie Lawn for helping me understand Schröder's paper.

I feel indebted to Professor Kathryn Hess, Professor Jack Morava, Professor Hongchi Shi, Professor Ziliang Zong, Professor Yonghui Wu, Professor Han Li, Professor Horng-Tzer Yau, Professor Zhuande Wang, Professor Xiufen Fang, Professor Peng Yang, Professor Xiaokui Yang, Professor Jeffrey Kuan, and Professor Melody Chan for their love and charity.

I would like to extend my sincere thanks for the mentorship from peers, including from Dr. Laura Fredrickson, Dr. Ning Kang, Dr. Aaron Fenyes, Dr. Timothy Magee, Dr. Yuecheng Zhu, Dr. Valentin Zakharevich, Dr. Allison N. Miller, Dr. Yan Zhou, Dr. Richard Hughes, and Dr. Meng Guo for many discussions on algebraic geometry; Dr. Matthew D. Novack, Dr. Yijing Wu, and Dr. Jiaoyang Huang for discussions on analysis; Dr. Lindsay Martin, Dr. Sean P. Carney, Dr. Yimin Zhong, Dr. Min Zhang (Harvard), Dr. Zeyuan Allen-Zhu, and Dr. Chengu Wang for discussions on computational mathematics; and Dr. Ke Liu, Dr. Jing Xing, and Dr. Junfei Zhao for discussions on bioinformatics.

I owe my deepest gratitude to my parents and grandparents who set me on the path of lifetime learning, encouraged my mathematical pursuits, and provided unfailing support at every moment.

Comparison Theorems of Phylogenetic Spaces and Algebraic Fans

Yingying Wu, Ph.D.
The University of Texas at Austin, 2019

Supervisor: ANDREW J. BLUMBERG

Rapid developments in high-throughput sequencing have accumulated a wealth of cancer genomics data [44, 12], which has led to the use of phylogenetic methods becoming an important direction in cancer research. In-depth understanding of the topological and metric structures of spaces of phylogenetic trees and phylogenetic networks helps to explain the inter- and intra-patient variability of tumor cells. The geometric structures of phylogenetic spaces also allow comparison between subclones of tumor cells to guide therapeutics discovery [11, 38]. Motivated by such research goals, my dissertation proves comparison theorems between phylogenetic spaces that represent evolutionary histories and algebraic fans over simplicial complexes which arise in the moduli space of smooth marked del Pezzo surfaces. I will discuss homeomorphisms and isometries between their projectivized spaces and simplicial complexes formed by root subsystems. Furthermore, embeddings between spaces of phylogenetic trees and networks, and those between the projectivized spaces of phylogenetic trees and networks, are introduced. As a result, algebraic fans that correspond to spaces of phylogenetic trees embed into the fans that correspond to spaces of phylogenetic networks. Likewise, simplicial complexes that correspond to the projectivized spaces of phylogenetic trees embed into the simplicial complexes that correspond to the projectivized spaces of phylogenetic networks. Identifying phylogenetic spaces with structures in mathematics expedites the discovery of the missing pieces in evolutionary models whose counterparts are naturally expected in mathematics, and it also equips investigations in phylogeny with more mathematical tools. The connection between genomic spaces and mathematical structures presented has the potential to address core challenges in oncology with recent advances in algebraic geometry [37], tropical geometry [50, 19, 45], and metric methods [7, 57, 5, 3, 53].

Table of Contents

Chapter 1

Introduction

Phylogenetic analysis studies the evolutionary relationships between different species, organisms, or other taxa to understand the evolution of life [10]. Darwin proposed phylogenetic trees in 1859 to represent the evolutionary relationships of reproducing individuals [13]; over a century later, Grant proposed phylogenetic networks in 1971 [24] to represent non-tree-like relationships such as species hybridization, bacterial gene transfer, and homologous recombination. Cancer is a genetic disease that progresses under Darwinian evolution via a process of diversification and selection for mutations that promote tumor cell proliferation and survival [41]. Evolutionary trees reconstruct the evolution of tumors, and they are also called "cancer evolutionary trees" [39] or "oncogenetic trees" [48]. Characterizing inter-tumor heterogeneity according to the patterns of evolutionary trees is an important strategy for developing targeted therapies and identifying metastatic potential, as well as preventing the emergence of drug resistance. The task is challenging due to genetic heterogeneity, complicated by the clonal evolutionary patterns, and lack of robust identification across patients.

Multi-region sequencing allows the temporal order of some genomic changes in a tumor to be inferred. Several reconstruction methods have been developed to resolve intratumor heterogeneity, including nearest-neighbor interchanging in 1978 [55], Robinson-Foulds distance in 1981 [46], quartet distance in 1985 [21], tree edit distance in 1989 [60], path length metrics in 1993 [52], branch length scores in 1994 [36], subtree transfer distance in 2001 [2], and Billera-Holmes-Vogtmann distance in 2001 [5], which is the central object of my study. Many studies have investigated the characterization of genomic data using tools from geometry and topology to study the evolution of viruses and cancer [58, 6, 20]. Speyer and Sturmfels proved that the tropical Grassmannian is the space of phylogenetic trees [50]; Sen and Dukkipati showed that the tropical linear spaces correspond to the phylogenetic trees

[19]. Devadoss and Morava formulated a smooth blowup to BHV_n from the compactified moduli stacks of stable genus 0 curves with n marked points $\mathcal{M}_{0,n}(\mathbb{R})$ [15]. Baez and Otter constructed an operad whose operations are the edge-labelled phylogenetic trees [4]. Therefore, studies on the tropicalization of classical moduli spaces [1, 45], projective variety [37], and category theory become meaningful tools for the studies of phylogenetic spaces.

In the following, I will introduce comparison theorems between phylogenetic spaces that represent evolutionary histories and algebraic fans over simplicial complexes which arise in the moduli space of smooth marked del Pezzo surfaces. I will discuss morphisms between the projectivized spaces and simplicial complexes formed by root subsystems. Furthermore, I will prove embeddings between spaces of phylogenetic trees and networks, and those between the projectivized spaces of phylogenetic trees and networks. As a result, algebraic fans that correspond to spaces of phylogenetic trees embed into the fans that correspond to spaces of phylogenetic networks, and simplicial complexes that correspond to the projectivized spaces of phylogenetic trees embed into the simplicial complexes that correspond to the projectivized spaces of phylogenetic networks. Identifying phylogenetic spaces with structures in mathematics equips investigations in phylogeny with more mathematical tools from algebraic geometry and tropical geometry [50, 19]. The homeomorphisms and isometries between spaces parameterizing genomic data and non-positively curved cubical complexes provide a conceptual way to understand spaces of phylogenetics in terms of simplicial fans formed by the root systems.

1.1 Topological Structures of Spaces of Phylogenetic Trees

We first discuss theorems that describe topological structures that are related to spaces of phylogenetic trees. According to the definition of Zairis-Khiabanian-Blumberg-Rabadan, a *cubical complex* is a metric space obtained by gluing together cubes via the data of isometries of faces, subject to the condition that two cubes are connected by at most a single face identification and no cube is glued to itself. The distance between x and y is the infimum of the lengths over all paths from x to y that can be expressed as the union of finitely many segments each contained within a cube [59]. And the *link of a vertex v* in a cubical complex C is obtained as the subset

$$\mathcal{L}(v) = \{z \in C \mid d(z, v) = \epsilon\}$$

for fixed $0 < \epsilon < 1$ [59]. Then we will show that:

Proposition 2.1.11: The link $\mathcal{L}(v)$ of a vertex v in a cubical complex C is a spherical complex.

We denote *root subsystems* $\Theta_n^{A_k|B_k}$ of D_n as below:

$$\Theta_n^{A_k|B_k} = \{\pm\epsilon_i \pm \epsilon_j \mid i \neq j; i, j \in A_k\},\ A_k \subset \{1, \dots, n\},\ \text{and } B_k = \{1, \dots, n\} \setminus A_k.$$

We define a simplicial complex $\mathcal{R}(\Delta_n)$ according to Hacking, Keel and Tevelev [25]. Let $\mathcal{D}_{k,2m}$ and $\mathcal{D}_{k,2m+1}$ be sets of root subsystems $\Theta_n^{A_i|B_i}$, such that
for $k = 1, \dots, m$:

$$\mathcal{D}_{k,2m+1} = \left\{\Theta_{2m+1}^{A_1|B_1}, \dots, \Theta_{2m+1}^{A_L|B_L}\right\},\ \text{where } L = \binom{2m+1}{k};$$

for $k = 1, \dots, m-1$:

$$\mathcal{D}_{k,2m} = \left\{\Theta_{2m}^{A_1|B_1}, \dots, \Theta_{2m}^{A_L|B_L}\right\},\ \text{where } L = \binom{2m}{k};$$

for $k = m$:

$$\mathcal{D}_{m,2m} = \left\{\Theta_{2m}^{A_1|B_1}, \dots, \Theta_{2m}^{A_L|B_L}\right\},\ \text{where } L = \binom{2m}{m}\Big/ 2.$$

We define the set $\mathcal{R}(D_{2m+1})$ and $\mathcal{R}(D_{2m})$ as below:

$$\mathcal{R}(D_{2m+1}) = \mathcal{D}_{2,2m+1} \sqcup \mathcal{D}_{3,2m+1} \sqcup \cdots \sqcup \mathcal{D}_{m,2m+1};$$
$$\mathcal{R}(D_{2m}) = \mathcal{D}_{2,2m} \sqcup \mathcal{D}_{3,2m} \sqcup \cdots \sqcup \mathcal{D}_{m-1,2m} \sqcup (\mathcal{D}_{m,2m} \times \mathcal{D}_{m,2m}).$$

Therefore, $\mathcal{R}(D_n)$ for $n = 2m$ or $n = 2m + 1$ is a set of root subsystems with elements from the sets of root subsystems $\mathcal{D}_{k,n}$ for k from 2 to m. Define the simplicial complex $\mathcal{R}(\Delta_n)$ with 0-simplices being elements in the set $\mathcal{R}(D_n)$. Higher-dimensional simplices $\Theta_n \subset \mathcal{R}(\Delta_n)$ are formed by elements in $\mathcal{R}(D_n)$, such that for each pair of $\Theta_n^{A_i|B_i}, \Theta_n^{A_j|B_j} \in \Theta_n$, one of the following criteria is fulfilled:

$$\Theta_n^{A_i|B_i} \perp \Theta_n^{A_j|B_j},\quad \Theta_n^{A_i|B_i} \subset \Theta_n^{A_j|B_j},\quad \text{or } \Theta_n^{A_j|B_j} \subset \Theta_n^{A_i|B_i}.$$

We define a strongly convex rational polyhedral cone $\tau_{\mathcal{F}}$ as a set:

$$\tau_{\mathcal{F}} = \left\{ r_1 \psi \left(\Theta_n^{A_1|B_1} \right) + \cdots + r_n \psi \left(\Theta_n^{A_k|B_k} \right) \in N(D_n)_{\mathbb{Q}} : r_i \geq 0 \right\}.$$

The fan $\mathcal{F}(D_n)$ in $N(D_n)_{\mathbb{Q}}$ is a collection of strongly convex rational polyhedral cones $\tau_{\mathcal{F}}$ in the real vector space $N(D_n)_{\mathbb{Q}}$ determined by the rays $\psi \left(\Theta_n^{A_i|B_i} \right)$ for each $\Theta_n^{A_i|B_i} \in \mathcal{R}(\Delta_n)$, where rays span a cone in $\mathcal{F}(D_n)$ if and only if the corresponding subsystems form a simplex in $\mathcal{R}(\Delta_n)$. Then we will prove:

Proposition 2.4.7: $\mathcal{F}(D_n)$ is a fan.

Lemma 2.4.10: $\mathcal{F}(D_n)$ is a cubical complex.

Then we will prove theorems that compare different spaces related to phylogenetic trees and their metric structures. We first define the Billera-Holmes-Vogtmann space of phylogenetic trees. An illustration of BHV_3 and $\mathbb{P}\,\mathrm{BHV}_3$ can be found in Fig. 1.1a, and an illustration of BHV_4 and $\mathbb{P}\,\mathrm{BHV}_4$ can be found in Fig. 2.1b.

Definition 1.1.1 (Billera-Holmes-Vogtmann space of phylogenetic trees)**.** The *Billera-Holmes-Vogtmann space of phylogenetic trees* is the space of isometry classes of rooted phylogenetic trees with m-labeled leaves, where the nonzero weights are on the internal branches, denoted as BHV_n. The space BHV_n is constructed by gluing together $(2n-3)!!$ positive orthants; each orthant corresponds to a particular tree topology, with the coordinates specifying the lengths of the edges. A point T_n in the interior of an orthant in BHV_n represents a binary tree. If any of the coordinates are 0, the tree is obtained by collapsing edges corresponding to the 0 coordinates from a binary tree. We glue orthants together, such that a tree is on the boundary between two orthants when it can be obtained by collapsing edges from either tree topology. The BHV_n space is a cubical complex built by tiling each orthant with unit cubes of dimension $n-2$ [59].

Definition 1.1.2 (Projective Billera-Holmes-Vogtmann space of phylogenetic trees)**.** The *projective Billera-Holmes-Vogtmann space of phylogenetic trees* $\mathbb{P}\,\mathrm{BHV}_n$ is the subspace of BHV_n consisting of points in each orthant, for which the sum of its $n-2$ internal edges t_i is 1. $\mathbb{P}\,\mathrm{BHV}_n$ inherits a simplicial structure from BHV_n, where the k-simplices of $\mathbb{P}\,\mathrm{BHV}_n$ are points projected from points in BHV_n with exactly $k+1$ nonzero edges, and the intersection of faces is determined by degenerated trees that share the same internal edges.

Then we will show that:

Proposition 2.1.12: There is an isomorphism of sets between $\mathbb{P}\,\mathrm{BHV}_n$ and the link of BHV_n at the origin defined by Zairis-Khiabanian-Blumberg-Rabadan.

Proposition 2.1.15: The link of BHV_n at the origin equipped with the Euclidean metric on each simplex is homeomorphic to the space equipped with the spherical metric.

Proposition 2.1.16: $\mathbb{P}\,\mathrm{BHV}_n$ is the projectivized space of BHV_n, and the cone over $\mathbb{P}\,\mathrm{BHV}_n$ is isomorphic to BHV_n as sets.

Proposition 2.1.21: The 0-cone over $\mathbb{P}\,\mathrm{BHV}_n$ is isometric to BHV_n.

Theorem 2.3.7: $\mathcal{R}(\Delta_n)$ and $\mathcal{P}(D_n)$ are duals: k-simplices of $\mathcal{R}(\Delta_n)$ are in bijection with $(n-k-4)$-dimensional faces of $\mathcal{P}(D_n)$; a k-simplex of $\mathcal{R}(\Delta_n)$ consists of a collection of root subsystems Θ forms a $(k+1)$-simplex of $\mathcal{R}(\Delta_n)$ with some root subsystem $\Theta_n^{A|B}$ if and only if the $(n-k-4)$-dimensional faces of $\mathcal{P}(D_n)$ that corresponds to Θ form an $(n-k-5)$-dimensional face of $\mathcal{P}(D_n)$ with a facet that corresponds to $\Theta_n^{A|B}$.

1.2 Comparison Theorems on Spaces of Phylogenetic Trees and Algebraic Fans

Then we compare spaces of phylogenetic trees that correspond to simplicial complexes and algebraic fans that arise in algebraic geometry:

Theorem 2.3.1: $\mathbb{P}\,\mathrm{BHV}_{n-1}$ and $\mathcal{R}(\Delta_n)$ are isomorphic abstract simplicial complexes.

Corollary 2.3.4: The geometric realization of $\mathbb{P}\,\mathrm{BHV}_{n-1}$ and $\mathcal{R}(\Delta_n)$ are homeomorphic.

Theorem 2.4.8: BHV_{n-1} and $\mathcal{F}(D_n)$ are isomorphic.

Main Theorem 2.4.11: BHV_{n-1} and $\mathcal{F}(D_n)$ are isometric with the cubical complex metric.

1.3 Topological Structures of Spaces of Phylogenetic Networks

We perform analogous analysis for the space of phylogenetic networks. We prove theorems on topological structures of spaces that are related to spaces of phylogenetic networks. We define $\mathcal{G}(D_n)$ as the convex hull of the collection of rays from $\mathcal{F}(D_n)$; these rays form a cone if and only if the corresponding boundary divisors all meet a single closed cell. We will show:

Proposition 3.2.3: $\mathcal{G}(D_n)$ is a fan.

Lemma 3.4.1: $\mathcal{G}(D_n)$ is a cubical complex.

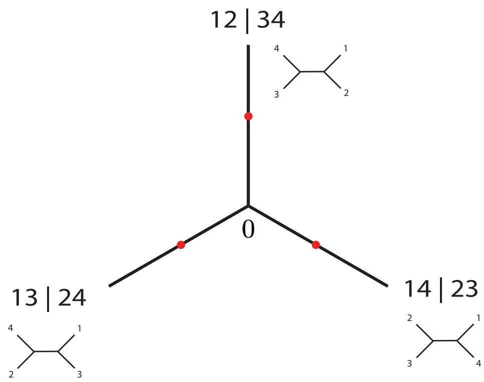
A *split network* $N = (V, E, \sigma, \lambda)$ that represents a set of splits $\mathcal{S}$ on $\mathcal{X}$ is given by a split graph $(G = (V, E), \sigma : E \to \mathcal{S})$ and a node labeling $\lambda : \mathcal{X} \to V$, such that

$$A = \bigcup_{v \in V_S^0} \lambda^{-1}(v) \text{ and } B = \bigcup_{v \in V_S^1} \lambda^{-1}(v)$$

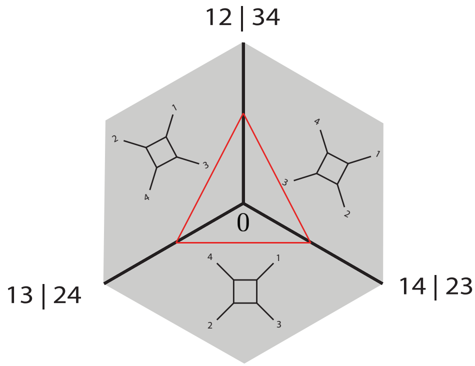
for every split $S = A \mid B$ in $\mathcal{S}$, where V_S^0 and V_S^1 refer to the sets of vertices in the connected component G_c^0 and G_c^1, respectively [29]. Phylogenetic network refers to any graph used to represent evolutionary relationships between a set of taxa [30].

Definition 1.3.1 (The space of circular split networks)**.** The *space of circular split networks* is the space of isometry classes of split networks with n-labeled leaves, where the nonzero weights are on the internal branches sharing the same circular ordering, denoted as CSN_n. The space CSN_n is constructed by gluing together $\frac{(n-1)!}{2}$ positive orthants; each orthant corresponds to a particular split network topology, with the coordinates specifying the lengths of the edges. A point C_n in the interior of an orthant in CSN_n represents a split network with all positive edge lengths. If any of the coordinates are 0, the split network is obtained by collapsing edges corresponding to the 0 coordinates from a split network with all positive length edges. We glue orthants together, such that a split network topology is on the boundary between two orthants when it can be obtained by collapsing edges from either network topology. The CSN_n space is a cubical complex built by tiling each orthant with unit cubes of dimension $\frac{n(n-3)}{2}$.

Definition 1.3.2 (Projective space of circular split networks)**.** The *projective space of circular split networks* $\mathbb{P}\,\mathrm{CSN}_n$ is the subspace of CSN_n consisting of points in each orthant, for which the sum of its $\frac{n(n-3)}{2}$ internal edges t_i is 1. $\mathbb{P}\,\mathrm{CSN}_n$ inherits a simplicial structure from CSN_n, where the k-simplices of $\mathbb{P}\,\mathrm{CSN}_n$ are points projected from points in CSN_n with exactly $k+1$ nonzero edges, and the intersection of faces is determined by collapsed edges.



(a) BHV_3 and $\mathbb{P}\mathrm{BHV}_3$.



(b) CSN_4 and $\mathbb{P}\mathrm{CSN}_4$.

Figure 1.1: (a) BHV_3: Each ray is a 1-dimensional orthant in BHV_3. **Dots**: $\mathbb{P}\,\mathrm{BHV}_3$. (b) CSN_4: Each ray represents a split; each orthant corresponds to a distinctive network topology. **Triangle**: $\mathbb{P}\,\mathrm{CSN}_4$.

In Fig. 1.1a, each ray is a 1-dimensional orthant in BHV_3, and each orthant represents a different tree topology, which corresponds to a distinctive non-trivial split. $\mathbb{P}\,\mathrm{BHV}_3$ is highlighted as red dots, each dot is of distance 1 from the origin. In Fig. 1.1b, each ray can be identified with a ray in BHV_3. Each orthat in CSN_4 is 2-dimensional, spanned by two rays that share the same circular ordering. $\mathbb{P}\,\mathrm{CSN}_4$ is represented as the triangle in the center of Fig. 1.1b; for each point on $\mathbb{P}\,\mathrm{CSN}_4$, their coordinates sum up to 1. We will show:

Proposition 3.1.10: There is an isomorphism of sets between $\mathbb{P}\,\mathrm{CSN}_n$ and the link of CSN_n at the origin defined by Zairis-Khiabanian-Blumberg-Rabadan.

Theorem 3.1.11: $\mathbb{P}\,\mathrm{CSN}_n$ is the projectivized space of CSN_n, and the cone over $\mathbb{P}\,\mathrm{CSN}_n$ is isomorphic to CSN_n as sets.

1.4 Embeddings from Spaces of Phylogenetic Trees to Networks

We will consider how spaces of phylogenetic trees are connected to spaces of phylogenetic networks:

Theorem 3.1.9: BHV_{n-1} is embedded into CSN_n. There are C_{n-2} types of tree topology in BHV_{n-1} embedded in each orthant of CSN_n, where C_{n-2} is the Catalan number. Each tree topology in BHV_{n-1} is embedded in the intersection of 2^{n-3} orthants of CSN_n.

Corollary 3.1.12: $\mathbb{P}\,\mathrm{BHV}_{n-1}$ is embedded into $\mathbb{P}\,\mathrm{CSN}_n$. There are C_{n-2} types of tree topology in $\mathbb{P}\,\mathrm{BHV}_{n-1}$ embedded in each orthant of $\mathbb{P}\,\mathrm{CSN}_n$. Each type of tree topology in $\mathbb{P}\,\mathrm{BHV}_{n-1}$ is embedded in the intersection of 2^{n-3} orthants of $\mathbb{P}\,\mathrm{CSN}_n$.

Theorem 3.2.4: The fan $\mathcal{F}(D_n) \subset N(D_n)_{\mathbb{Q}}$ embeds into the *remarkable fan* $\mathcal{G}(D_n)$.

Corollary 3.3.3: $\mathcal{R}(\Delta_n)$ is embedded into $\mathcal{Q}(\Delta_n)$. There are C_{n-2} maximal dimension simplices in $\mathcal{R}(\Delta_n)$ embedded in each maximal dimension simplex of $\mathcal{Q}(\Delta_n)$. Each maximal dimension simplex in $\mathcal{R}(\Delta_n)$ is embedded in the intersection of 2^{n-3} maximal dimension simplices of $\mathcal{Q}(\Delta_n)$.

1.5 Comparison Theorems on Spaces of Phylogenetic Networks and Algebraic Fans

Finally, we compare spaces arising from phylogenetic networks that correspond to simplicial complexes and algebraic fans arising in algebraic geometry:

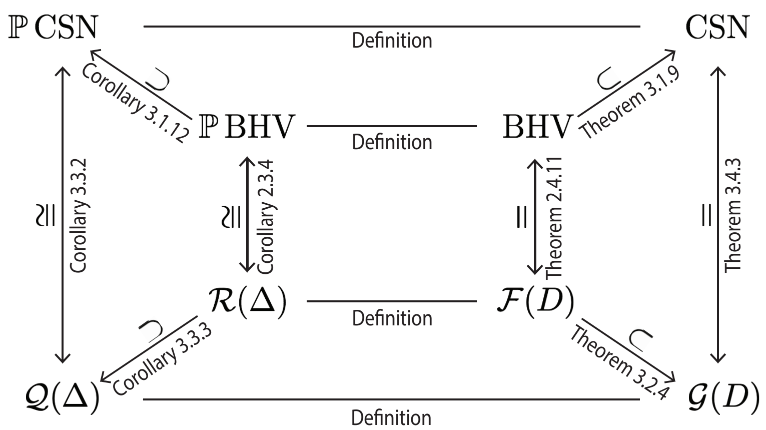
Theorem 3.3.1: $\mathbb{P}\,\mathrm{CSN}_n$ and $\mathcal{Q}(D_n)$ are isomorphic abstract simplicial complexes.

Corollary 3.3.2: The geometric realization of $\mathbb{P}\,\mathrm{CSN}_n$ and $\mathcal{Q}(\Delta_n)$ are homeomorphic.

Theorem 3.4.2: CSN_n and $\mathcal{G}(D_n)$ are isomorphic.

Main Theorem 3.4.3: CSN_n and $\mathcal{G}(D_n)$ are isometric cubical complexes.

We summarize comparison theorems of phylogenetic spaces and algebraic fans in Fig. 1.2. The relations we discovered are highly symmetrical: Spaces related to phylogenetic trees are located on the internal cycle in the chart, and spaces related to phylogenetic networks are located on the outer cycle. The upper half of the diagram refers to spaces that directly arise in genetics, and the lower-half refers to spaces that arise in algebraic geometry. Meanwhile, spaces on the right half of the chart can be generated by their counterparts on the left, i.e, BHV and CSN are the 0-cone over their projectivized spaces $\mathbb{P}\,\mathrm{CSN}$ and $\mathbb{P}\,\mathrm{BHV}$, and $\mathcal{F}(\Delta)$ and $\mathcal{G}(\Delta)$ are the algebraic cone over the simplicial complexes $\mathcal{R}(\Delta)$ and $\mathcal{Q}(\Delta)$. The diagonals represent embeddings of spaces of phylogenetic trees and their counterparts into spaces of phylogenetic networks. These highly symmetrical relations among these spaces reveal the intrinsic connection between genetic spaces and topological spaces, and we will discuss these spaces in greater detail.



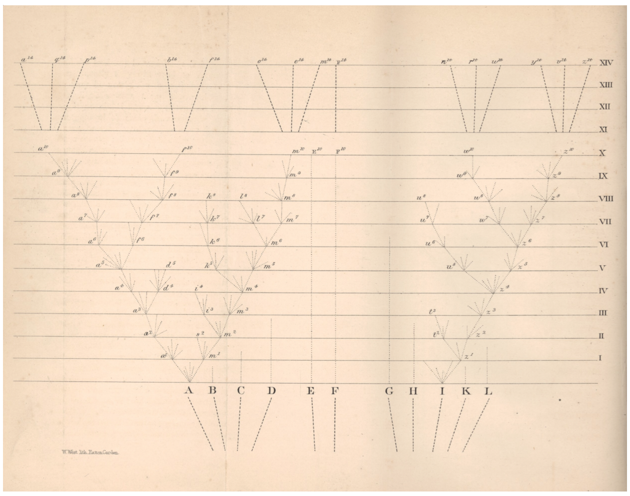
$\simeq$: isomorphism $\cong$: Homeomorphism $=$: Isometry $\subset$: inclusion

Figure 1.2: Comparison theorems of phylogenetic spaces and algebraic fans.

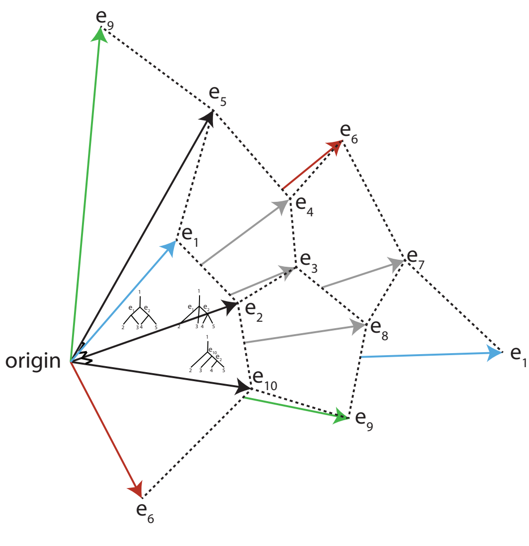
Chapter 2

Comparison Theorems on Spaces of Phylogenetic Trees

Evolution occurs in populations of reproducing individuals, and phylogenetic analysis studies the evolutionary relationships between different species, organisms, or other taxa to understand the evolution of life [10, 29]. In 1859, Darwin proposed the paradigm to describe the modified descendants of the diversified structure of a species [13], as shown in Fig. 2.1a, later formulated as phylogenetic trees. In the diagram, A to L represent the species of a large genus, while resembling each other in unequal degrees represented in the diagram by the letters standing at unequal distances. Let (A) represent a common, widely-diffused, and varying species belonging to a large genus. The little fan of diverging dotted lines of unequal lengths proceeding from (A) represents its varying offspring, where the variation is slight but diversified, consecutive but following long intervals of time [13].



(a) Phylogenetic trees, Darwin 1859



(b) Space of phylogenetic trees BHV_4

Figure 2.1: Paradigm to represent evolutionary relationships and the space it resides in.

2.1 Spaces of Phylogenetic Trees and Projectivized Spaces

This section introduces the spaces of phylogenetic trees formulated by Billera, Holmes, and Vogtmann. Fig. 1.1a illustrates BHV_3, the space of phylogenetic trees of dimension 3, with the dots depicting its projective space $\mathbb{P}\,\mathrm{BHV}_3$. Fig. 2.1b is the space of phylogenetic trees of dimension 4, with the pair of e_1 vectors identified, the pair of e_6 vectors identified, and the pair of e_9 vectors identified. The dotted lines represent the projective space of phylogenetic trees $\mathbb{P}\,\mathrm{BHV}_4$. Denote by $I^n = [0,1] \times \cdots \times [0,1] \subset \mathbb{R}^n$ the n-dimensional unit cube with induced metric. An elementary interval is a subset $I \subsetneq \mathbb{R}$ of the form

$$I = [\ell, \ell + 1] \quad \text{or} \quad I = [\ell, \ell]$$

for some $\ell \in \mathbb{Z}$. An elementary cube Q is the finite product of elementary intervals, i.e.,

$$Q = I_1 \times I_2 \times \cdots \times I_d \subsetneq \mathbb{R}^d$$

where $I_1, I_2, \ldots, I_d$ are elementary intervals. A codimension k face of the cube I^n is determined by fixing k coordinates to be in the set $\{0, 1\}$. In the following, we present a definition of a cubical complex from standard CAT(0) theory [8].

Definition 2.1.1 (Cubical complex)**.** A *cubical complex* K is the quotient of a disjoint union of cubes $X = \coprod_\Lambda I^{n_\lambda}$ by an equivalence relation $\sim$. The restrictions $p_\lambda : I^{n_\lambda} \to K$ of the natural projection $p : X \to K = X/\sim$ are required to satisfy:

1. for every $\lambda \in \Lambda$, the map p_λ is injective;

2. if $p_\lambda(I^{n_\lambda}) \cap p_{\lambda'}(I^{n_{\lambda'}}) \neq \emptyset$ then there is an isometry $h_{\lambda,\lambda'}$ from a face $T_\lambda \subset I^{n_\lambda}$ onto a face $T_{\lambda'} \subset I^{n_{\lambda'}}$, such that $p_\lambda(x) = p_{\lambda'}(x')$ if and only if $x' = h_{\lambda,\lambda'}(x)$ [8].

Definition 2.1.2 (Phylogenetic tree)**.** A *phylogenetic tree* with n leaves is the equivalence class of a weighted, connected graph with no cycles, having n distinguished labeled vertices up to rotations and a root. The labeled vertices and the root are of degree 1 and are called *leaves.* All the other vertices are of degree ≥ 3 [5].

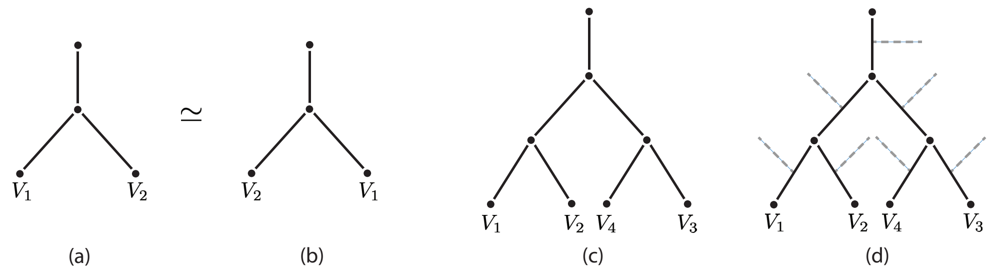


Figure 2.2: $(2n-3)!!$ kinds of trees with n labels.

Schröder proved that the number of representations of parentheses on n items while changing the order within parentheses are considered the same is $(2n-3)!!$ [47]. Such representations can be identified with types of rooted phylogenetic trees with n-labeled leaves of degree 1, while all other vertices are of degree 3. In the following, we present an inductive proof.

Proposition 2.1.3 (Number of full-fledged trees)**.** There are $(2n-3)!! = (2n-3)\cdot(2n-5)\cdot\cdots\cdot 3\cdot 1$ types of rooted phylogenetic trees with n-labeled leaves of degree 1; all other vertices are of degree 3. Such trees reside in the interior of an orthant in BHV_n, and we call them full-fledged rooted phylogenetic trees.

Proof. We proceed the proof by induction. Since rooted phylogenetic trees are the equivalence classes up to rotation, there is only one kind of rooted phylogenetic tree with two labels when $n = 2$, depicted in Fig. 2.2(a) and (b). For the inductive hypothesis, we assume that there are $(2n-3)!!$ types of full-fledged rooted phylogenetic trees, as illustrated in Fig. 2.2(c). For the inductive step, consider adding the $(n+1)$-th leaf, based on the tree with n existing leaves. There are $n-2$ internal edges, n leaves, and one root, totaling $2n-1$ ways to add a new leaf. Therefore, there are $(2n-3)!!(2n-1) = (2(n+1)-3)!!$ types of full-fledged rooted phylogenetic trees, illustrated as Fig. 2.2(d). □

Definition 2.1.4 (Split)**.** Consider any bipartition of a set of taxa $\mathcal{X}$ labeling the leaves of a phylogenetic tree into two non-empty subsets A and B, such that $A\cap B = \emptyset$ and $A\cup B = \mathcal{X}$. If each subset has at least two elements, the partition is called a *split,* denoted by $S_n^{A|B}$. Otherwise, it is called a trivial split.

Definition 2.1.5 (Circular ordering)**.** We call such an ordering $\{x_1, \ldots, x_n\}$ a *circular ordering* for $\mathcal{S}_n$, as it holds that $\{x_1, \ldots, x_n\}$ is a circular ordering for $\mathcal{S}_n$ if and only if both $\{x_n, x_{n-1}, \ldots, x_1\}$ and $\{x_2, x_3, \ldots, x_n, x_1\}$ are circular orderings.

Removing an internal edge of a phylogenetic tree gives two connected components, each with a set of leaves; therefore, each split in a phylogenetic tree corresponds to an internal edge. We use $\mathfrak{S}_n$ to denote the set of splits in BHV_n. $|\mathfrak{S}_n| = 2^n - n - 1$ because there are a total of 2^n choices over n taxa, minus n trivial splits and one star, i.e., a degenerated phylogenetic tree with no internal edge.

Definition 2.1.6 (Compatible)**.** Two partitions $A|B$ and $A'|B'$ are defined to be *compatible* if one of the subsets

$$A \cap A',\ A \cap B',\ A' \cap B,\ B \cap B'$$

is empty.

Different edges of the same tree correspond to compatible partitions, and each binary tree in BHV_n corresponds to the maximal dimensional orthant, which consists of $n-2$ splits; therefore, we represent each tree T_n of a certain orthant in BHV_n by a coordinate system

$$\mathcal{S}_n^{n-2} = \left\{S_n^{A_1|B_1}, \ldots, S_n^{A_{n-2}|B_{n-2}}\right\}$$

for a set of compatible splits $\mathcal{S}_n^{n-2} \subset \mathfrak{S}_n$, and each phylogenetic tree in a specific orthant can be parametrized by

$$T_n = \left(t_1 S_n^{A_1|B_1}, \ldots, t_{n-2} S_n^{A_{n-2}|B_{n-2}}\right)$$

for $t_i \in \mathbb{R}_{\geq 0}$.

Definition 2.1.7 (BHV metric)**.** The *distance* between any two points of BHV_n in the same orthant is the Euclidean distance. If two points are in different orthants, we join them by a sequence of straight segments, with each segment lying in a single orthant. The distance between the two points is the minimum of the lengths of such segmented paths joining the two points [5], denoted as $d_{\text{BHV}}(x, y)$.

The metric structure of the BHV space remains an active area of inquiry. Ardila-Owen-Sullivant [3] found that BHV is an orthant space, which is a special case of cubical complexes, such that any set of orthants is arranged around a common origin. Therefore, BHV is equipped with a cubical complex metric, and the geodesics in BHV space can be computed in polynomial time [42, 43]. Holmes studied the estimation and validation of phylogenetic trees in statistical terms [27], the statistical testing of phylogenies [28], and the evaluation of robustness and performing statistical inferences [26]. Zairis-Khiabanian-Blumberg-Rabadan studied the evolutionary moduli space and introduced the projective tree space [58]. In evolutionary applications, rescaling edge lengths should not change the relationship between the branches [58]; the definition of the *projective Billera-Holmes-Vogtmann space of phylogenetic trees* is motivated by this consideration.

Definition 2.1.8 (Spherical simplex)**.** Fix a vertex v in a cubical complex C and a cube $C_i \simeq I^m \subset C$, such that v is a vertex of C_i. For fixed $\epsilon > 0$, the *spherical simplex* associated with (C_i, v) is the subset

$$S(C_i, v) = \{z \in C_i \mid d(z, v) = \epsilon\}.$$

The set $S(C_i, v)$ has a metric induced by the Euclidean angle metric. The faces of $S(C_i, v)$ are defined as the intersection of $S(C_i, v)$ with faces of C_i [59].

These are the all right spherical simplices associated with faces of C_i. The term *all right* indicates the fact that the edges have the length $\epsilon\pi/2$ [8]. Without the metric, the all right spherical simplices form an abstract simplicial complex. Billera-Holmes-Vogtmann viewed the link of the origin as the set of points, such that the sum of the coordinates is equal to one, which is $\mathbb{P}\,\mathrm{BHV}$ in Definition 1.1.2. Zairis-Khiabanian-Blumberg-Rabadan defined the link of vertex as a subset of points with equal distance ϵ for some $0 < \epsilon < 1$:

Definition 2.1.9 (Zairis-Khiabanian-Blumberg-Rabadan link of vertex)**.** The *link of a vertex* v in a cubical complex C is obtained as the subset

$$\mathcal{L}_C(v) = \{z \in C \mid d(z, v) = \epsilon\}$$

for fixed $0 < \epsilon < 1$ [59].

Therefore, the link of BHV_n at the origin is the set

$$\mathcal{L}_{\mathrm{BHV}}(0) = \{z \in \mathrm{BHV}_n \mid d(z, 0) = \epsilon\}$$

for some fixed $0 < \epsilon < 1$.

Definition 2.1.10 (Spherical complex)**.** A *spherical complex* K in $\mathbb{R}^N$ is a collection of spherical simplices in $\mathbb{R}^N$, such that:

1. Every face of a spherical simplex of K is in K.
2. The intersection of any two spherical simplices of K is a face of each of them.

Proposition 2.1.11 (Link of a vertex is a spherical complex)**.** *The link $\mathcal{L}(v)$ of a vertex v in a cubical complex C is a spherical complex.*

Proof. By definition, for a cube $C_i \simeq I^m$ in a cubical complex C, the faces of a spherical simplex $S(C_i, v)$, where v is a vertex of C_i, are defined as the intersection of $S(C_i, v)$ with faces of C_i. Therefore, for fixed $\epsilon > 0$, the intersection of $S(C_i, v)$ with faces of C_i is the subset

$$S(\partial C_i, v) = \{z \in \partial C_i \mid d(z, v) = \epsilon\}.$$

However, $\partial C_i \simeq I^{m'} \subset C$; therefore, $S(\partial C_i, v)$ is also a spherical simplex. By the definition of the link of a vertex v, the intersection of two spherical simplices is:

$$\begin{aligned} S(C_i, v) \cap S(C_i', v) &= \{z \in C_i \mid d(z, v) = \epsilon\} \cap \{z \in C_i' \mid d(z, v) = \epsilon\} \\ &= \{z \in C_i \cap C_i' \mid d(z, v) = \epsilon\}. \end{aligned}$$

Since $C_i \cap C_i' \simeq I^m \subset C$, $S(C_i, v) \cap S'(C_i', v)$ is also a spherical simplex. Hence, the intersection of any two spherical simplices of K is a face of each of them. Therefore, the link $\mathcal{L}(v)$ of a vertex v is a spherical complex. $\square$

Thus, the link of the origin in the cubical complex BHV_n is a spherical complex. In the following, we will show it is bijective to $\mathbb{P}\,\mathrm{BHV}_n$.

Proposition 2.1.12 ($\mathbb{P}\,\mathrm{BHV}_n$ and link of BHV_n isomorphism)**.** There is an isomorphism of sets between $\mathbb{P}\,\mathrm{BHV}_n$ and the link of BHV_n at the origin defined by Zairis-Khiabanian-Blumberg-Rabadan.

Proof. According to the definition of Zairis-Khiabanian-Blumberg-Rabadan, the link of the origin in BHV_n is the set of points

$$T_n = \left(t_1 S_n^{A_1|B_1}, \dots, t_{n-2} S_n^{A_{n-2}|B_{n-2}}\right),$$

where $t_i \geq 0$ and

$$\mathcal{S}_n^{n-2} = \left\{S_n^{A_1|B_1}, \dots, S_n^{A_{n-2}|B_{n-2}}\right\}$$

is a set of compatible splits $\mathcal{S}_n^{n-2} \subset \mathfrak{S}_n$, such that $\sum_{i=1}^{n-2} t_i^2 = \epsilon^2$. Meanwhile, by Definition 1.1.2, $\mathbb{P}\,\mathrm{BHV}_n$ consists of the points, such that the sum of internal edges t_i satisfies $\sum_{i=1}^{n-2} t_i = 1$. Both of these sets are parameterizable using angular coordinates within each orthant; therefore, we define a map:

$$\rho : \mathbb{P}\,\mathrm{BHV}_n \longrightarrow \mathcal{L}_{\mathrm{BHV}}(0),$$
$$(l, \varphi_1, \dots, \varphi_{n-3}) \mapsto (\epsilon, \varphi_1, \dots, \varphi_{n-3}),$$

where φ_ks are angular coordinates and ϵ is determined by the relation that the sum of the squares of the coordinates is equal to one. Since ρ is a bijection, $\mathcal{L}_{\mathrm{BHV}}(0)$ and $\mathbb{P}\,\mathrm{BHV}_n$ are isomorphic. $\square$

Definition 2.1.13 (Metric on the link of origin)**.** Each tree $T_n \in \mathrm{BHV}_n$ lies on a ray from the origin. The intersection of this ray with the link of origin is called the projection of T_n onto the link, and is denoted $t(T_n)$. The angle between T_n and T_n', denoted $\angle(T_n, T_n')$, is the distance between $t(T_n)$ and $t(T_n')$ in the spherical metric $\|\cdot\|_{\mathbb{S}}$ on the link.

This new metric on simplices extends to a natural metric on the entire link, in which each simplex is a right-angled spherical simplex with all edges of length $\frac{\pi}{2}$, and we piece together these metrics on each spherical simplex by defining paths and taking a length metric as follows:

Definition 2.1.14 (The l^1-path metric of a simplicial complex)**.** Let X be a simplicial complex. For any two points $x, y \in X$, we define a path in X from x to y to be a sequence of points $x = a_0, a_1, \dots, a_r = y$ in X, such that for every $i = 1, \dots, r$ there is a simplex σ_i of X that contains both a_{i-1} and a_i. The length of such a path is the sum

$$\sum_{i=1}^{r} d_{\sigma_i}(a_{i-1}, a_i),$$

for some simplicial metric d_{σ_i} on σ_i. Then we define a metric on X by setting $d_X(x, y)$ to be the infimum of lengths of all such paths from x to y. We call this metric the l^1-path metric of a simplicial complex.

Each simplex σ_i of $\mathbb{P}\,\mathrm{BHV}_n$ is embedded in an orthant spanned by a collection of compatible splits and corresponds to a tree topology. Since each orthant of $\mathbb{P}\,\mathrm{BHV}_n$ is Euclidean, the canonical metric on σ_i is the Euclidean metric, denoted $\|\cdot\|_{\mathbb{E}}$. However, imposing the spherical metric on $\mathbb{P}\,\mathrm{BHV}_n$ is not the same as imposing the Euclidean metric. For example, the link of BHV_4 at the origin is the Peterson graph. The distance between any two adjacent vertices in the Peterson graph under the spherical metric is $\frac{\pi}{2}$, but under the Euclidean metric it is $\sqrt{2}$. However, the map between the link of BHV_n at the origin with d_{σ_i} being the Euclidean metric and the metric space $\mathcal{L}_{\mathrm{BHV}}(0)$ with d_{σ_i} being the spherical metric is continuous:

Proposition 2.1.15 (Euclidean metric and spherical metric homeomorphism)**.** The link of BHV_n at the origin equipped with the Euclidean metric on each simplex is homeomorphic to the space equipped with the spherical metric.

Proof. Pick $T \in \mathcal{L}_{\mathrm{BHV}}(0)$ and a sufficiently small neighborhood U of T. If T is an interior point, a point $T' \in U$ is contained in the same orthant as T; if T is a boundary point, then the orthant that contains T' has different choices, but the orthant that contains T' also contains T. Therefore, we only need to show that $\|T - T'\|_{\mathbb{E}}$ is bounded by $\|T - T'\|_{\mathbb{S}}$ when T and T' are in the same orthant. Denote the angle between T and T' to be θ. Then

$$\|T - T'\|_{\mathbb{S}} = \angle(T, T') = \theta.$$

$\mathcal{L}_{\mathrm{BHV}}(0)$ is a subset of BHV_n, and each orthant of BHV_n is $\mathbb{R}_{\geq 0}^{n-2}$. Denote $T = (t_1, \dots, t_{n-2})$ and $T' = (t'_1, \dots, t'_{n-2})$. Then

$$\|T - T'\|_{\mathbb{E}}^2 = \sum_{i=1}^{n-2} (t_i - t'_i)^2. \tag{2.1}$$

By the law of cosines,

$$\|T - T'\|_{\mathbb{E}}^2 = c^2 + c^2 - 2c^2 \cos\theta. \tag{2.2}$$

Combining Equation 2.1 and 2.2 gives

$$\sum_{i=1}^{n-2} t_i t'_i = c^2 \cos\theta.$$

For all $\epsilon > 0$, take $\delta = 2c^2 - 2c^2 \cos\epsilon$, then

$$\|T - T'\|_{\mathbb{E}}^2 < \delta \Rightarrow \theta < \epsilon.$$

On the other side, for all $\epsilon' > 0$, take $\delta' = \arccos \frac{2c^2 - (\epsilon')^2}{2c^2}$. Because T, T' are in a small neighborhood U of T, θ is small. Therefore, $\arccos\theta$ decreases in this neighborhood. Thus,

$$\theta < \delta' \Rightarrow \cos\theta > \frac{2c^2 - (\epsilon')^2}{2c^2} \Rightarrow 2c^2 - 2c^2\cos\theta < (\epsilon')^2 \Rightarrow \|T - T'\|_{\mathbb{E}} < \epsilon'.$$

□

$\mathbb{P}\,\mathrm{BHV}_n$ is the quotient space of BHV_n by the equivalence relation of scaling by $\mathbb{R}_{\geq 0}$; a similar treatment can be found in Zairis-Khiabanian-Blumberg-Rabadan [58]. Billera-Holmes-Vogtmann assigned a particular choice to the representatives of $\mathbb{P}\,\mathrm{BHV}_n$ in Definition 1.1.2, which is the collection of coordinates $(x_1, \dots, x_{n-2})$, such that

$$\sum_{i=1}^{n-2} x_i = 1.$$

This collection of coordinates is a representative of the equivalence class

$$\sum_{i=1}^{n-2} x_i = k, k \in \mathbb{R}_{\geq 0}.$$

Then the collection of representatives is $\mathbb{P}\,\mathrm{BHV}_n$, and the union of equivalence classes is BHV_n.

According to Proposition 2.1.15, the link of origin is homeomorphic to $\mathbb{P}\,\mathrm{BHV}_n$. Billera-Holmes-Vogtmann pointed out that the entire tree space is an infinite cone on the link of origin [5]; this implies that the tree space is isomorphic to the cone over $\mathbb{P}\,\mathrm{BHV}_n$. We present a proof in the context of the projectivized space of phylogenetic trees.

Proposition 2.1.16 (BHV_n and $\mathbb{P}\,\mathrm{BHV}_n$ connection theorem)**.** $\mathbb{P}\,\mathrm{BHV}_n$ is the projectivized space of BHV_n, and the cone over $\mathbb{P}\,\mathrm{BHV}_n$ is isomorphic to BHV_n as sets.

Proof. Denote a point T_n of BHV_n by a set of compatible splits $\mathcal{S}_n^{n-2} = \left\{S_n^{A_1|B_1}, \ldots, S_n^{A_{n-2}|B_{n-2}}\right\}$:

$$T_n = \left(t_1 S_n^{A_1|B_1}, \cdots, t_{n-2} S_n^{A_{n-2}|B_{n-2}}\right), t_i \geq 0.$$

By definition, the image of the following quotient map is $\mathbb{P}\,\mathrm{BHV}_n$:

$$\pi : \mathrm{BHV}_n \setminus \{0\} \longrightarrow \mathrm{BHV}_n \setminus \{0\},$$
$$\left(t_1 S_n^{A_1|B_1}, \cdots, t_{n-2} S_n^{A_{n-2}|B_{n-2}}\right) \mapsto \left(t_1 S_n^{A_1|B_1}, \cdots, t_{n-2} S_n^{A_{n-2}|B_{n-2}}\right) / \sum_{i=1}^{n-2} t_i.$$

Therefore, $\mathbb{P}\,\mathrm{BHV}_n$ is the projectivized space of BHV_n, with the simplicial structure of $\mathbb{P}\,\mathrm{BHV}_n$ inherited from the cubical complex BHV_n.

On the other hand, $\mathbb{P}\,\mathrm{BHV}_n$ is the quotient space of BHV_n, and we denote each element in $\mathbb{P}\,\mathrm{BHV}_n$ by

$$\overline{T}_n = \left(\hat{t}_1 S_n^{A_1|B_1}, \cdots, \hat{t}_{n-2} S_n^{A_{n-2}|B_{n-2}}\right)$$

with $\hat{t}_i \geq 0$ and $\sum_{i=1}^{n-2} \hat{t}_i = 1$. Each element in $\mathbb{P}\,\mathrm{BHV}_n$ is a representative of the equivalence class of rescaled trees in BHV_n. We define the 0-cone over $\mathbb{P}\,\mathrm{BHV}_n$ as

$$\mathrm{BHV}_n^{\star} = (\mathbb{P}\,\mathrm{BHV}_n \times \mathbb{R}_{\geq 0}) / (\mathbb{P}\,\mathrm{BHV}_n \times 0).$$

Each point in $\mathrm{BHV}_n^{\star}$ can be represented by

$$\overline{T}_n^k = \left(k, \hat{t}_1 S_n^{A_1|B_1}, \cdots, \hat{t}_{n-2} S_n^{A_{n-2}|B_{n-2}}\right)$$

for $k \geq 0$. Consider the map

$$\xi : \mathrm{BHV}_n^\star \longrightarrow \mathrm{BHV}_n,$$
$$\left(k, \hat{t}_1 S_n^{A_1|B_1}, \cdots, \hat{t}_{n-2} S_n^{A_{n-2}|B_{n-2}}\right) \mapsto \left(k\hat{t}_1 S_n^{A_1|B_1}, \cdots, k\hat{t}_{n-2} S_n^{A_{n-2}|B_{n-2}}\right).$$

The map ξ takes a point $\overline{T}_n^k$ in $\mathrm{BHV}_n^\star$ to a point T_n in BHV_n with the rescaling factor k. Since the map ξ is a bijection, the 0-cone over $\mathbb{P}\,\mathrm{BHV}_n$ is isomorphic to BHV_n as sets. $\square$

Furthermore, BHV_n is congruent to the cone over $\mathbb{P}\,\mathrm{BHV}_n$ with the 0-cone metric. For a formal proof, we need to define the κ-cone over a metric space, starting from the model spaces M_κ^n.

Definition 2.1.17 (Model spaces M_κ^n)**.** Given a real number κ, we denote by M_κ^n the following metric spaces [8].

1. If $\kappa = 0$ then M_0^n is Euclidean space $\mathbb{E}^n$;
2. If $\kappa > 0$ then M_κ^n is obtained from the sphere $\mathbb{S}^n$ by multiplying the distance function by the constant $1/\sqrt{\kappa}$;
3. If $\kappa < 0$ then M_κ^n is obtained from hyperbolic space $\mathbb{H}^n$ by multiplying the distance function by $1/\sqrt{-\kappa}$.

Definition 2.1.18 (The diameter of M_κ^n)**.** We write D_κ to denote the *diameter* of M_κ^n. More precisely, $D_\kappa := \pi/\sqrt{\kappa}$ for $\kappa > 0$ and $D_\kappa := \infty$ for $\kappa \leq 0$ [8].

Definition 2.1.19 (The κ-cone over a metric space)**.** Given a metric space Y and a real number κ, the *κ-cone* $X = C_\kappa Y$ over Y is the metric space defined as follows: If $\kappa \leq 0$ then, as a set, X is the quotient of $[0, \infty) \times Y$ by the equivalence relation given by:

$$(t, y) \sim (t', y')$$

if $(t = t' = 0)$ or $(t = t' > 0$ and $y = y')$. If $\kappa > 0$ then X is the quotient of $[0, D_\kappa/2] \times Y$ by the same relation [8].

Definition 2.1.20 (The distance between two points on a 0-cone)**.** The *distance* between two points $x = ty$ and $x' = t'y'$ in a κ-cone X for $\kappa = 0$ is defined as:

$$d_\kappa(x, x')^2 = t^2 + (t')^2 - 2tt' \cos(d_\pi(y, y')),$$

for $d_\pi(y, y') = \min\{\pi, d(y, y')\}$ [8].

Connecting a point $T_n \in \mathrm{BHV}_n$ to the origin by a straight line segment, then connecting the origin to $T'_n \in \mathrm{BHV}_n$ by another straight line segment is a *cone path* from T_n to T'_n [5]. The cone path is a geodesic if and only if the angle between T_n and T'_n is at least π [5, 8].

Proposition 2.1.21 (BHV_n and 0-cone isometry)**.** The 0-cone over $\mathbb{P}\,\mathrm{BHV}_n$ is isometric to BHV_n.

Proof. Denote $\overline{T}_n \in \mathbb{P}\,\mathrm{BHV}_n$ by sets of compatible splits, such that:

$$\overline{T}_n = \left(\hat{t}_1 S_n^{A_1|B_1}, \cdots, \hat{t}_{n-2} S_n^{A_{n-2}|B_{n-2}}\right), \hat{t}_i \geq 0 \text{ and } \sum_{i=1}^{n-2} \hat{t}_i = 1.$$

Denote $\overline{T}'_n \in \mathbb{P}\,\mathrm{BHV}_n$ by

$$\overline{T}'_n = \left(\hat{t}'_1 S_n^{A'_1|B'_1}, \cdots, \hat{t}'_{n-2} S_n^{A'_{n-2}|B'_{n-2}}\right), \hat{t}'_i \geq 0 \text{ and } \sum_{i=1}^{n-2} \hat{t}'_i = 1.$$

Let $\mathrm{BHV}_n^\star$ be the 0-cone over $\mathbb{P}\,\mathrm{BHV}_n$. By Definition 2.1.20, the distance between two points

$$T_n^\star = \left(k, \hat{t}_1 S_n^{A_1|B_1}, \cdots, \hat{t}_{n-2} S_n^{A_{n-2}|B_{n-2}}\right), \hat{t}_i \geq 0$$

and

$$T_n'^\star = \left(k', \hat{t}'_1 S_n^{A'_1|B'_1}, \cdots, \hat{t}'_{n-2} S_n^{A'_{n-2}|B'_{n-2}}\right), \hat{t}'_i \geq 0$$

in a 0-cone is:

$$d_\kappa\left(T_n^\star, T_n'^\star\right)^2 = k^2 + (k')^2 - 2kk' \cos\left(d_\pi\left(\overline{T}_n, \overline{T}'_n\right)\right)$$

for $d_\pi\left(\overline{T}_n, \overline{T}'_n\right) = \min\left\{\pi, d\left(\overline{T}_n, \overline{T}'_n\right)\right\}$. According to Definition 2.1.7, if two points of BHV_n locate in the same orthant, then the geodesic between these two points is the Euclidean distance, agreeing with the metric on a 0-cone.

If two points $T_n, T'_n \in \mathrm{BHV}_n$ are in different orthants, we join them by a sequence of straight segments, with each segment lying in a single orthant, and the distance between the two points is the minimum of the lengths of such segmented paths joining the two points, according to the metric defined on the BHV_n space. If the angle between T_n and T'_n is less than π, then

$$d_{\mathrm{BHV}}\left(T_n, T'_n\right) = d_\kappa\left(T_n^\star, T_n'^\star\right)$$

since the minimum length of segments joining T_n and T'_n is the length of the opposite side of θ in the triangle spanned by T_n and T'_n with $\angle\left(T_n, T'_n\right) = \theta$.

If the angle between T_n and T'_n is equal to or greater than π, then the cone path is a geodesic between T_n and T'_n, which is $\|T_n\| + \|T'_n\|$. Meanwhile, $d_\pi\left(\overline{T}_n, \overline{T}'_n\right) = \pi$, and therefore,

$$d_\kappa\left(T_n^\star, T_n'^\star\right) = t + t' = d_{\mathrm{BHV}}\left(T_n, T'_n\right).$$

Therefore, the metric on the 0-cone over $\mathbb{P}\,\mathrm{BHV}_n$ is the same as the BHV metric; hence, the 0-cone over $\mathbb{P}\,\mathrm{BHV}_n$ is isometric to BHV_n. □

2.2 The Algebraic Fan $\mathcal{F}(D_n)$ Over Simplicial Complexes $\mathcal{R}(\Delta_n)$

In the following section, we will show the isomorphisms between the spaces of phylogenetic trees and these simplicial complexes. We first introduce root subsystems of type D and simplicial complexes whose 0-simplices are the root subsystems of type D. We consider the root system of type D of the form

$$D_n = \{\pm\epsilon_i \pm \epsilon_j \mid 1 \leq i < j \leq n\},$$

where $\epsilon_1, \ldots, \epsilon_n$ denote the usual orthonormal unit vectors which form a basis of $\mathbb{R}^n$. We denote *root subsystems* $\Theta_n^{A_k|B_k}$ of D_n as follows:

$$\Theta_n^{A_k|B_k} = \{\pm\epsilon_i \pm \epsilon_j \mid i \neq j; i, j \in A_k\},\ A_k \subset \{1, \ldots, n\},\ \text{and } B_k = \{1, \ldots, n\} \setminus A_k.$$

We define a simplicial complex $\mathcal{R}(\Delta_n)$ according to Hacking-Keel-Tevelev [25]. Let $\mathcal{D}_{k,2m}$ and $\mathcal{D}_{k,2m+1}$ be sets of root subsystems $\Theta_n^{A_i|B_i}$, such that:

for $k = 1, \ldots, m$:

$$\mathcal{D}_{k,2m+1} = \left\{\Theta_{2m+1}^{A_1|B_1}, \ldots, \Theta_{2m+1}^{A_L|B_L}\right\}, \text{ where } L = \binom{2m+1}{k};$$

for $k = 1, \ldots, m-1$:

$$\mathcal{D}_{k,2m} = \left\{\Theta_{2m}^{A_1|B_1}, \ldots, \Theta_{2m}^{A_L|B_L}\right\}, \text{ where } L = \binom{2m}{k};$$

for $k = m$:

$$\mathcal{D}_{m,2m} = \left\{\Theta_{2m}^{A_1|B_1}, \ldots, \Theta_{2m}^{A_L|B_L}\right\}, \text{ where } L = \binom{2m}{m} \Big/ 2.$$

Definition 2.2.1 (The set $\mathcal{R}(D_n)$)**.** We define the set $\mathcal{R}(D_{2m+1})$ and $\mathcal{R}(D_{2m})$ as below:

$$\mathcal{R}(D_{2m+1}) = \mathcal{D}_{2,2m+1} \sqcup \mathcal{D}_{3,2m+1} \sqcup \cdots \sqcup \mathcal{D}_{m,2m+1};$$
$$\mathcal{R}(D_{2m}) = \mathcal{D}_{2,2m} \sqcup \mathcal{D}_{3,2m} \sqcup \cdots \sqcup \mathcal{D}_{m-1,2m} \sqcup (\mathcal{D}_{m,2m} \times \mathcal{D}_{m,2m}).$$

Therefore, $\mathcal{R}(D_n)$ for $n = 2m$ or $n = 2m+1$ is a set of root subsystems with elements from the sets of root subsystems $\mathcal{D}_{k,n}$ for k from 2 to m.

Definition 2.2.2 (The simplicial complex $\mathcal{R}(\Delta_n)$)**.** Define the simplicial complex $\mathcal{R}(\Delta_n)$ with 0-simplices being elements in the set $\mathcal{R}(D_n)$. Higher-dimensional simplices $\Theta_n \subset \mathcal{R}(\Delta_n)$ are formed by root subsystems, such that for each pair of $\Theta_n^{A_i|B_i}, \Theta_n^{A_j|B_j} \in \Theta_n$, one of the following criteria is fulfilled:

$$\Theta_n^{A_i|B_i} \perp \Theta_n^{A_j|B_j}, \quad \Theta_n^{A_i|B_i} \subset \Theta_n^{A_j|B_j}, \quad \text{or} \quad \Theta_n^{A_j|B_j} \subset \Theta_n^{A_i|B_i}.$$

$\mathcal{R}(\Delta_n)$ is a simplicial complex by passage to subsets: for a collection of root subsystems Θ_n that forms a simplex, each element in Θ_n must be perpendicular to or have an inclusion relation with all other elements in Θ_n. This defining relation is true for any subset of Θ_n; hence, the faces of each simplex in $\mathcal{R}(\Delta_n)$ is also a simplex of $\mathcal{R}(\Delta_n)$. Pairs of elements in the non-empty intersection of two collections of root subsystems also satisfy one of the three criteria pairwise; hence, the intersection is a face of each.

2.3 Homeomorphism Between $\mathbb{P}\,\mathrm{BHV}_{n-1}$ and $\mathcal{R}(\Delta_n)$

Theorem 2.3.1 (Main Theorem: $\mathbb{P}\,\mathrm{BHV}_{n-1}$ and $\mathcal{R}(\Delta_n)$ isomorphism)**.** $\mathbb{P}\,\mathrm{BHV}_{n-1}$ and $\mathcal{R}(\Delta_n)$ are isomorphic abstract simplicial complexes.

Proof. We use $\mathfrak{S}_{n-1}$ to denote the set of splits in BHV_{n-1}:

$$\mathfrak{S}_{n-1} = \left\{ S_{n-1}^{A_1|B_1}, \ldots, S_{n-1}^{A_M|B_M} \right\},$$

where $M = |\mathfrak{S}_{n-1}| = 2^n - n - 1$. The set of vertices of $\mathbb{P}\,\mathrm{BHV}_{n-1}$ is the same as the set of internal edges in BHV_{n-1}. Consider the map ω between $\mathfrak{S}_{n-1} \subset \mathbb{P}\,\mathrm{BHV}_{n-1}$ and the set $\mathcal{R}(D_n)$:

$$\begin{aligned} \omega : \mathfrak{S}_{n-1} &\longrightarrow \mathcal{R}(D_n), \\ S_{n-1}^{A_i|B_i} &\mapsto \Theta_n^{A_i|B_i}. \end{aligned}$$

The map ω is an isomorphism on the 0-simplices of $\mathbb{P}\,\mathrm{BHV}_{n-1}$ and $\mathcal{R}(D_n)$. To show that there is an isomorphism between the abstract simplicial complexes $\mathbb{P}\,\mathrm{BHV}_{n-1}$ and $\mathcal{R}(\Delta_n)$, we prove that a collection of 0-simplices forming a k-simplex of $\mathbb{P}\,\mathrm{BHV}_{n-1}$ also forms a k-simplex of $\mathcal{R}(\Delta_n)$, and vice versa. Since

$$\Theta_n^{A_m|B_m} = \{\pm\epsilon_i \pm \epsilon_j \mid i \neq j; i, j \in A_m\},\ A_m \subset \{1, \ldots, n\},\ \text{and}\ B_m = \{1, \ldots, n\} \setminus A_m$$

for $k = 0$, ω maps 0-simplices of $\mathbb{P}\,\mathrm{BHV}_{n-1}$ bijectively to those of $\mathcal{R}(\Delta_n)$. The k-simplices of $\mathbb{P}\,\mathrm{BHV}_{n-1}$ are projected from BHV_{n-1} with exactly $k+1$ nonzero edges, denoted by a collection of 0-simplices

$$\mathcal{S}_{n-1}^{k+1} = \left\{ S_{n-1}^{A_1|B_1}, \ldots, S_{n-1}^{A_{k+1}|B_{k+1}} \right\}.$$

For $k > 0$, we show that the collection of 0-simplices of $\mathcal{R}(\Delta_n)$

$$\Theta_n = \left\{ \omega\left(S_{n-1}^{A_1|B_1}\right), \ldots, \omega\left(S_{n-1}^{A_{k+1}|B_{k+1}}\right) \right\}$$

forms a k-simplex of $\mathcal{R}(\Delta_n)$. Two partitions $S_n^{A_i|B_i}$ and $S_n^{A_j|B_j}$ are defined to be compatible if one of the subsets

$$A_i \cap A_j, \qquad A_i \cap B_j, \qquad A_j \cap B_i, \qquad \text{or} \qquad B_i \cap B_j$$

is empty. Since a tree in BHV_m consists of internal edges that correspond to compatible partitions, any two elements $S_{n-1}^{A_i|B_i}, S_{n-1}^{A_j|B_j} \in \mathcal{S}_{n-1}^{k+1}$ are compatible and satisfy the above criteria. In the following, we will show that if one of the subsets is empty, then one of the criteria in Definition 2.2.2 is fulfilled, and vice versa. If $|A_i| \leq |B_i|$ and $|A_j| \leq |B_j|$, then

$$\begin{aligned}
&A_i \cap A_j = \emptyset \Rightarrow \Theta_n^{A_i|B_i} \perp \Theta_n^{A_j|B_j};\\
&A_i \cap B_j = \emptyset \Rightarrow \Theta_n^{A_i|B_i} \subset \Theta_n^{A_j|B_j}, \text{ i.e., } A_i \subset A_j;\\
&A_j \cap B_i = \emptyset \Rightarrow \Theta_n^{A_j|B_j} \subset \Theta_n^{A_i|B_i}, \text{ i.e., } A_j \subset A_i;\\
&B_i \cap B_j = \emptyset \Rightarrow \Theta_n^{A_i|B_i} \times \Theta_n^{A_j|B_j} = \Theta_n^{A_j|B_j} \times \Theta_n^{A_i|B_i}, \text{ i.e., } S_{n-1}^{A_i|B_i} = S_{n-1}^{A_j|B_j}.
\end{aligned}$$

In particular, $B_i \cap B_j = \emptyset$ implies that the splits $S_{n-1}^{A_i|B_i}$ and $S_{n-1}^{A_j|B_j}$ are different representations of the same split, and hence are not eligible for comparision, i.e., this case is vacuous. If $|A_i| \leq |B_i|$ and $|A_j| > |B_j|$, then

$$\begin{aligned}
&A_i \cap A_j = \emptyset \Rightarrow \Theta_n^{A_i|B_i} \subset \Theta_n^{B_j|A_j}, \text{ i.e., } A_i \subset B_j;\\
&A_i \cap B_j = \emptyset \Rightarrow \Theta_n^{A_i|B_i} \perp \Theta_n^{B_j|A_j};\\
&A_j \cap B_i = \emptyset \Rightarrow \Theta_n^{A_i|B_i} = \Theta_n^{A_j|B_j}, \text{ is impossible since } |A_j| > |B_j|;\\
&B_i \cap B_j = \emptyset \Rightarrow \Theta_n^{B_j|A_j} \subset \Theta_n^{A_i|B_i}, \text{ i.e., } B_j \subset A_i.
\end{aligned}$$

This shows that if $|A_i| \leq |B_i|$, then satisfying one of the compatibility conditions implies one of the criteria for forming a simplex in $\mathcal{R}(\Delta_n)$ is satisfied. By symmetry, this is also true for $|B_i| \leq |A_i|$ by interchanging i and j. Now we will show that if one of the criteria for forming a simplex in $\mathcal{R}(\Delta_n)$ is satisfied, then one of the subsets

$$A_i \cap A_j, \qquad A_i \cap B_j, \qquad A_j \cap B_i, \qquad \text{or} \qquad B_i \cap B_j$$

is empty. If $\Theta_n^{A_i|B_i} \perp \Theta_n^{A_j|B_j}$, then $A_i \cap A_j = \emptyset$; If $\Theta_n^{A_i|B_i} \subset \Theta_n^{A_j|B_j}$, then $A_i \cap B_j = \emptyset$; if $\Theta_n^{A_j|B_j} \subset \Theta_n^{A_i|B_i}$, then $A_j \cap B_i = \emptyset$. Since the 0-simplices of $\mathbb{P}\,\mathrm{BHV}_{n-1}$ and $\mathcal{R}(D_n)$ are isomorphic, and higher-dimensional simplices in $\mathbb{P}\,\mathrm{BHV}_{n-1}$ and $\mathcal{R}(D_n)$ are formed in the same way, abstract complexes $\mathbb{P}\,\mathrm{BHV}_{n+1}$ and $\mathcal{R}(\Delta_n)$ are isomorphic. $\square$

Definition 2.3.2 (Vertex scheme)**.** If K is a simplicial complex, let V be the vertex set of K. Let $\mathcal{K}$ be the collection of all subsets $\{a_0, \ldots, a_n\}$ of V, such that the vertices $a_0, \ldots, a_n$ span a simplex of K. The collection $\mathcal{K}$ is called the *vertex scheme* of K [40].

Definition 2.3.3 (Geometric realization)**.** If the abstract simplicial complex $\mathcal{S}$ is isomorphic with the vertex scheme of the simplicial complex K, we call K a *geometric realization* of $\mathcal{S}$. It is uniquely determined up to a linear isomorphism [40].

According to Munkres [40], isomorphisms between abstract complexes induce homeomorphisms between their geometric realizations. Theorem 2.3.1 proved that abstract complexes $\mathbb{P}\,\mathrm{BHV}_{n+1}$ and $\mathcal{R}(\Delta_n)$ are isomorphic. Therefore, $\mathbb{P}\,\mathrm{BHV}_{n+1}$ and the geometric realization of $\mathcal{R}(\Delta_n)$ are homeomorphic. As a result:

Corollary 2.3.4 ($\mathbb{P}\,\mathrm{BHV}_{n-1}$ and $\mathcal{R}(\Delta_n)$ homeomorphism)**.** The geometric realization of $\mathbb{P}\,\mathrm{BHV}_{n-1}$ and $\mathcal{R}(\Delta_n)$ are homeomorphic.

Definition 2.3.5 (The polytope $P_{\sigma_i}(D_n)$)**.** Let σ_i be in the symmetric group S_n. A polytope $P_{\sigma_i}(D_n)$ is diffeomorphic to the closure of each cell in $\overline{\mathcal{M}}_{0,n}$ as a manifold with corners corresponding to the circular ordering $\{\sigma_i(1), \ldots, \sigma_i(n)\}$. Faces of $P_{\sigma_i}(D_n)$ correspond to the boundary divisors of $\overline{\mathcal{M}}_{0,n}$ which meet the closure of the cell. Each $P_{\sigma_i}(D_n)$ is identified with an associahedron with each vertex representing the triangulation of an n-gon, with one facet for each diagonal, depicted in Fig. 2.3.

Definition 2.3.6 (The polytope $\mathcal{P}(D_n)$)**.** $\mathcal{P}(D_n)$ is the collection of polytopes $P_{\sigma_i}(D_n)$ identified along common faces:

$$\mathcal{P}(D_n) = \coprod_{\sigma_i} P_{\sigma_i}(D_n)/\sim .$$

According to Definition 2.1.5, a circular ordering identifies $\{x_1, \ldots, x_n\}$ with circular orderings $\{x_n, x_{n-1}, \ldots, x_1\}$ and $(x_2, x_3, \ldots, x_n, x_1)$. Therefore, there are $(n-1)!/2$ copies of $P_{\sigma_i}(D_n)$ in $\mathcal{P}(D_n)$, created by S_n acting on labeled vertices modulo the dihedral group. The 0-dimensional simplices in $\mathbb{P}\,\mathrm{BHV}_{n-1}$ are in bijection with the $(n-4)$-dimensional spaces (top-dimensional) in $\mathcal{P}_{\sigma_i}(D_n)$; a k-dimensional simplex in $\mathbb{P}\,\mathrm{BHV}_{n-1}$ is formed by $k+1$ compatible bipartitions, and each bipartition corresponds to a facet of an associahedron. We summarize this dual relation between $\mathcal{R}(\Delta_n)$ and $\mathcal{P}(D_n)$ as the following:

Theorem 2.3.7 ($\mathcal{R}(\Delta_n)$ and $\mathcal{P}(D_n)$ are duals)**.** $\mathcal{R}(\Delta_n)$ and $\mathcal{P}(D_n)$ are duals: k-simplices of $\mathcal{R}(\Delta_n)$ are in bijection with $(n-k-4)$-dimensional faces of $\mathcal{P}(D_n)$; a k-simplex of $\mathcal{R}(\Delta_n)$

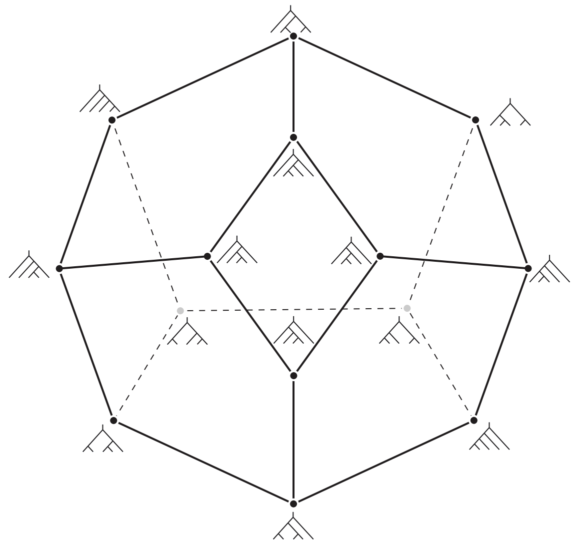


Figure 2.3: The associahedron $P_{\sigma_i}(D_5)$.

consists of a collection of root subsystems Θ forms a $(k+1)$-simplex of $\mathcal{R}(\Delta_n)$ with some root subsystem $\Theta_n^{A|B}$ if and only if the $(n-k-4)$-dimensional faces of $\mathcal{P}(D_n)$ that corresponds to Θ form an $(n-k-5)$-dimensional face of $\mathcal{P}(D_n)$ with a facet that corresponds to $\Theta_n^{A|B}$.

Proof. The k-simplices of $\mathcal{R}(\Delta_n)$ are isomorphic to the k-simplices of $\mathbb{P}\,\mathrm{BHV}_{n-1}$, which are $k+1$ compatible bipartitions. These $k+1$ compatible bipartitions correspond to $k+1$ diagonals in an n-gon. By Definition 2.3.5, faces of $P_{\sigma_i}(D_n)$ can be identified with an associahedron with each vertex representing the triangulation of an n-gon, with one facet for each diagonal. This implies that k-dimensional faces of $P_{\sigma_i}(D_n)$ correspond to an n-gon with $n-k-3$ diagonals, and $(n-k-4)$-dimensional faces of $P_{\sigma_i}(D_n)$ correspond to an n-gon with $k+1$ diagonals, which are $k+1$ compatible splits identified with k-simplices of $\mathcal{R}(\Delta_n)$. Identifying faces from $P_{\sigma_i}(D_n)$ to form $\mathcal{P}(D_n)$ does not affect the dimensionality of the space; therefore, k-simplices of $\mathcal{R}(\Delta_n)$ are in bijection with $(n-k-4)$-dimensional faces of $\mathcal{P}(D_n)$,

A k-simplex of $\mathcal{R}(\Delta_n)$ that corresponds to a collection of $k+1$ root subsystems Θ forms a $(k+1)$-simplex if and only if the new root subsystem $\Theta_n^{A|B}$ has a perpendicular or inclusion relation with each element in Θ. This condition is the same as a new split being compatible with each split in a set of compatible splits $\mathcal{S}_{n-1}$ that corresponds to Θ.

A bipartition compatible with other bipartitions corresponds to a diagonal of an n-gon that does not intersect with other existing diagonals of the n-gon. Therefore, adding a new diagonal to $(n-k-4)$-dimensional faces of $\mathcal{P}(D_n)$ creates an $(n-k-5)$-dimensional face of $\mathcal{P}(D_n)$.

$\square$

2.4 Isometry Between BHV_{n-1} and $\mathcal{F}(D_n)$

A lattice N is isomorphic to $\mathbb{Z}^n$ for some n, and we define the real vector space $N_{\mathbb{R}} = N \otimes_{\mathbb{Z}} \mathbb{R}$. First we define the concept of a convex polyhedral cone.

Definition 2.4.1 (Convex polyhedral cone)**.** A *convex polyhedral cone* is a set

$$\sigma = \{r_1 v_1 + \cdots r_s v_s \in N_{\mathbb{R}} : r_i \geq 0\}$$

generated by any finite set of vectors $v_1, \ldots, v_s$ in $N_{\mathbb{R}}$. A *strongly convex rational polyhedral cone* σ in $N_{\mathbb{R}}$ is a cone with apex at the origin, generated by a finite number of vectors; *rational* means that it is generated by vectors in the lattices, and *strong convexity* means that it contains no line through the origin [22].

Let $M = \mathrm{Hom}(N, \mathbb{Z})$ denote the dual lattice, with dual pairing denoted $\langle , \rangle$.

Definition 2.4.2 (Dual cone [22])**.** If σ is a cone in N, the *dual cone* $\sigma^\vee$ is the set of vectors in $M_{\mathbb{R}} = M \otimes_{\mathbb{Z}} \mathbb{R}$ that are nonnegative on σ, i.e., it is the set of equations of supporting hyperplanes, i.e.,

$$\sigma^\vee = \{u \in V^* : \langle u, v\rangle \geq 0 \text{ for all } v \in \sigma\}.$$

Definition 2.4.3 (Supporting hyperplane)**.** Given a (convex) set $S \subset \mathbb{R}^n$, we say that a hyperplane $H \subset \mathbb{R}^n$ is a *supporting hyperplane* of S if all of S is contained in one of the halfspaces bounded by H and if H contains at least one boundary point of S [22].

Definition 2.4.4 (Face of a cone)**.** The *face* τ of a cone σ is the intersection of σ with any supporting hyperplane:

$$\tau = \sigma \cap u^\perp = \{v \in \sigma : \langle u, v\rangle = 0\}$$

for some u in $\sigma^\vee$ [22].

Definition 2.4.5 (Fan)**.** A *fan* in N is a collection of strongly convex rational polyhedral cones σ in the real vector space $N_{\mathbb{R}}$, such that every face of a cone in the fan is also a cone in the fan, and the intersection of two cones in the fan is a face of each [22].

Denote $D_n = \{\alpha \in \Lambda \mid \alpha^2 = -2\}$ as the root system of type D, where Λ is a negative-definite $\mathbb{Z}$-lattice spanned by D_n. Consider the linear map

$$\begin{aligned}\phi : \operatorname{Sym}^2 \Lambda^\vee &\longrightarrow \mathbb{Z}^{\mathcal{A}_1(D_n)}\\ f &\mapsto \sum_{A_1 \in \mathcal{A}_1(D_n)} f(A_1)[A_1],\end{aligned}$$

where $\operatorname{Sym}^2 \Lambda^\vee$ is the space of quadratic forms on Λ, and $f(A_1)$ is equal to the value of f on one of the opposite roots of A_1. Define $N(D_n)$ as Coker ϕ, and let ψ be the canonical map to quotient module. Let $\psi : \mathbb{Z}^{D_n^+} \to N(D_n)$ be the dual map. For any root subsystem $J \subset D_n$, let

$$\psi(J) := \sum_{\alpha \in J \cap D_n^+} \psi(\alpha).$$

In the following, we present a formal definition of the algebraic fan $\mathcal{F}(D_n)$ according to Hacking-Keel-Tevelev [25].

Definition 2.4.6 (The fan $\mathcal{F}(D_n)$)**.** Define strongly convex rational polyhedral cones $\tau_{\mathcal{F}}$ in $N_{\mathbb{R}}$ as a set:

$$\tau_{\mathcal{F}} = \left\{ r_1 \psi\left(\Theta_n^{A_1|B_1}\right) + \cdots + r_n \psi\left(\Theta_n^{A_k|B_k}\right) \in N(D_n)_{\mathbb{Q}} : r_i \geq 0 \right\}.$$

The fan $\mathcal{F}(D_n)$ in $N(D_n)_{\mathbb{Q}}$ is a collection of strongly convex rational polyhedral cones $\tau_{\mathcal{F}}$ in the real vector space $N(D_n)_{\mathbb{Q}}$ determined by the rays $\psi\left(\Theta_n^{A_i|B_i}\right)$ for each $\Theta_n^{A_i|B_i} \in \mathcal{R}(D_n)$, where rays span a cone in $\mathcal{F}(D_n)$ if and only if the corresponding subsystems form a simplex in $\mathcal{R}(\Delta_n)$.

A simplicial complex K in $\mathbb{R}^N$ is a collection of simplices in $\mathbb{R}^N$, such that every face of a simplex of K is in K and the intersection of any two simplices of K is a face of each of them; meanwhile, every face of a cone in the fan is also a cone in the fan, and the intersection of two cones in the fan is a face of each, which is analogous to a simplicial complex [22]. Meanwhile, Zairis-Khiabanian-Blumberg-Rabadan defined cubical complexes as metric spaces obtained by gluing together cubes via the data of isometries of faces, subject to the condition that two cubes are connected by at most a single face identification and no cube is glued to itself [59]. In the following theorem, we show that $\mathcal{F}(D_n)$ inherits a simplicial structure from $\mathcal{R}(\Delta_n)$ and therefore is a fan.

Proposition 2.4.7. $\mathcal{F}(D_n)$ is a fan.

Proof. To prove $\mathcal{F}(D_n)$ is a fan, we need to show that every face of a cone in $\mathcal{F}(D_n)$ is also a cone in $\mathcal{F}(D_n)$, and the intersection of two cones in $\mathcal{F}(D_n)$ is a face of each. Consider a face τ of a cone $\sigma \subset \mathcal{F}(D_n)$. By definition, τ is the intersection of σ with any supporting hyperplane:

$$\tau = \sigma \cap u^{\perp} = \{v \in \sigma : \langle u, v\rangle = 0\}$$

for some u in $\sigma^{\vee}$. Because τ is the intersection of σ with a supporting hyperplane, τ is a subset of σ and spanned by a collection of pairwise compatible root subsystems. Since the root subsystems spanning σ are pairwise compatible, the root subsystems spanning τ are also pairwise compatible. Recall that rays span a cone in $\mathcal{F}(D_n)$ if and only if the corresponding subsystems form a simplex in $\mathcal{R}(\Delta_n)$. Therefore, τ is also a cone in the fan.

Since a cone is a face of itself [22], we have:

$$\sigma = \sigma \cap u^{\perp} = \{v \in \sigma : \langle u, v\rangle = 0\}$$

for some u in $\sigma^{\vee}$, and

$$\tilde{\sigma} = \tilde{\sigma} \cap \tilde{u}^{\perp} = \{v \in \tilde{\sigma} : \langle \tilde{u}, v\rangle = 0\}$$

for some $\tilde{u}$ in $\tilde{\sigma}^{\vee}$.

Let $\tilde{\tau}$ be the intersection of two cones $\sigma, \tilde{\sigma} \subset \mathcal{F}(D_n)$, then

$$\tilde{\tau} = (\sigma \cap u^{\perp}) \cap (\tilde{\sigma} \cap \tilde{u}^{\perp}) = \sigma \cap (\tilde{\sigma} \cap u^{\perp} \cap \tilde{u}^{\perp}).$$

By definition,

$$\sigma^{\vee} = \{u \in V^* : \langle u, v\rangle \geq 0 \text{ for all } v \in \sigma\}.$$

Therefore, $\tilde{\sigma} \cap u^{\perp} \cap \tilde{u}^{\perp}$ is in $\sigma^{\vee}$. So $\tilde{\tau}$ is a face of σ. And similarly,

$$\tilde{\tau} = \tilde{\sigma} \cap (\sigma \cap u^{\perp} \cap \tilde{u}^{\perp}).$$

So $\tilde{\tau}$ is a face of $\tilde{\sigma}$. Therefore, τ' is a face of the cones σ and σ', respectively, and we conclude that $\mathcal{F}(D_n)$ is a fan.

□

The definition of the fan $\mathcal{F}(D_n)$ determined by $\mathcal{R}(\Delta_n)$ induces an isomorphism between $\mathcal{F}(D_n)$ and $\mathcal{R}(\Delta_n)$. This relation is analogous to that between $\mathbb{P}\,\mathrm{BHV}_n$ and BHV_n. BHV_n is the cone over $\mathbb{P}\,\mathrm{BHV}_n$, and $\mathcal{F}(D_n)$ is the cone over $\mathcal{R}(\Delta_n)$, both consisting of rays emanating from the suspension point to infinity. Since $\mathcal{R}(\Delta_n)$ is isomorphic to $\mathbb{P}\,\mathrm{BHV}_{n-1}$, this isomorphism also passes down to $\mathcal{F}(D_n)$ and BHV_{n-1}.

Theorem 2.4.8 (BHV_{n-1} and $\mathcal{F}(D_n)$ isomorphism)**.** BHV_{n-1} and $\mathcal{F}(D_n)$ are isomorphic.

Proof. Each k-dimensional orthant in BHV_{n-1} can be parametrized by a set of compatible splits $\mathcal{S}_{n-1}^{k} = \left\{ S_{n-1}^{A_1|B_1}, \ldots, S_{n-1}^{A_k|B_k} \right\}$:

$$\sigma_{n-1}^{k} = \left\{ t_1 S_{n-1}^{A_1|B_1} + \cdots + t_k S_{n-1}^{A_k|B_k} \in \mathrm{BHV}_{n-1} : t_i \geq 0 \right\}.$$

Each k-dimensional cone of the fan $\mathcal{F}(D_n)$ can be parametrized by

$$\tau_{\mathcal{F}}^{k} = \left\{ r_1 \psi\left(\Theta_n^{A_1|B_1}\right) + \cdots + r_n \psi\left(\Theta_n^{A_k|B_k}\right) \in N(D_n)_{\mathbb{Q}} : r_i \geq 0 \right\}$$

for a collection of root subsystems $\Theta_n^k = \left\{ \Theta_n^{A_1|B_1}, \ldots, \Theta_n^{A_k|B_k} \right\}$ in $\mathcal{R}(D_n)$. Therefore, there is an isomorphism between a k-dimensional orthant σ_{n-1}^{k} in BHV_{n-1} and a k-dimensional cone $\tau_{\mathcal{F}}^{k}$ of $\mathcal{F}(D_n)$ obtained by extending $\omega : \mathfrak{S}_{n-1} \longrightarrow \mathcal{R}(D_n)$ as in the following:

$$\hat{\omega} : \sigma_{n-1}^{k} \longrightarrow \tau_{\mathcal{F}}^{k},$$
$$r_1 S_{n-1}^{A_1|B_1} + \cdots + r_k S_{n-1}^{A_k|B_k} \mapsto r_1 \Theta_n^{A_1|B_1} + \cdots + r_k \Theta_n^{A_k|B_k}.$$

We notice that the fan $\mathcal{F}(D_n)$ is constructed by abstract cones $\tau_{\mathcal{F}}^{k}$ according to $\mathcal{R}(\Delta_n)$, and the cubical complex BHV_{n-1} is formed by emanating rays over $\mathbb{P}\,\mathrm{BHV}_{n-1}$. Since $\mathcal{R}(\Delta_n)$ and

$\mathbb{P}\,\mathrm{BHV}_{n-1}$ are isomorphic, the isomorphism $\hat{\omega}$ extends to BHV_{n-1} orthants and maximal dimensional cones in $\mathcal{F}(D_n)$:

$$\Omega : \mathrm{BHV}_{n-1} \longrightarrow \mathcal{F}(D_n),$$
$$r_1 S_{n-1}^{A_1|B_1} + \cdots + r_{n-3} S_{n-1}^{A_{n-3}|B_{n-3}} \mapsto r_1 \Theta_n^{A_1|B_1} + \cdots + r_{n-3} \Theta_n^{A_{n-3}|B_{n-3}}.$$

Different orthants of BHV_{n-1} are represented by different collections of compatible splits, and different cones of $\mathcal{F}(D_n)$ are represented by different collections of root subsystems. The invertibility of the map Ω follows from that of ω, and, therefore, Ω is an isomorphism between $\mathcal{F}(D_n)$ and BHV_{n-1} as sets. □

The cubical complex metric is similar to the l^1-path metric on a simplicial complex:

Definition 2.4.9 (The l^1-path metric of a cubical complex)**.** Let C be a cubical complex. For any two points $x, y \in C$, we define a path in C from x to y to be a sequence of points $x = a_0, a_1, \ldots, a_r = y$ in C, such that for every $i = 1, \ldots, r$ there is a cube σ_i of C that contains both a_{i-1} and a_i. The length of such a path is the sum

$$\sum_{i=1}^{r} d_{\sigma_i}(a_{i-1}, a_i),$$

where d_{σ_i} denotes the metric on the cube σ_i, i.e., the Euclidean metric. Then we define a metric on C by setting $d_C(x, y)$ to be the infimum of lengths of all such paths from x to y. We call this metric the l^1-path metric of a cubical complex.

Ardila-Owen-Sullivant [3] found that BHV is an orthant space, which is a special case of cubical complexes, so BHV is naturally equipped with the cubical complex metric. By Definition 2.4.6, $\mathcal{F}(D_n)$ is a collection of strongly convex rational polyhedral cones determined by the rays $\psi(\Theta_n)$ for each $\Theta \in \mathcal{R}(\Delta_n)$, where rays span a cone in $\mathcal{F}(D_n)$ if and only if the corresponding subsystems form a simplex in $\mathcal{R}(\Delta_n)$. The correspondence of $\mathcal{F}(D_n)$ with simplices in $\mathcal{R}(\Delta_n)$ formed by subsystems imposes a simplicial structure on $\mathcal{F}(D_n)$.

Lemma 2.4.10. $\mathcal{F}(D_n)$ is a cubical complex.

Proof. $\mathcal{F}(D_n)$ is a collection of strongly convex rational polyhedral cones

$$\tau_{\mathcal{F}}^k = \left\{ r_1 \psi\left(\Theta_n^{A_1|B_1}\right) + \cdots + r_k \psi\left(\Theta_n^{A_k|B_k}\right) \in N(D_n)_{\mathbb{Q}} : r_i \geq 0 \right\}.$$

$\mathcal{F}(D_n)$ is obtained by tiling with unit cubes, and is the quotient of a disjoint union of cubes $X = \coprod_\Lambda I^{n_\lambda}$ by an equivalence relation $\sim$. Consider the restrictions $p_\lambda : I^{n_\lambda} \to \mathcal{F}(D_n)$ of the natural projection $p : X \to \mathcal{F}(D_n) = X/\sim$. Because $\mathcal{F}(D_n)$ is obtained by tiling with unit cubes, no cube is glued to itself. Thus, the map p_λ is injective for every $\lambda \in \Lambda$.

When gluing cones together, rays span a cone in $\mathcal{F}(D_n)$ if and only if the corresponding subsystems form a simplex in $\mathcal{R}(\Delta_n)$, and two cones are glued together along the face spanned by the root subsystems they share. As a resut, for every cone σ_λ, the map

$$p_\lambda : \coprod_\Lambda \sigma_\lambda \to \coprod_\Lambda \sigma_\lambda / \sim$$

with the equivalence relation represents the identification of two faces spanned by the same root subsystems. Faces of tiling unit cubes in $\mathcal{F}(D_n)$ are glued via an isometry $h_{\lambda,\lambda'}$ from a face $T_\lambda \subset I^{n_\lambda}$ onto a face $T_{\lambda'} \subset I^{n_{\lambda'}}$, such that $p_\lambda(x) = p_{\lambda'}(x')$ if and only if $x' = h_{\lambda,\lambda'}(x)$. Therefore, $\mathcal{F}(D_n)$ is a cubical complex.

□

$\mathcal{F}(D_n)$ is a cubical complex, so the l^1-path metric on a cubical complex is canonical on $\mathcal{F}(D_n)$. Points in BHV_{n-1} and $\mathcal{F}(D_n)$ are isomorphic by Theorem 2.4.8; therefore, the cubical complex metric induces an isometry between BHV_{n-1} and $\mathcal{F}(D_n)$:

Theorem 2.4.11 (BHV_{n-1} and $\mathcal{F}(D_n)$ isometry)**.** BHV_{n-1} and $\mathcal{F}(D_n)$ are isometric with the cubical complex metric.

Proof. Denote by

$$\mathcal{S}_{n-1} = \left\{ S_{n-1}^{A_1|B_1}, \ldots, S_{n-1}^{A_{n-3}|B_{n-3}} \right\}$$

a set of compatible splits that spans an orthant in BHV_{n-1}, and denote by

$$\Theta_n = \{\Theta_n^{A_1|B_1}, \ldots, \Theta_n^{A_{n-3}|B_{n-3}}\} \subset \mathcal{R}(\Delta_n)$$

a set of root subsystems that spans a maximal dimensional cone in $\mathcal{F}(D_n)$. Theorem 2.4.8 proved that the map

$$\Omega : \mathrm{BHV}_{n-1} \longrightarrow \mathcal{F}(D_n),$$
$$r_1 S_{n-1}^{A_1|B_1} + \cdots + r_{n-3} S_{n-1}^{A_{n-3}|B_{n-3}} \mapsto r_1 \Theta_n^{A_1|B_1} + \cdots + r_{n-3} \Theta_n^{A_{n-3}|B_{n-3}}$$

is a bijection between BHV_{n-1} and $\mathcal{F}(D_n)$. We show that the cubical complexes BHV_{n-1} and $\mathcal{F}(D_n)$ are glued in the same way, so that the distance between two points $T_{n-1}, \tilde{T}_{n-1} \in \mathrm{BHV}_{n-1}$ is equal to the distance between the images of these two points in $\mathcal{F}(D_n)$ under the map Ω. By definition, rays identified with $\Theta_n^{A_i|B_i}$ span a cone in $\mathcal{F}(D_n)$ if and only if the root subsystems that correspond to the rays form a simplex in $\mathcal{R}(\Delta_n)$, which implies that the splits corresponding to the root subsystems span an orthant; and it is the same for the degenerated orthants and the faces of cone. Therefore, cubes in the cubical complex BHV_{n-1} and $\mathcal{F}(D_n)$ are bijective.

Now we show that these identified cubes in BHV_{n-1} and $\mathcal{F}(D_n)$ are glued in the same way. Since the intersection of Θ_n and $\tilde{\Theta}_n$ is a subset of each, each pair of elements in $\Theta_n \cap \tilde{\Theta}_n$ also satisfies this criteria; hence, the intersection spans a cone in $\mathcal{F}(D_n)$. Consider two orthants spanned by two sets of compatible splits $\mathcal{S}_{n-1}$ and $\tilde{\mathcal{S}}_{n-1}$ corresponding to Θ_n and $\tilde{\Theta}_n$. Since the intersection of two sets of compatible splits is also a set of compatible splits, $\mathcal{S}_{n-1} \cap \tilde{\mathcal{S}}_{n-1}$ spans a degenerated orthant that corresponds to the face of a cone in $\mathcal{F}(D_n)$. Hence, the disjoint union of cubes in $\mathcal{F}(D_n)$ and BHV_{n-1} quotient in the same way.

Now we have proved that $\mathcal{F}(D_n)$ and BHV_{n-1} consist of isomorphic cubes equipped with the l^1-path metric of a cubical complex and that they are glued in the same way. Therefore, for any two points $T_{n-1}, \tilde{T}_{n-1} \in \mathrm{BHV}_{n-1}$ and their images in $\mathcal{F}(D_n)$ under the map Ω, the paths between T_{n-1} and $\tilde{T}_{n-1}$ and the paths between $\Omega(T_{n-1})$ and $\Omega(\tilde{T}_{n-1})$ can be identified. In particular, the length of the shortest path between T_{n-1} and $\tilde{T}_{n-1}$ and that between $\Omega(T_{n-1})$ and $\Omega(\tilde{T}_{n-1})$ are equal. Thus,

$$\|T_{n-1} - \tilde{T}_{n-1}\|_{d_{\mathrm{BHV}}} = \|\Omega(T_{n-1}) - \Omega(\tilde{T}_{n-1})\|_{d_{\mathcal{F}}}.$$

$\square$

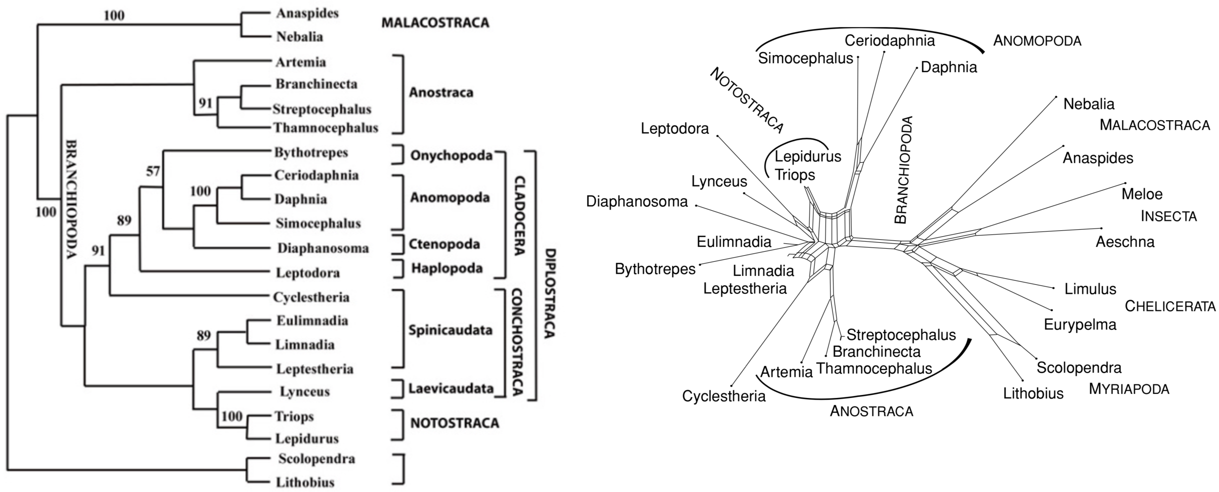
The following section discusses generalizations of spaces of phylogenetic trees, called spaces of phylogenetic networks, and introduces comparison theorems that connect spaces of phylogenetic networks with algebraic geometry.

Chapter 3

Comparison Theorems on Spaces of Phylogenetic Networks

Phylogenetic networks are unrooted graphs, such that every split is represented by an array of parallel edges, where the internal nodes in the phylogenetic network represent ancestral species, and nodes with more than two parents correspond to reticulate events such as hybridization or recombination [29]. Compared to phylogenetic trees, phylogenetic networks more accurately reflect reticulate events such as hybridization, horizontal gene transfer, recombination, or gene duplication and loss [9, 49, 17, 32, 34, 33, 31]. Fig. 3.1 compares phylogenetic trees with phylogenetic networks for the same dataset by Wägele and Mayer [54] to demonstrate how the network exhibits conflicting signals that are not represented in the tree.

A set of compatible splits can be represented by an unrooted phylogenetic tree; an example is shown in Fig. 3.1a. However, in natural selection, even assuming that gene conversion between paralogous genes and horizontal transfer of genes between species are rare, non-phylogenetic signals still cannot always be recognized. Therefore, random data [56] and the presence of paralogous sequences introduce noise in the form of conflicting signals. For example, *Cyclestheria* appears as sister taxon to Cladocera in the tree, as shown in Fig. 3.1a; however, *Cyclestheria* is usually classified as genus of Spinicaudata, while there are also morphological characters indicating that *Cyclestheria* might not be a spinicaudatan genus. The bifurcations are contradicted by conflicting patterns, and there are no edges distinctly better than conflicting signals that allow safe placement of Notostraca, Spinicaudata, *Lynceus,* or *Cyclestheria.* The phylogenetic network exhibits conflicting signals that are not represented in the tree, as shown in Fig. 3.1b.



(a) **Original tree topology [54].** (b) **Neighbor-net graph [54].**

Figure 3.1: **Comparison of phylogenetic trees with phylogenetic networks for the same dataset.** 25 species of Branchiopoda and outgroups from 18SrDNA sequences.

3.1 Spaces of Phylogenetic Networks and Projectivized Spaces

In this section, we introduce spaces of phylogenetic networks with a brief introduction from the viewpoint of biologists [29], followed by the mathematical model of phylogenetic networks as cubical complexes according to Devadoss and Petti, called the circular split network (CSN) [16]. Then we will discuss relations between spaces of phylogenetic networks and algebraic fans, projectivized spaces of phylogenetic networks and simplicial complexes, and embeddings among these spaces.

Definition 3.1.1 (Edge coloring)**.** Let $G = (V, E)$ be a finite connected graph, and let K denote a finite set of labels called colors. An *edge coloring* is a mapping $\sigma : E \to K$ that has the property that no two adjacent edges are given the same color [29].

Definition 3.1.2 (Properly colored)**.** Consider an undirected path,

$$P = (u_0, e_1 = \{u_0, u_1\}, u_1, e_2 = \{u_1, u_2\}, \ldots, e_t = \{u_{t-1}, u_t\}, u_t),$$

containing exactly $\mathrm{len}(P) = t$ different edges. We denote the set of colors that occur in P by $\sigma(P)$, defined as

$$\sigma(P) = \{\sigma(e_1), \ldots, \sigma(e_t)\},$$

and call P *properly colored,* if all edges in P have different colors, that is, if $|\sigma(P)| = t$ [30].

Definition 3.1.3 (Isometric coloring)**.** An edge coloring σ of a finite connected graph $G = (V, E)$ is an *isometric coloring,* if for any two nodes, all shortest paths between them are properly colored and use exactly the same set of colors [30].

Definition 3.1.4 (Split graph)**.** A *split graph* consists of a finite, simple, connected, bipartite graph $G = (V, E)$, together with an edge coloring $\sigma : E \to K$ that is surjective and isometric [29].

Deleting all edges of any given color in a split graph produces two connected components, initially proved by Dress and Huson [18]. This property ensures split networks are well-defined, with nodes labeled by a set of taxa $\mathcal{X}$ and edges labeled by a set of splits $\mathcal{S}$.

Definition 3.1.5 (Split network)**.** Let $\mathcal{S}$ be a set of splits on $\mathcal{X}$. A *split network* $N = (V, E, \sigma, \lambda)$ that represents $\mathcal{S}$ is given by a split graph $(G = (V, E), \sigma : E \to \mathcal{S})$ and a node labeling $\lambda : \mathcal{X} \to V$, such that

$$A = \bigcup_{v \in V_S^0} \lambda^{-1}(v) \text{ and } B = \bigcup_{v \in V_S^1} \lambda^{-1}(v)$$

for every split $S = A \mid B$ in $\mathcal{S}$, where V_S^0 and V_S^1 refer to the sets of vertices in the connected component G_c^0 and G_c^1, respectively [29].

In other words, deletion of all edges of color S produces a graph consisting of precisely two connected components, one containing all nodes labeled with elements of A and the other containing all nodes labeled with elements of B. Phylogenetic network refers to any graph used to represent evolutionary relationships between a set of taxa [30]. More recently, Devadoss and Petti presented a mathematical model of phylogenetic networks, called a circular split network, where a split network is specified with a circular ordering that is compatible with most of the splits [16], given in Definition 1.3.1. Orthants in CSN_n are glued together along cells that represent split systems that are compatible with the orderings of their respective chambers of $\overline{\mathcal{M}}_{0,n}$ [16].

CSN_n is a cubical complex realized by subdividing each orthant into cubes. BHV_{n-1} consists of one root and a set $\mathcal{X}$ of $n-1$ labeled leaves, and CSN_n consists of a set $\mathcal{X}' = \mathcal{X} \cup \{v\}$ with some labeled leaf v. CSN_n and BHV_{n-1} share the same set of bipartitions after identifying with the root in BHV_{n-1}. The set of bipartitions has cardinality $2^{n-1}-n-1$, and we use $\mathring{\mathfrak{S}}_n$ to denote the set of splits in CSN_n. Each phylogenetic network $P_n \in \mathrm{CSN}_n$ consists of at most $M = \frac{n(n-3)}{2}$ splits; therefore, we represent each network $P_n \in \mathrm{CSN}_n$ by a coordinate system

$$\mathring{\mathcal{S}}_n = \left\{ \dot{S}_n^{A_1|B_1}, \ldots, \dot{S}_n^{A_M|B_M} \right\}$$

for a set of splits $\mathring{\mathcal{S}}_n \subset \mathring{\mathfrak{S}}_n$, such that

$$P_n = \left(t_1 \dot{S}_n^{A_1|B_1}, \ldots, t_M \dot{S}_n^{A_M|B_M} \right)$$

for some $t_i \geq 0$.

Definition 3.1.6 (CSN metric)**.** The *distance* between any two points of CSN_n in the same orthant is the Euclidean distance. If two points are in different orthants, we join them by a sequence of straight segments, with each segment lying in a single orthant. The distance between the two points is the minimum of the lengths of such segmented paths joining the two points, denoted as $d_{\mathrm{CSN}_n}(x, y)$.

Although BHV_{n-1} and CSN_n share the same set of bipartitions, the bipartitions span a phylogenetic tree or a phylogenetic network under different criteria, resulting in different maximal dimensional cells in BHV_{n-1} and CSN_n. In CSN_n, bipartitions form higher-dimensional simplices if they share the same circular ordering; in BHV_{n-1}, bipartitions are required to be compatible, in addition to sharing the same circular ordering, in order to form higher-dimensional simplices. This forms a natural inclusion from BHV_{n-1} into CSN_n. In the following, we first present a definition of the Catalan number C_n, which can be regarded as the number of non-isomorphic ordered trees with $n+1$ vertices [51]. Then we will discuss the combinatorial properties of the embedding of BHV_{n-1} into CSN_n.

Definition 3.1.7 (Catalan number)**.** The *Catalan numbers* are named after the Belgian mathematician Eugène Charles Catalan. The n-th Catalan number is

$$C_n = \frac{1}{n+1}\binom{2n}{n} = \frac{(2n)!}{(n+1)!\,n!} = \prod_{k=2}^{n} \frac{n+k}{k} \qquad \text{for } n \geq 0.$$

Lemma 3.1.8 (Factorial identity)**.**

$$\frac{(2n-4)\cdot\cdots\cdot(n-1)}{2(2n-5)!!}=2^{n-3}.$$

Proof. The proof is by induction. $\square$

Theorem 3.1.9 (BHV_{n-1} embedding into CSN_n)**.** BHV_{n-1} is embedded into CSN_n. There are C_{n-2} types of tree topology in BHV_{n-1} embedded in each orthant of CSN_n, where C_{n-2} is the Catalan number. Each type of tree topology in BHV_{n-1} is embedded in the intersection of 2^{n-3} orthants of CSN_n.

Proof. Identifying the root of BHV_{n-1} with $v=\mathfrak{X}'\setminus\mathfrak{X}$, which is the difference between the set of taxa in CSN_n and BHV_{n-1}, induces the identity map λ between the set of bipartitions $\mathfrak{S}_{n-1}$ of BHV_{n-1} and $\mathring{\mathfrak{S}}_n$ of CSN_n with $A_i=A_i'$, $A_i=B_i'$, $A_i'=B_i$, or $A_i'=B_i'$:

$$\begin{aligned}\lambda:\mathfrak{S}_{n-1}&\longrightarrow\mathring{\mathfrak{S}}_n,\\ S_{n-1}^{A_i|B_i}&\mapsto\dot{S}_n^{A_i'|B_i'}.\end{aligned}$$

Extend λ to Λ as the following:

$$\begin{gathered}\Lambda:\mathrm{BHV}_{n-1}\longrightarrow\mathrm{CSN}_n,\\ \left(r_1S_{n-1}^{A_1|B_1},\cdots,r_{n-3}S_{n-1}^{A_{n-3}|B_{n-3}}\right)\mapsto\left(r_{i_1}S_n^{A_{i_1}|B_{i_1}},\cdots,r_{i_M}S_n^{A_{i_M}|B_{i_M}}\right),\end{gathered}$$

where $M=\frac{n(n-3)}{2}$, $r_i=r_j$ if $A_i=A_j$, $A_i=B_j$, $B_i=A_j$, or $B_i=B_j$, and $r_k=0$ otherwise. Since Λ is injective, BHV_{n-1} is embedded into CSN_n.

The number of non-isomorphic ordered trees with $n+1$ vertices is C_n [51], and each non-isomorphic tree with certain circular ordering corresponds to a tree topology; therefore, there are C_{n-2} types of tree topology in BHV_{n-1} embedded in each orthant of CSN_n. Since each tree in BHV_{n-1} is up to rotation along the internal edges, but each rotation leads to a different orthant in CSN_n, each tree topology is in the intersection of 2^{n-3} different CSN_n orthants, where $n-3$ is the number of internal edges for a tree in the interior of BHV_{n-1}. $\square$

The number of CSN_n orthants that each BHV_{n-1} orthant intersects is the number of tree topologies in each CSN_n orthant C_{n-2}, multiplied by the total number of orthants $\frac{(n-1)!}{2}$ in CSN_n, and then divided by the number of tree topologies in BHV_{n-1}:

$$\frac{C_{n-2} \cdot \frac{(n-1)!}{2}}{(2n-5)!!} = \frac{(2n-4) \cdot \dots \cdot (n-1)}{2(2n-5)!!} = 2^{n-3},$$

which yields the same result as Theorem 3.1.9. The last identity is according to Lemma 3.1.8.

Following the definition of the link of vertex by Zairis-Khiabanian-Blumberg-Rabadan, we define the link of CSN_n at the origin as the set

$$\mathcal{L}_{\mathrm{CSN}}(0) = \{z \in \mathrm{CSN}_n \mid d(z, 0) = \epsilon\}$$

for some fixed $0 < \epsilon < 1$. Therefore, the link of the origin in CSN_n is the set of points $P_n \in \mathrm{CSN}_n$, such that $\sum_{i=1}^{n(n-3)/2} t_i^2 = \epsilon^2$. By Proposition 2.1.11, the link $\mathcal{L}(v)$ of a vertex v in a cubical complex C is a spherical complex. Since CSN_n is a cubical complex, as a result, the link of the origin in CSN_n is a spherical complex.

Proposition 3.1.10 ($\mathbb{P}\,\mathrm{CSN}_n$ and the link of CSN_n isomorphism)**.** There is an isomorphism of sets between $\mathbb{P}\,\mathrm{CSN}_n$ and the link of CSN_n at the origin defined by Zairis-Khiabanian-Blumberg-Rabadan.

Proof. By Definition 1.3.2, $\mathbb{P}\,\mathrm{CSN}_n$ consists of the points with the sum of internal edges t_i satisfying $\sum_{i=1}^{n(n-3)/2} t_i = 1$. Parameterizing this set of points by angular coordinates in each orthant, we have:

$$\mu : \mathbb{P}\,\mathrm{CSN}_n \longrightarrow \mathcal{L}_{\mathrm{CSN}}(0),$$
$$\left(l, \varphi_1, \dots, \varphi_{\frac{n(n-3)}{2}-1}\right) \mapsto \left(\epsilon, \varphi_1, \dots, \varphi_{\frac{n(n-3)}{2}-1}\right),$$

where φ_k are angular coordinates, and ϵ is determined by the relation that the sum of the squares of the coordinates is equal to one. Therefore, $\mathcal{L}_{\mathrm{CSN}}(0)$ and $\mathbb{P}\,\mathrm{CSN}_n$ are isomorphic. □

Theorem 3.1.11 (CSN_n and $\mathbb{P}\,\mathrm{CSN}_n$ connection theorem)**.** $\mathbb{P}\,\mathrm{CSN}_n$ is the projectivized space of CSN_n, and the cone over $\mathbb{P}\,\mathrm{CSN}_n$ is isomorphic to CSN_n as sets.

Proof. Let $M = \frac{n(n-3)}{2}$. Denote a point P_n of CSN_n by a set of compatible splits $\overset{\circ}{\mathcal{S}}_n$:

$$P_n = \left(t_1 \dot{S}_n^{A_1|B_1}, \cdots, t_M \dot{S}_n^{A_M|B_M}\right), t_i \geq 0.$$

The image of the following quotient map is $\mathbb{P}\,\mathrm{CSN}_n$:

$$\zeta : \mathrm{CSN}_n \setminus \{0\} \longrightarrow \mathrm{CSN}_n \setminus \{0\},$$
$$\left(t_1 \dot{S}_n^{A_1|B_1}, \cdots, t_M \dot{S}_n^{A_M|B_M}\right) \mapsto \left(t_1 \dot{S}_n^{A_1|B_1}, \cdots, t_M \dot{S}_n^{A_M|B_M}\right) / \sum_{i=1}^{M} t_i.$$

Therefore, $\mathbb{P}\,\mathrm{CSN}_n$ is the projectivized space of CSN_n, with the simplicial structure of $\mathbb{P}\,\mathrm{CSN}_n$ inherited from the cubical complex CSN_n.

On the other hand, $\mathbb{P}\,\mathrm{CSN}_n$ is the quotient space of CSN_n, and we denote each element in $\mathbb{P}\,\mathrm{CSN}_n$ by

$$\overline{P}_n = \left(\hat{t}_1 \dot{S}_n^{A_1|B_1}, \cdots, \hat{t}_M \dot{S}_n^{A_M|B_M}\right)$$

with $\hat{t}_i \geq 0$ and $\sum_{i=1}^{M} \hat{t}_i = 1$. Each element in $\mathbb{P}\,\mathrm{CSN}_n$ is a representative of the equivalence class of rescaled phylogenetic networks in CSN_n. We define the 0-cone over $\mathbb{P}\,\mathrm{CSN}_n$ as

$$\mathrm{CSN}_n^{\star} = (\mathbb{P}\,\mathrm{CSN}_n \times \mathbb{R}_{\geq 0}) \,/\, (\mathbb{P}\,\mathrm{CSN}_n \times 0).$$

Each point in $\mathrm{CSN}_n^{\star}$ can be represented by

$$\overline{P}_n^k = \left(k, \hat{t}_1 \dot{S}_n^{A_1|B_1}, \cdots, \hat{t}_M \dot{S}_n^{A_M|B_M}\right)$$

for $k \geq 0$. Consider the map

$$\Xi : \mathrm{CSN}_n^{\star} \longrightarrow \mathrm{CSN}_n,$$
$$\left(k, \hat{t}_1 \dot{S}_n^{A_1|B_1}, \cdots, \hat{t}_M \dot{S}_n^{A_M|B_M}\right) \mapsto \left(k\hat{t}_1 \dot{S}_n^{A_1|B_1}, \cdots, k\hat{t}_M \dot{S}_n^{A_M|B_M}\right).$$

The map Ξ takes a point $\overline{P}_n^k$ in $\mathrm{CSN}_n^{\star}$ to a point P_n in CSN_n with a rescaling factor k. Therefore, the 0-cone over $\mathbb{P}\,\mathrm{CSN}_n$ is isomorphic to CSN_n as sets. □

Identifying the set of bipartitions of $\mathbb{P}\,\mathrm{BHV}_{n-1}$ and $\mathbb{P}\,\mathrm{CSN}_n$ also induces an inclusion from $\mathbb{P}\,\mathrm{BHV}_n$ into $\mathbb{P}\,\mathrm{CSN}_n$.

Corollary 3.1.12 ($\mathbb{P}\,\mathrm{BHV}_{n-1}$ embedding into $\mathbb{P}\,\mathrm{CSN}_n$)**.** $\mathbb{P}\,\mathrm{BHV}_{n-1}$ is embedded into $\mathbb{P}\,\mathrm{CSN}_n$. There are C_{n-2} types of tree topology in $\mathbb{P}\,\mathrm{BHV}_{n-1}$ embedded in each orthant of $\mathbb{P}\,\mathrm{CSN}_n$. Each type of tree topology in $\mathbb{P}\,\mathrm{BHV}_{n-1}$ is embedded in the intersection of 2^{n-3} orthants of $\mathbb{P}\,\mathrm{CSN}_n$.

Proof. According to Definition 1.1.2 and Definition 1.3.2, $\mathbb{P}\,\mathrm{BHV}_n$ consists of the points T_n for which the sum of its $n-2$ internal edges t_i is 1, and $\mathbb{P}\,\mathrm{CSN}_n$ consists of the circular split networks for which the sum of its internal edges t_i is 1. Since rescaling does not affect tree topology, each tree topology in BHV_{n-1} is embedded in an orthant of CSN_n if and only if that tree topology in $\mathbb{P}\,\mathrm{BHV}_{n-1}$ is embedded in the respective $\mathbb{P}\,\mathrm{CSN}_n$ orthant. Therefore, the results in Theorem 3.1.9 extend to $\mathbb{P}\,\mathrm{BHV}_{n-1}$ and $\mathbb{P}\,\mathrm{CSN}_n$. □

3.2 The Remarkable Fan $\mathcal{G}(D_n)$ Over Simplicial Complexes $\mathcal{Q}(\Delta_n)$

In this section, we introduce algebraic fans that are isomorphic to spaces of phylogenetic networks and inclusion relations between algebraic fans and phylogenetic spaces.

Definition 3.2.1 (The simplicial complex $\mathcal{Q}(\Delta_n)$)**.** Define the simplicial complex $\mathcal{Q}(\Delta_n)$ with 0-simplices being elements in the set $\mathcal{R}(D_n)$. Higher-dimensional simplices

$$\Theta_n^k = \left\{\Theta_n^{A_1|B_1}, \ldots, \Theta_n^{A_k|B_k}\right\} \subset \mathcal{Q}(\Delta_n)$$

are formed by root subsystems if and only if their corresponding boundary divisors all meet a single closed cell, i.e., they correspond to faces of $P_{\sigma_i}(D_n)$.

Definition 3.2.2 (The remarkable fan $\mathcal{G}(D_n)$)**.** Define strongly convex rational polyhedral cones $\tau_{\mathcal{G}}$ in $N_{\mathbb{R}}$ as a set:

$$\tau_{\mathcal{G}} = \left\{r_1\psi\left(\Theta_n^{A_1|B_1}\right) + \cdots + r_n\psi\left(\Theta_n^{A_k|B_k}\right) \in N(D_n)_{\mathbb{Q}} : r_i \geq 0\right\}.$$

The *remarkable fan* $\mathcal{G}(D_n)$ in $N(D_n)_{\mathbb{Q}}$ is a collection of strongly convex rational polyhedral cones $\tau_{\mathcal{G}}$ in the real vector space $N(D_n)_{\mathbb{Q}}$ determined by the rays $\psi\left(\Theta_n^{A_i|B_i}\right)$ for each $\Theta_n^{A_i|B_i} \in \Theta_n$, where rays span a cone in $\mathcal{G}(D_n)$ if and only if the corresponding boundary divisors all meet a single closed cell, i.e., they correspond to faces of $P_{\sigma_i}(D_n)$, as defined in Section 2.3.

The proof that $\mathcal{G}(D_n)$ is a fan is similar to Proposition 2.4.7.

Proposition 3.2.3. $\mathcal{G}(D_n)$ is a fan.

Proof. Consider a face τ of a cone $\sigma \subset \mathcal{G}(D_n)$, which is the intersection of σ with any supporting hyperplane:

$$\tau = \sigma \cap u^{\perp} = \{v \in \sigma : \langle u, v \rangle = 0\}$$

for some u in $\sigma^{\vee}$. Therefore, τ is a subset of σ and is spanned by a collection of root subsystems. Since τ is a subset of σ, the root subsystems spanning τ also correspond to faces of $P_{\sigma_i}(D_n)$. Therefore, τ is also a cone in the fan.

Now we show that the intersection of two cones in $\mathcal{G}(D_n)$ is a face of each. Let σ denote a cone in $\mathcal{G}(D_n)$ spanned by a collection of root subsystems $\mathcal{S}$, and let $\tilde{\sigma}$ denote another cone in $\mathcal{G}(D_n)$ spanned by a collection of root subsystems $\tilde{\mathcal{S}}$. Assuming $\sigma \cap \tilde{\sigma} \neq \emptyset$, the hyperplane u spanned by $\mathcal{S} \cap \tilde{\mathcal{S}}$ is a supporting hyperplane, such that

$$\tau = \sigma \cap u^{\perp} = \{v \in \sigma : \langle u, v \rangle = 0\}.$$

Therefore, τ is a face of σ. Similarly, the hyperplane $\tilde{u}$ spanned by $\mathcal{S} \cap \tilde{\mathcal{S}}$ is a supporting hyperplane, such that

$$\tau = \tilde{\sigma} \cap \tilde{u}^{\perp} = \{v \in \tilde{\sigma} : \langle \tilde{u}, v \rangle = 0\}.$$

Therefore, τ is also a face of $\tilde{\sigma}$ also. Hence, $\mathcal{F}(D_n)$ is a fan.

$\square$

By definition, the fan $\mathcal{F}(D_n) \subset N(D_n)_{\mathbb{Q}}$ is the collection of cones determined by the rays $\psi\left(\Theta_n^{A_i|B_i}\right)$ for each $\Theta_n^{A_i|B_i} \in \mathcal{R}(\Delta_n)$, where rays span a cone in $\mathcal{F}(D_n)$ if and only if the corresponding root subsystems form a simplex in $\mathcal{R}(\Delta_n)$. This implies that bipartitions that correspond to the root subsystems are compatible. For a collection of bipartitions to be compatible, they first need to share the same circular ordering. Since rays span a cone in $\mathcal{G}(D_n)$ if they share the same circular ordering, this fact induces an inclusion relation between $\mathcal{F}(D_n)$ and $\mathcal{G}(D_n)$. In the following, we present a proof of the conjecture by Hacking-Keel-Tevelev on the inclusion relation between $\mathcal{F}(D_n)$ and $\mathcal{G}(D_n)$ [25].

Theorem 3.2.4 ($\mathcal{G}(D_n)$ contains $\mathcal{F}(D_n)$)**.** The fan $\mathcal{F}(D_n) \subset N(D_n)_{\mathbb{Q}}$ embeds into the *remarkable fan* $\mathcal{G}(D_n)$.

Proof. By definition, the fan $\mathcal{F}(D_n) \subset N(D_n)_{\mathbb{Q}}$ is the collection of cones determined by the rays $\psi\left(\Theta_n^{A_i|B_i}\right)$ for each $\Theta_n^{A_i|B_i} \in \mathcal{R}(\Delta_n)$. Consider the map

$$\begin{aligned} \iota : \mathcal{F}(D_n) &\longrightarrow \mathcal{G}(D_n), \\ \sum_{i=1}^{n-3} r_i \psi\left(\Theta_n^{A_i|B_i}\right) &\mapsto \sum_{i=1}^{M} r_i \psi\left(\Theta_n^{A_i|B_i}\right) \end{aligned}$$

for $M = \frac{n(n-3)}{2}$. Since rays span a cone in $\mathcal{F}(D_n)$ if and only if the corresponding subsystems form a simplex in $\mathcal{R}(\Delta_n)$, by Theorem 2.3.1, this implies these rays form a simplex if and only if they are compatible bipartitions. Compatible bipartitions can be viewed as internal edges of a labeled tree, which has a fixed circular ordering. Therefore, compatible bipartitions can be arranged to share the same circular ordering. Rays form a cone $\tau_{\mathcal{G}}$ in the fan $\mathcal{G}$ if and only if they correspond to faces of $P_{\sigma_i}(D_n)$, i.e., they share the same circular ordering. Therefore, each cone $\tau_{\mathcal{F}}$ in the fan $\mathcal{F}$ is a subset of a cone $\tau_{\mathcal{G}}$ in the fan $\mathcal{G}$. Two cones in the fan $\mathcal{G}$ are identified along their common faces, which are also common faces in the fan $\mathcal{F}$; hence, they are identified in the preimage under ι as well. Thus, ι is an injection from $\tau_{\mathcal{F}}$ into $\tau_{\mathcal{G}}$, and the fan $\mathcal{F}(D_n) \subset N(D_n)_{\mathbb{Q}}$ embeds into the *remarkable fan* $\mathcal{G}(D_n)$. $\square$

3.3 Homeomorphism Between $\mathbb{P}\,\mathrm{CSN}_n$ and $\mathcal{Q}(\Delta_n)$

$\mathcal{Q}(\Delta_n)$ is a simplicial complex: for a collection of root subsystems Θ_n that forms a simplex, each subset of Θ_n must also share a circular ordering, i.e., they correspond to faces of $P_{\sigma_i}(D_n)$ for certain circular ordering σ_i. So every face of a simplex in $\mathcal{Q}(\Delta_n)$ is also a simplex in $\mathcal{Q}(\Delta_n)$. Also, a non-empty intersection of two simplices is a subset of each; therefore, it is a face of each.

Similar to $\mathbb{P}\,\mathrm{BHV}_{n-1}$ and $\mathcal{R}(\Delta_n)$, $\mathbb{P}\,\mathrm{CSN}_n$ and $\mathcal{Q}(\Delta_n)$ are also isomorphic as abstract simplicial complexes and their geometric realizations are homeomorphic.

Theorem 3.3.1 ($\mathbb{P}\,\mathrm{CSN}_n$ and $\mathcal{Q}(\Delta_n)$ isomorphism)**.** $\mathbb{P}\,\mathrm{CSN}_n$ and $\mathcal{Q}(\Delta_n)$ are isomorphic abstract simplicial complexes.

Proof. The set of splits in CSN_n are denoted by:

$$\mathring{\mathfrak{S}}_n = \left\{ \dot{S}_n^{A_1|B_1}, \ldots, \dot{S}_n^{A_M|B_M} \right\}$$

for $M = |\mathfrak{S}_{n-1}| = 2^n - n - 1$. Consider the map υ between $\mathring{\mathfrak{S}}_n \subset \mathbb{P}\,\mathrm{CSN}_n$ and $\mathcal{R}(D_n)$:

$$\begin{aligned} \upsilon : \mathring{\mathfrak{S}}_n &\longrightarrow \mathcal{R}(D_n), \\ \dot{S}_{n-1}^{A_i|B_i} &\mapsto \Theta_n^{A_i|B_i}. \end{aligned}$$

The map I identifies 0-simplices of $\mathbb{P}\,\mathrm{CSN}_n$ and $\mathcal{R}(D_n)$.

By Definition 1.3.1, the internal branches of split networks share the same circular ordering; simplices in $\mathcal{Q}(\Delta_n)$ are formed by root subsystems if and only if they correspond to faces of $P_{\sigma_i}(D_n)$ for a certain σ_i, which coincides with sharing the same circular ordering. Therefore, the 0-simplices of $\mathbb{P}\,\mathrm{CSN}_n$ and $\mathcal{Q}(\Delta_n)$ form higher-dimensional simplices in the same way. Thus, abstract complexes $\mathbb{P}\,\mathrm{CSN}_n$ and $\mathcal{Q}(\Delta_n)$ are isomorphic. $\square$

Since the induced simplicial map between two abstract complexes is a homeomorphism, $\mathbb{P}\,\mathrm{CSN}_n$ and the geometric realization of $\mathcal{Q}(\Delta_n)$ are also homeomorphic. As a result,

Corollary 3.3.2 ($\mathbb{P}\,\mathrm{CSN}_{n-1}$ and $\mathcal{Q}(\Delta_n)$ homeomorphism)**.** The geometric realization of $\mathbb{P}\,\mathrm{CSN}_n$ and $\mathcal{Q}(\Delta_n)$ are homeomorphic.

Because $\mathbb{P}\,\mathrm{BHV}_{n-1}$ is embedded into $\mathbb{P}\,\mathrm{CSN}_n$, the geometric realization of $\mathbb{P}\,\mathrm{BHV}_{n-1}$ and $\mathcal{R}(\Delta_n)$ are homeomorphic by Corollary 2.3.4, and the geometric realization of $\mathbb{P}\,\mathrm{CSN}_n$ and $\mathcal{Q}(\Delta_n)$ are homeomorphic by Corollary 3.3.2, we conclude that:

Corollary 3.3.3 ($\mathcal{R}(\Delta_n)$ embedding into $\mathcal{Q}(\Delta_n)$)**.** $\mathcal{R}(\Delta_n)$ is embedded into $\mathcal{Q}(\Delta_n)$. There are C_{n-2} maximal dimension simplices in $\mathcal{R}(\Delta_n)$ embedded in each maximal dimension simplex of $\mathcal{Q}(\Delta_n)$. Each maximal dimension simplex in $\mathcal{R}(\Delta_n)$ is embedded in the intersection of 2^{n-3} maximal dimension simplices of $\mathcal{Q}(\Delta_n)$.

3.4 Isometry Between CSN_n and $\mathcal{G}(D_n)$

The isomorphism between CSN_n and $\mathcal{G}(D_n)$ is extended from the isomorphism between the set of bipartitions of CSN_n and rays $\psi(\Theta)_{\mathcal{G}} \subset \mathcal{G}(D_n)$.

Lemma 3.4.1. $\mathcal{G}(D_n)$ is a cubical complex.

Proof. $\mathcal{G}(D_n)$ is a collection of strongly convex rational polyhedral cones

$$\sigma_{\mathcal{G}} = \left\{ r_1\psi\left(\Theta_n^{A_1|B_1}\right) + \cdots + r_M\psi\left(\Theta_M^{A_M}\right) \in N(D_n)_{\mathbb{Q}} : r_i \geq 0 \right\}, M = \frac{n(n-3)}{2}.$$

$\mathcal{G}(D_n)$ is obtained by tiling with unit cubes, and it is the quotient of a disjoint union of cubes $X = \coprod_{\Lambda} I^{n_\lambda}$ by an equivalence relation $\sim$. Consider the restrictions $p_\lambda : I^{n_\lambda} \to \mathcal{G}(D_n)$ of the natural projection $p : X \to \mathcal{G}(D_n) = X/\sim$. Because $\mathcal{G}(D_n)$ is obtained by tiling with unit cubes, for every $\lambda \in \Lambda$ the map p_λ is injective.

When gluing cones together, rays span a cone in $\mathcal{G}(D_n)$ if and only if the corresponding subsystems meet at a single closed cell. Two cones are glued together along the face spanned by the root subsystems they share, which is the isometry $h_{\lambda,\lambda'}$ from a face $T_\lambda \subset I^{n_\lambda}$ onto a face $T_{\lambda'} \subset I^{n_{\lambda'}}$, such that $p_\lambda(x) = p_{\lambda'}(x')$ if and only if $x' = h_{\lambda,\lambda'}(x)$. Therefore, $\mathcal{G}(D_n)$ is a cubical complex. □

Theorem 3.4.2 (CSN_n and $\mathcal{G}(D_n)$ isomorphism)**.** CSN_n and $\mathcal{G}(D_n)$ are isomorphic.

Proof. Each k-dimensional orthant in CSN_n can be parametrized by a set of splits $\mathring{\mathcal{S}}_n = \left\{ \dot{S}_n^{A_1|B_1}, \ldots, \dot{S}_n^{A_k|B_k} \right\}$ that shares the same circular ordering:

$$\sigma_n^k = \left\{ t_1\dot{S}_n^{A_1|B_1} + \cdots + t_k\dot{S}_n^{A_k|B_k} \in \mathrm{CSN}_n : t_i \geq 0 \right\}.$$

Each k-dimensional cone of the fan $\mathcal{G}(D_n)$ can be parametrized by

$$\tau_{\mathcal{G}}^k = \left\{ r_1\psi\left(\Theta_n^{A_1|B_1}\right) + \cdots + r_n\psi\left(\Theta_n^{A_k|B_k}\right) \in N(D_n)_{\mathbb{Q}} : r_i \geq 0 \right\}$$

for a collection of root subsystems $\Theta_n^k = \left\{ \Theta_n^{A_1|B_1}, \ldots, \Theta_n^{A_k|B_k} \right\}$ that are 0-simplices in $\mathcal{R}(\Delta_n)$. Therefore, there is an isomorphism between a k-dimensional orthant σ_n^k in CSN_n and a k-

dimensional cone $\tau_{\mathcal{G}}^k$ of $\mathcal{G}(D_n)$ induced by $\upsilon : \mathring{\mathfrak{S}}_n \longrightarrow \mathcal{R}(D_n)$ as the following:

$$\hat{\upsilon} : \sigma_n^k \longrightarrow \tau_{\mathcal{G}}^k$$
$$\sum_{i=1}^{k} r_i \dot{S}^{A_i|B_i} \mapsto \sum_{i=1}^{k} r_i \psi\left(\Theta_n^{A_i|B_i}\right).$$

But $\mathbb{P}\,\mathrm{CSN}_n$ and $\mathcal{Q}(\Delta_n)$ are isomorphic abstract simplicial complexes; therefore, orthants in $\mathbb{P}\,\mathrm{CSN}_n$ and cones in $\mathcal{Q}(\Delta_n)$ are identified in the same way. So $\hat{\upsilon}$ can be extended to CSN_n and $\mathcal{G}(D_n)$:

$$\Upsilon : \mathrm{CSN}_n \longrightarrow \mathcal{G}(D_n),$$
$$\sum_{i=1}^{n(n-3)/2} r_i \dot{S}^{A_i|B_i} \mapsto \sum_{i=1}^{n(n-3)/2} r_i \psi\left(\Theta_n^{A_i|B_i}\right).$$

The invertibility of the map $\hat{\Upsilon}$ follows from that of υ, and, therefore, Υ is an isomorphism between CSN_n and $\mathcal{G}(D_n)$ as sets. □

Any set of orthants in CSN_n is arranged around a common origin; therefore, CSN_n is an orthant space. Orthants in CSN_n are formed by splits sharing the same circular ordering, and they are glued together along the face spanned by splits they share. We equip CSN_n with cubical complex metric. $\mathcal{G}(D_n)$ is also a cubical complex, where rays span a cone in $\mathcal{G}(D_n)$ if and only if the corresponding subsystems form a simplex in $\mathcal{Q}(\Delta_n)$. Since CSN_n and $\mathcal{G}(D_n)$ are isomorphic, the cubical complex metric induces an isometry between CSN_n and $\mathcal{G}(D_n)$. The proof is as follows:

Theorem 3.4.3 (CSN_n and $\mathcal{G}(D_n)$ isometry)**.** CSN_n and $\mathcal{G}(D_n)$ are isometric cubical complexes.

Proof. We show that the cubical complexes CSN_n and $\mathcal{G}(D_n)$ are glued in the same way, so that the distance between two points $P_n, \tilde{P}_n \in \mathrm{CSN}_n$ is equal to the distance between the images of these two points in $\mathcal{G}(D_n)$ under the map Υ. By definition, rays identified with $\Theta_n^{A_i|B_i}$ span a cone in $\mathcal{G}(D_n)$ if and only if the corresponding subsystems form a simplex in $\mathcal{Q}(\Delta_n)$. This condition implies that they can be identified with a circular ordering. Meanwhile, a set of splits form a simplex in $\mathbb{P}\,\mathrm{CSN}_n$ and span an orthant in CSN_n when they correspond to faces of $P_{\sigma_i}(D_n)$ for some σ_i, which also implies that they can be identified with a circular ordering. Therefore, cubes in the cubical complex CSN_n and $\mathcal{G}(D_n)$ are bijective.

Cubes in CSN_n and $\mathcal{G}(D_n)$ are also identified in the same way. Consider two collections of rays Θ_n and $\tilde{\Theta}_n$ that span a cone in $\mathcal{G}(D_n)$. By definition, they correspond to faces of $P_{\sigma_i}(D_n)$ and $P_{\sigma_j}(D_n)$, and Θ_n and $\tilde{\Theta}_n$ are glued along common faces, i.e., the face spanned by root subsystems that are both contained in $P_{\sigma_i}(D_n)$ and $P_{\sigma_j}(D_n)$. On the other hand, orthants in CSN_n are glued along the face spanned by common splits in different circular orderings, which implies that the orthants in CSN_n and simplices in $\mathcal{G}(D_n)$ are identified in the same way. Therefore, for any two points $P_n, \tilde{P}_n \in \mathrm{CSN}_n$ and their images in $\mathcal{G}(D_n)$ under the map Υ, the paths between P_n and $\tilde{P}_n$ and the paths between $\Upsilon(P_n)$ and $\Upsilon(\tilde{P}_n)$ are identical, so the lengths of the shortest path are also equal. Thus,

$$\|P_n - \tilde{P}_n\|_{d_{\mathrm{CSN}}} = \|\Upsilon(P_n) - \Upsilon(\tilde{P}_n)\|_{d_{\mathcal{G}}}.$$

$\square$

Chapter 4

Summary

Phylogenetic techniques play an important role in the research of cancer evolution. The geometrical and topological structures of phylogenetic spaces unveil the intrinsic properties hidden in the ocean of data. The pathogenesis and evolutionary histories provide a quantitative characterization of tumor initiation and maintenance, and therefore, are valuable for understanding carcinogenesis and cancer metastasis. A classical approach to represent the evolutionary trajectories across patients uses phylogenetic trees of each tumor and then compares the n trees; another methodology developed much later is a generalization of phylogenetic trees, called phylogenetic networks [24]. Metric structures of phylogenetic spaces assist the identification of predictors of the malignancy evolution and cluster other genetic architecture. The geometric structure of phylogenetic spaces elucidates the differences between evolutionary histories, which may guide therapeutic intervention [11].

With recent developments in the acquisition of biological data and progress in genetics, biology has become a data-rich discipline; for example, biologists have wielded CRISPR to track a mammal's development from a single egg into an embryo with millions of cells [35], which creates an even greater demand for a deeper understanding of geometric structures that represent evolutionary histories. Developments in technology bring even more pressing challenges to phylogenetic studies. The phylogenetic trees are distinct within and between patients [14, 23], but clonal alternations are only partially ordered in most cases through phylogenetic analysis, yielding a complicated evolutionary model. Furthermore, genomic instability, interpatient variability, and technical noise introduce stochasticity and complexity to the evolutionary process.

In my dissertation, I studied evolutionary models and their mathematical structures. Modeling the evolutionary histories and identifying structures which arise in theoretical mathematics connect biology with mathematics. We discovered the space of phylogenetic trees BHV_{n-1} is isometric to the algebraic fan $\mathcal{F}(D_n)$, and the space of phylogenetic networks CSN_n is isometric to the algebraic fan $\mathcal{G}(D_n)$. Also, the projectivized space of phylogenetic trees $\mathbb{P}\,\mathrm{BHV}$ and the projectivized space of phylogenetic networks $\mathbb{P}\,\mathrm{CSN}$ are homeomorphic to the geometric realizations of simplicial complexes $\mathcal{R}(\Delta_n)$ and $\mathcal{Q}(D_n)$, respectively. Furthermore, $\mathcal{G}(D_n)$ contains $\mathcal{F}(D_n)$; BHV_{n-1} embeds into CSN_n; $\mathbb{P}\,\mathrm{BHV}_{n-1}$ embeds into $\mathbb{P}\,\mathrm{CSN}_n$; and $\mathcal{R}(\Delta_n)$ embeds into $\mathcal{Q}(D_n)$, as exhibited in Fig. 1.2.

Molecular events driving cancer evolution may deepen our understanding of the mechanisms of drug resistance and cancer metastasis. As longitudinal samplings of cohorts become available, the mathematical structures of genomic histories may enhance investigations into the molecular events underlying oncogenesis, metastasis, and acquired resistance to therapeutics. Thus, the evolutionary trajectories that modeled according to the spaces of phylogenetic trees BHV_n and phylogenetic networks CSN_n are promising quantitative characterizations for patient stratification, cancer progression, and therapeutic responses.

Appendices

Appendix A

Simplicial Complexes Formed by Root Subsystems D_n

A.1 Simplicial complexes $\mathcal{R}(\Delta_n)$ spanned by root subsystems of D_n

This section gives some explicit construction of simplicial complexes $\mathcal{R}(\Delta_n)$ spanned by root subsystems of D_n and their correspondences with spaces of phylogenetic trees. We consider the root system of type D of the form

$$D_n = \{\pm\epsilon_i \pm \epsilon_j \mid 1 \leq i < j \leq n\},$$

where $\epsilon_1, \ldots, \epsilon_n$ denote the usual orthonormal unit vectors which form a basis of $\mathbb{R}^n$. We denote *root subsystems* $\Theta_n^{A_k|B_k}$ of D_n as below:

$$\Theta_n^{A_k|B_k} = \{\pm\epsilon_i \pm \epsilon_j \mid i \neq j; i, j \in A_k\}, \ A_k \subset \{1, \ldots, n\}, \text{ and } B_k = \{1, \ldots, n\} \setminus A_k.$$

The simplicial complex $\mathcal{R}(\Delta_4)$. The root subsystems $\mathcal{D}_{2,4}$ of root system D_4 spanned by canonical basis ϵ_i, ϵ_j is represented as

$$\mathcal{D}_{2,4} = \left\{\Theta_4^{\{\epsilon_1,\epsilon_2\}|\{\epsilon_3,\epsilon_4\}}, \Theta_4^{\{\epsilon_1,\epsilon_3\}|\{\epsilon_2,\epsilon_4\}}, \Theta_4^{\{\epsilon_1,\epsilon_4\}|\{\epsilon_2,\epsilon_3\}}, \Theta_4^{\{\epsilon_2,\epsilon_3\}|\{\epsilon_1,\epsilon_4\}}, \Theta_4^{\{\epsilon_2,\epsilon_4\}|\{\epsilon_1,\epsilon_3\}}, \Theta_4^{\{\epsilon_3,\epsilon_4\}|\{\epsilon_1,\epsilon_2\}}\right\},$$

with

$$\begin{aligned}
\Theta_4^{\{\epsilon_1,\epsilon_2\}|\{\epsilon_3,\epsilon_4\}} &= \{\pm\epsilon_1 \pm \epsilon_2\}, \ \Theta_4^{\{\epsilon_1,\epsilon_3\}|\{\epsilon_2,\epsilon_4\}} = \{\pm\epsilon_1 \pm \epsilon_3\}, \\
\Theta_4^{\{\epsilon_1,\epsilon_4\}|\{\epsilon_2,\epsilon_3\}} &= \{\pm\epsilon_1 \pm \epsilon_4\}, \ \Theta_4^{\{\epsilon_2,\epsilon_3\}|\{\epsilon_1,\epsilon_4\}} = \{\pm\epsilon_2 \pm \epsilon_3\}, \\
\Theta_4^{\{\epsilon_2,\epsilon_4\}|\{\epsilon_1,\epsilon_3\}} &= \{\pm\epsilon_2 \pm \epsilon_4\}, \ \Theta_4^{\{\epsilon_3,\epsilon_4\}|\{\epsilon_1,\epsilon_2\}} = \{\pm\epsilon_3 \pm \epsilon_4\}.
\end{aligned}$$

There are $\binom{4}{2}/2 = 3$ types of decomposition of $\mathcal{D}_{2,4} \times \mathcal{D}_{2,4}$:

$$\begin{aligned}
\mathcal{R}(D_4) = \mathcal{D}_{2,4} \times \mathcal{D}_{2,4} = \ & \Big\{\Theta_4^{\{\epsilon_1,\epsilon_2\}|\{\epsilon_3,\epsilon_4\}} \times \Theta_4^{\{\epsilon_3,\epsilon_4\}|\{\epsilon_1,\epsilon_2\}}, \\
& \Theta_4^{\{\epsilon_1,\epsilon_3\}|\{\epsilon_2,\epsilon_4\}} \times \Theta_4^{\{\epsilon_2,\epsilon_4\}|\{\epsilon_1,\epsilon_3\}}, \\
& \Theta_4^{\{\epsilon_1,\epsilon_4\}|\{\epsilon_2,\epsilon_3\}} \times \Theta_4^{\{\epsilon_2,\epsilon_3\}|\{\epsilon_1,\epsilon_4\}}\Big\},
\end{aligned}$$

which are the 0-simplices of the simplicial complex $\mathcal{R}(\Delta_4)$. Since there is no perpendicular nor inclusion relation in these 0-simplices, they are the top-dimensional simplices of $\mathcal{R}(\Delta_4)$. Each element in $\mathcal{R}(D_4)$ corresponds to a split in BHV_3, and therefore $\mathcal{R}(D_4)$ is isomorphic to the space of phylogenetic trees BHV_3.

The simplicial complex $\mathcal{R}(\Delta_5)$. As a set, $\mathcal{R}(D_5)$ consists of root subsystems $\mathcal{D}_{2,5}$ of root system D_5, and there are $\binom{5}{2}$ different types of root subsystems $\mathcal{D}_{2,5}$ of root system D_5. Since A and B complement each other, we introduce a shorthand Θ_n^A to represent $\Theta_n^{A|B}$. By definition, we have:

$$\begin{aligned}\mathcal{R}(D_5) = \mathcal{D}_{2,5} = \Big\{ &\Theta_5^{\{\epsilon_1,\epsilon_2\}}, \Theta_5^{\{\epsilon_1,\epsilon_3\}}, \Theta_5^{\{\epsilon_1,\epsilon_4\}}, \Theta_5^{\{\epsilon_1,\epsilon_5\}}, \Theta_5^{\{\epsilon_2,\epsilon_3\}}, \\ &\Theta_5^{\{\epsilon_2,\epsilon_4\}}, \Theta_5^{\{\epsilon_2,\epsilon_5\}}, \Theta_5^{\{\epsilon_3,\epsilon_4\}}, \Theta_5^{\{\epsilon_3,\epsilon_5\}}, \Theta_5^{\{\epsilon_4,\epsilon_5\}} \Big\}.\end{aligned}$$

Since these root subsystems $\mathcal{D}_{2,5}$ are the 0-simplices of $\mathcal{R}(D_5)$, there is no inclusion relation among them, but only perpendicular relation. Denote $\left[\Theta_5^{A_i}, \Theta_5^{A_j}\right]^{\perp}$ as the 1-simplex formed by perpendicular relation between two root subsystems $\Theta_5^{A_i}, \Theta_5^{A_j} \in \mathcal{D}_{2,5}$; the 1-simplices of $\mathcal{R}(\Delta_5)$ are

$$\begin{aligned}\Big\{ &\left[\Theta_5^{\{\epsilon_1,\epsilon_2\}}, \Theta_5^{\{\epsilon_3,\epsilon_4\}}\right]^{\perp}, \left[\Theta_5^{\{\epsilon_1,\epsilon_2\}}, \Theta_5^{\{\epsilon_3,\epsilon_5\}}\right]^{\perp}, \left[\Theta_5^{\{\epsilon_1,\epsilon_2\}}, \Theta_5^{\{\epsilon_4,\epsilon_5\}}\right]^{\perp}, \\ &\left[\Theta_5^{\{\epsilon_1,\epsilon_3\}}, \Theta_5^{\{\epsilon_2,\epsilon_4\}}\right]^{\perp}, \left[\Theta_5^{\{\epsilon_1,\epsilon_3\}}, \Theta_5^{\{\epsilon_2,\epsilon_5\}}\right]^{\perp}, \left[\Theta_5^{\{\epsilon_1,\epsilon_3\}}, \Theta_5^{\{\epsilon_4,\epsilon_5\}}\right]^{\perp}, \\ &\left[\Theta_5^{\{\epsilon_1,\epsilon_4\}}, \Theta_5^{\{\epsilon_2,\epsilon_3\}}\right]^{\perp}, \left[\Theta_5^{\{\epsilon_1,\epsilon_4\}}, \Theta_5^{\{\epsilon_2,\epsilon_5\}}\right]^{\perp}, \left[\Theta_5^{\{\epsilon_1,\epsilon_4\}}, \Theta_5^{\{\epsilon_3,\epsilon_5\}}\right]^{\perp}, \\ &\left[\Theta_5^{\{\epsilon_1,\epsilon_5\}}, \Theta_5^{\{\epsilon_2,\epsilon_3\}}\right]^{\perp}, \left[\Theta_5^{\{\epsilon_1,\epsilon_5\}}, \Theta_5^{\{\epsilon_2,\epsilon_4\}}\right]^{\perp}, \left[\Theta_5^{\{\epsilon_1,\epsilon_5\}}, \Theta_5^{\{\epsilon_3,\epsilon_4\}}\right]^{\perp}, \\ &\left[\Theta_5^{\{\epsilon_2,\epsilon_3\}}, \Theta_5^{\{\epsilon_4,\epsilon_5\}}\right]^{\perp}, \left[\Theta_5^{\{\epsilon_2,\epsilon_4\}}, \Theta_5^{\{\epsilon_3,\epsilon_5\}}\right]^{\perp}, \left[\Theta_5^{\{\epsilon_2,\epsilon_5\}}, \Theta_5^{\{\epsilon_3,\epsilon_4\}}\right]^{\perp} \Big\}.\end{aligned}$$

These 1-simplices are top-dimensional, and the total number of 1-simplices is 15, which corresponds to 15 orthants in BHV_4.

The simplicial complex $\mathcal{R}(\Delta_6)$. By definition,

$$\mathcal{R}(D_6) = \mathcal{D}_{2,6} \sqcup (\mathcal{D}_{3,6} \times \mathcal{D}_{3,6}).$$

The cardinality of the set $\mathcal{D}_{2,6}$ in $\mathcal{R}(D_6)$ is $\binom{6}{2} = 15$, which corresponds to the following 0-simplices:

$$\mathcal{R}(D_5) = \mathcal{D}_{2,5} = \Big\{\Theta_6^{\{\epsilon_1,\epsilon_2\}}, \Theta_6^{\{\epsilon_1,\epsilon_3\}}, \Theta_6^{\{\epsilon_1,\epsilon_4\}}, \Theta_6^{\{\epsilon_1,\epsilon_5\}}, \Theta_6^{\{\epsilon_1,\epsilon_6\}}, \Theta_6^{\{\epsilon_2,\epsilon_3\}}, \Theta_6^{\{\epsilon_2,\epsilon_4\}}, \Theta_6^{\{\epsilon_2,\epsilon_5\}}, \Theta_6^{\{\epsilon_2,\epsilon_6\}}, \Theta_6^{\{\epsilon_3,\epsilon_4\}}, \Theta_6^{\{\epsilon_3,\epsilon_5\}}, \Theta_6^{\{\epsilon_3,\epsilon_6\}}, \Theta_6^{\{\epsilon_4,\epsilon_5\}}, \Theta_6^{\{\epsilon_4,\epsilon_6\}}, \Theta_6^{\{\epsilon_5,\epsilon_6\}}\Big\}.$$

The cardinality of the set $\mathcal{D}_{3,6} \times \mathcal{D}_{3,6}$ in $\mathcal{R}(D_6)$ is $\binom{6}{3}\big/\, 2 = 10$, which corresponds to the following 0-simplices:

$$\begin{aligned}\mathcal{D}_{3,6} \times \mathcal{D}_{3,6} = \Big\{&\Theta_6^{\{\epsilon_1,\epsilon_2,\epsilon_3\}} \times \Theta_6^{\{\epsilon_4,\epsilon_5,\epsilon_6\}}, \Theta_6^{\{\epsilon_1,\epsilon_2,\epsilon_4\}} \times \Theta_6^{\{\epsilon_3,\epsilon_5,\epsilon_6\}}, \Theta_6^{\{\epsilon_1,\epsilon_2,\epsilon_5\}} \times \Theta_6^{\{\epsilon_3,\epsilon_4,\epsilon_6\}},\\ &\Theta_6^{\{\epsilon_1,\epsilon_2,\epsilon_6\}} \times \Theta_6^{\{\epsilon_3,\epsilon_4,\epsilon_5\}}, \Theta_6^{\{\epsilon_1,\epsilon_3,\epsilon_4\}} \times \Theta_6^{\{\epsilon_2,\epsilon_5,\epsilon_6\}}, \Theta_6^{\{\epsilon_1,\epsilon_3,\epsilon_5\}} \times \Theta_6^{\{\epsilon_2,\epsilon_4,\epsilon_6\}},\\ &\Theta_6^{\{\epsilon_1,\epsilon_3,\epsilon_6\}} \times \Theta_6^{\{\epsilon_2,\epsilon_4,\epsilon_5\}}, \Theta_6^{\{\epsilon_1,\epsilon_4,\epsilon_5\}} \times \Theta_6^{\{\epsilon_2,\epsilon_3,\epsilon_6\}}, \Theta_6^{\{\epsilon_1,\epsilon_4,\epsilon_6\}} \times \Theta_6^{\{\epsilon_2,\epsilon_3,\epsilon_5\}},\\ &\Theta_6^{\{\epsilon_1,\epsilon_5,\epsilon_6\}} \times \Theta_6^{\{\epsilon_2,\epsilon_3,\epsilon_4\}}\Big\}.\end{aligned}$$

Therefore, the number of 0-simplices of $\mathcal{R}(D_6)$ is 25. The number of 1-simplices formed by perpendicular relations among elements in $\mathcal{D}_{2,6}$ is

$$\binom{6}{2}\binom{4}{2}\Big/\, 2 = 45.$$

The number of 1-simplices formed by inclusion relations from $\mathcal{D}_{2,6}$ in $\mathcal{D}_{3,6} \times \mathcal{D}_{3,6}$ is

$$\binom{6}{2}\binom{4}{3} = 60.$$

Therefore, the total number of 1-simplices in $\mathcal{R}(\Delta_6)$ is 105, which corresponds to degenerated trees in BHV_5. The 2-simplices in $\mathcal{R}(\Delta_6)$ can be formed by three elements in $\mathcal{D}_{2,6}$ with disjoint basis, totaling

$$\binom{6}{2}\binom{4}{2}\binom{2}{2}\Big/\, 3! = 15$$

2-simplices. The set $\mathcal{D}_{3,6} \times \mathcal{D}_{3,6}$ has 10 elements; the other way to form 2-simplices in $\mathcal{R}(\Delta_6)$ is to choose two elements from $\mathcal{D}_{2,6}$ so that each element is included in one side of $\mathcal{D}_{3,6} \times \mathcal{D}_{3,6}$, totaling 90 types of 2-simplices. Therefore, the number of 2-simplices of $\mathcal{R}(\Delta_6)$, which are the top-dimensional simplices of $\mathcal{R}(\Delta_6)$, is 105. These simplices correspond to 105 orthants in BHV_5.

A.2 Simplicial complexes $\mathcal{Q}(\Delta_n)$ spanned by root subsystems of D_n

The simplicial complex $\mathcal{Q}(\Delta_4)$. The root subsystems $\mathcal{D}_{2,4}$ of root system D_4 spanned by canonical basis ϵ_i, ϵ_j is represented as

$$\mathcal{D}_{2,4} = \left\{\Theta_4^{\{\epsilon_1,\epsilon_2\}|\{\epsilon_3,\epsilon_4\}}, \Theta_4^{\{\epsilon_1,\epsilon_3\}|\{\epsilon_2,\epsilon_4\}}, \Theta_4^{\{\epsilon_1,\epsilon_4\}|\{\epsilon_2,\epsilon_3\}}, \Theta_4^{\{\epsilon_2,\epsilon_3\}|\{\epsilon_1,\epsilon_4\}}, \Theta_4^{\{\epsilon_2,\epsilon_4\}|\{\epsilon_1,\epsilon_3\}}, \Theta_4^{\{\epsilon_3,\epsilon_4\}|\{\epsilon_1,\epsilon_2\}}\right\}.$$

The set $\mathcal{R}(D_4)$ has $\binom{4}{2}/2 = 3$ elements:

$$\begin{aligned}\mathcal{R}(D_4) = \mathcal{D}_{2,4} \times \mathcal{D}_{2,4} = \; & \Big\{\Theta_4^{\{\epsilon_1,\epsilon_2\}|\{\epsilon_3,\epsilon_4\}} \times \Theta_4^{\{\epsilon_3,\epsilon_4\}|\{\epsilon_1,\epsilon_2\}}, \\ & \Theta_4^{\{\epsilon_1,\epsilon_3\}|\{\epsilon_2,\epsilon_4\}} \times \Theta_4^{\{\epsilon_2,\epsilon_4\}|\{\epsilon_1,\epsilon_3\}}, \\ & \Theta_4^{\{\epsilon_1,\epsilon_4\}|\{\epsilon_2,\epsilon_3\}} \times \Theta_4^{\{\epsilon_2,\epsilon_3\}|\{\epsilon_1,\epsilon_4\}}\Big\},\end{aligned}$$

which are the 0-simplices of the simplicial complex $\mathcal{Q}(\Delta_4)$. The 1-simplices of $\mathcal{Q}(\Delta_4)$ are formed by root subsystems in $\mathcal{R}(D_4)$ that share the same circular ordering, and they are:

$$\left[\Theta_4^{\{\epsilon_1,\epsilon_2\}|\{\epsilon_3,\epsilon_4\}} \times \Theta_4^{\{\epsilon_3,\epsilon_4\}|\{\epsilon_1,\epsilon_2\}}, \Theta_4^{\{\epsilon_1,\epsilon_3\}|\{\epsilon_2,\epsilon_4\}} \times \Theta_4^{\{\epsilon_2,\epsilon_4\}|\{\epsilon_1,\epsilon_3\}}\right],$$

$$\left[\Theta_4^{\{\epsilon_1,\epsilon_2\}|\{\epsilon_3,\epsilon_4\}} \times \Theta_4^{\{\epsilon_3,\epsilon_4\}|\{\epsilon_1,\epsilon_2\}}, \Theta_4^{\{\epsilon_1,\epsilon_4\}|\{\epsilon_2,\epsilon_3\}} \times \Theta_4^{\{\epsilon_2,\epsilon_3\}|\{\epsilon_1,\epsilon_4\}}\right],$$

$$\left[\Theta_4^{\{\epsilon_1,\epsilon_3\}|\{\epsilon_2,\epsilon_4\}} \times \Theta_4^{\{\epsilon_2,\epsilon_4\}|\{\epsilon_1,\epsilon_3\}}, \Theta_4^{\{\epsilon_1,\epsilon_4\}|\{\epsilon_2,\epsilon_3\}} \times \Theta_4^{\{\epsilon_2,\epsilon_3\}|\{\epsilon_1,\epsilon_4\}}\right].$$

Each orthant in $\mathcal{Q}(D_4)$ has dimension $n(n-3)/2 = 2$, and there are $(n-1)!/2 = 3$ orthants, as illustrated in Fig. 1.1b.

The simplicial complex $\mathcal{Q}(\Delta_5)$. As a set, $\mathcal{R}(D_5)$ consists of root subsystems D_2 of root system D_5, and there are $\binom{5}{2}$ different types of root subsystems $\mathcal{D}_{2,5}$ of D_5. We continue to use the shorthand Θ_n^A to represent $\Theta_n^{A|B}$:

$$\begin{aligned}\mathcal{R}(D_5) = \mathcal{D}_{2,5} = \Big\{&\Theta_5^{\{\epsilon_1,\epsilon_2\}}, \Theta_5^{\{\epsilon_1,\epsilon_3\}}, \Theta_5^{\{\epsilon_1,\epsilon_4\}}, \Theta_5^{\{\epsilon_1,\epsilon_5\}}, \Theta_5^{\{\epsilon_2,\epsilon_3\}}, \\ &\Theta_5^{\{\epsilon_2,\epsilon_4\}}, \Theta_5^{\{\epsilon_2,\epsilon_5\}}, \Theta_5^{\{\epsilon_3,\epsilon_4\}}, \Theta_5^{\{\epsilon_3,\epsilon_5\}}, \Theta_5^{\{\epsilon_4,\epsilon_5\}}\Big\}.\end{aligned}$$

Each orthant in $\mathcal{Q}(D_5)$ has dimension $n(n-3)/2 = 5$, and there are $(n-1)!/2 = 12$ orthants. We use $\mathcal{O}_i^{\{\sigma_i(1),\ldots,\sigma_i(n)\}}$ to denote the orthant spanned by root subsystems sharing the circular ordering σ_i. Orthants in $\mathcal{Q}(D_5)$ are spanned by the following root subsystems:

$$\mathcal{O}_1^{\{1,2,3,4,5\}} = \left\{\Theta_5^{\{\epsilon_1,\epsilon_2\}}, \Theta_5^{\{\epsilon_2,\epsilon_3\}}, \Theta_5^{\{\epsilon_3,\epsilon_4\}}, \Theta_5^{\{\epsilon_4,\epsilon_5\}}, \Theta_5^{\{\epsilon_1,\epsilon_5\}}\right\},$$

$$\mathcal{O}_2^{\{1,2,3,5,4\}} = \left\{\Theta_5^{\{\epsilon_1,\epsilon_2\}}, \Theta_5^{\{\epsilon_2,\epsilon_3\}}, \Theta_5^{\{\epsilon_3,\epsilon_5\}}, \Theta_5^{\{\epsilon_4,\epsilon_5\}}, \Theta_5^{\{\epsilon_1,\epsilon_4\}}\right\},$$

$$\mathcal{O}_3^{\{1,2,4,3,5\}} = \left\{\Theta_5^{\{\epsilon_1,\epsilon_2\}}, \Theta_5^{\{\epsilon_2,\epsilon_4\}}, \Theta_5^{\{\epsilon_3,\epsilon_4\}}, \Theta_5^{\{\epsilon_3,\epsilon_5\}}, \Theta_5^{\{\epsilon_1,\epsilon_5\}}\right\},$$

$$\mathcal{O}_4^{\{1,2,4,5,3\}} = \left\{\Theta_5^{\{\epsilon_1,\epsilon_2\}}, \Theta_5^{\{\epsilon_2,\epsilon_4\}}, \Theta_5^{\{\epsilon_4,\epsilon_5\}}, \Theta_5^{\{\epsilon_3,\epsilon_5\}}, \Theta_5^{\{\epsilon_1,\epsilon_3\}}\right\},$$

$$\mathcal{O}_5^{\{1,2,5,3,4\}} = \left\{\Theta_5^{\{\epsilon_1,\epsilon_2\}}, \Theta_5^{\{\epsilon_2,\epsilon_5\}}, \Theta_5^{\{\epsilon_3,\epsilon_5\}}, \Theta_5^{\{\epsilon_3,\epsilon_4\}}, \Theta_5^{\{\epsilon_1,\epsilon_4\}}\right\},$$

$$\mathcal{O}_6^{\{1,2,5,4,3\}} = \left\{\Theta_5^{\{\epsilon_1,\epsilon_2\}}, \Theta_5^{\{\epsilon_2,\epsilon_5\}}, \Theta_5^{\{\epsilon_4,\epsilon_5\}}, \Theta_5^{\{\epsilon_3,\epsilon_4\}}, \Theta_5^{\{\epsilon_1,\epsilon_3\}}\right\},$$

$$\mathcal{O}_7^{\{1,5,2,3,4\}} = \left\{\Theta_5^{\{\epsilon_1,\epsilon_5\}}, \Theta_5^{\{\epsilon_2,\epsilon_5\}}, \Theta_5^{\{\epsilon_2,\epsilon_3\}}, \Theta_5^{\{\epsilon_3,\epsilon_4\}}, \Theta_5^{\{\epsilon_1,\epsilon_4\}}\right\},$$

$$\mathcal{O}_8^{\{1,5,2,4,3\}} = \left\{\Theta_5^{\{\epsilon_1,\epsilon_5\}}, \Theta_5^{\{\epsilon_2,\epsilon_5\}}, \Theta_5^{\{\epsilon_2,\epsilon_4\}}, \Theta_5^{\{\epsilon_3,\epsilon_4\}}, \Theta_5^{\{\epsilon_1,\epsilon_3\}}\right\},$$

$$\mathcal{O}_9^{\{1,5,3,2,4\}} = \left\{\Theta_5^{\{\epsilon_1,\epsilon_5\}}, \Theta_5^{\{\epsilon_3,\epsilon_5\}}, \Theta_5^{\{\epsilon_2,\epsilon_3\}}, \Theta_5^{\{\epsilon_2,\epsilon_4\}}, \Theta_5^{\{\epsilon_1,\epsilon_4\}}\right\},$$

$$\mathcal{O}_{10}^{\{1,5,3,4,2\}} = \left\{\Theta_5^{\{\epsilon_1,\epsilon_5\}}, \Theta_5^{\{\epsilon_3,\epsilon_5\}}, \Theta_5^{\{\epsilon_3,\epsilon_4\}}, \Theta_5^{\{\epsilon_2,\epsilon_4\}}, \Theta_5^{\{\epsilon_1,\epsilon_2\}}\right\},$$

$$\mathcal{O}_{11}^{\{1,5,4,2,3\}} = \left\{\Theta_5^{\{\epsilon_1,\epsilon_5\}}, \Theta_5^{\{\epsilon_4,\epsilon_5\}}, \Theta_5^{\{\epsilon_2,\epsilon_4\}}, \Theta_5^{\{\epsilon_2,\epsilon_3\}}, \Theta_5^{\{\epsilon_1,\epsilon_3\}}\right\},$$

$$\mathcal{O}_{12}^{\{1,5,4,3,2\}} = \left\{\Theta_5^{\{\epsilon_1,\epsilon_5\}}, \Theta_5^{\{\epsilon_4,\epsilon_5\}}, \Theta_5^{\{\epsilon_3,\epsilon_4\}}, \Theta_5^{\{\epsilon_2,\epsilon_3\}}, \Theta_5^{\{\epsilon_1,\epsilon_2\}}\right\}.$$

The simplicial complex $\mathcal{Q}(\Delta_6)$. By definition,

$$\mathcal{R}(D_6) = \mathcal{D}_{2,6} \sqcup (\mathcal{D}_{3,6} \times \mathcal{D}_{3,6}).$$

The cardinality of the set $\mathcal{D}_{2,6}$ in $\mathcal{R}(D_6)$ is $\binom{6}{2} = 15$, and the cardinality of the set $\mathcal{D}_{3,6} \times \mathcal{D}_{3,6}$ in $\mathcal{R}(D_6)$ is $\binom{6}{3}/\,2 = 10$. Each orthant in $\mathcal{Q}(D_6)$ has dimension $n(n-3)/2 = 9$, and there are $(n-1)!/2 = 60$ orthants.

Appendix B

Glossary of Maps

B.1. Denote φ_k as angular coordinates from polar coordinates. ρ is an isomorphism:

$$\rho : \mathbb{P}\,\mathrm{BHV}_n \longrightarrow \mathcal{L}_{\mathrm{BHV}}(0),$$
$$(1, \varphi_1, \ldots, \varphi_{n-3}) \mapsto (\epsilon, \varphi_1, \ldots, \varphi_{n-3}).$$

B.2. π is a map from $\mathrm{BHV}_n \setminus \{0\}$, and its image is $\mathbb{P}\,\mathrm{BHV}_n$:

$$\pi : \mathrm{BHV}_n \setminus \{0\} \longrightarrow \mathrm{BHV}_n \setminus \{0\},$$
$$\left(t_1 S_n^{A_1|B_1}, \cdots, t_{n-2} S_n^{A_{n-2}|B_{n-2}}\right) \mapsto \left(t_1 S_n^{A_1|B_1}, \cdots, t_{n-2} S_n^{A_{n-2}|B_{n-2}}\right) / \sum_{i=1}^{n-2} t_i.$$

B.3. ξ is a bijection from the 0-cone $\mathrm{BHV}_n^\star$ over $\mathbb{P}\,\mathrm{BHV}_n$ to BHV_n:

$$\xi : \mathrm{BHV}_n^\star \longrightarrow \mathrm{BHV}_n,$$
$$\left(k, \hat{t}_1 S_n^{A_1|B_1}, \cdots, \hat{t}_{n-2} S_n^{A_{n-2}|B_{n-2}}\right) \mapsto \left(k\hat{t}_1 S_n^{A_1|B_1}, \cdots, k\hat{t}_{n-2} S_n^{A_{n-2}|B_{n-2}}\right).$$

B.4. ω is a bijection between the vertex set of $\mathbb{P}\,\mathrm{BHV}_{n-1}$ to the set $\mathcal{R}(D_n)$:

$$\omega : \mathfrak{S}_{n-1} \longrightarrow \mathcal{R}(D_n),$$
$$S_{n-1}^{A_i|B_i} \mapsto \Theta_n^{A_i|B_i}.$$

B.5. ϕ is a linear map between the space of quadratic forms on Λ, a negative-definite $\mathbb{Z}$-lattice spanned by the root system of type D_n:

$$\phi : \mathrm{Sym}^2 \Lambda^\vee \longrightarrow \mathbb{Z}^{\mathcal{A}_1(D_n)}$$
$$f \mapsto \sum_{A_1 \in \mathcal{A}_1(D_n)} f(A_1)[A_1],$$

where $f(A_1)$ is equal to the value of f on one of the opposite roots of A_1.

B.6. ω induces an isomorphism $\hat{\omega}$ between a k-dimensional orthant σ_{n-1}^k in BHV_{n-1} and a k-dimensional cone $\tau_{\mathcal{F}}^k$ of $\mathcal{F}(D_n)$:

$$\begin{gathered}\hat{\omega} : \sigma_{n-1}^k \longrightarrow \tau_{\mathcal{F}}^k, \\ r_1 S_{n-1}^{A_1|B_1} + \cdots + r_k S_{n-1}^{A_k|B_k} \mapsto r_1 \Theta_n^{A_1|B_1} + \cdots + r_k \Theta_n^{A_k|B_k}.\end{gathered}$$

B.7. The isomorphism $\hat{\omega}$ extends to BHV_{n-1} and $\mathcal{F}(D_n)$:

$$\begin{gathered}\Omega : \mathrm{BHV}_{n-1} \longrightarrow \mathcal{F}(D_n), \\ r_1 S_{n-1}^{A_1|B_1} + \cdots + r_{n-3} S_{n-1}^{A_{n-3}|B_{n-3}} \mapsto r_1 \Theta_n^{A_1|B_1} + \cdots + r_{n-3} \Theta_n^{A_{n-3}|B_{n-3}}.\end{gathered}$$

B.8. λ is the identity map between the set of bipartitions $\mathfrak{S}_{n-1}$ of BHV_{n-1} and $\mathring{\mathfrak{S}}_n$ of CSN_n:

$$\begin{gathered}\lambda : \mathfrak{S}_{n-1} \longrightarrow \mathring{\mathfrak{S}}_n \\ S_{n-1}^{A_i|B_i} \mapsto \dot{S}_n^{A_i'|B_i'}.\end{gathered}$$

B.9. Λ is an injection extended from λ:

$$\begin{gathered}\Lambda : \mathrm{BHV}_{n-1} \longrightarrow \mathrm{CSN}_n, \\ \left(r_1 S_{n-1}^{A_1|B_1}, \cdots, r_{n-3} S_{n-1}^{A_{n-3}|B_{n-3}}\right) \mapsto \left(r_{i_1} S_n^{A_{i_1}|B_{i_1}}, \cdots, r_{i_M} S_n^{A_{i_M}|B_{i_M}}\right), \ M = \frac{n(n-3)}{2}.\end{gathered}$$

B.10. μ is the isomorphism between $\mathbb{P}\,\mathrm{CSN}_n$ and $\mathcal{L}_{\mathrm{CSN}}(0)$:

$$\begin{gathered}\mu : \mathbb{P}\,\mathrm{CSN}_n \longrightarrow \mathcal{L}_{\mathrm{CSN}}(0), \\ \left(l, \varphi_1, \ldots, \varphi_{\frac{n(n-3)}{2}-1}\right) \mapsto \left(\epsilon, \varphi_1, \ldots, \varphi_{\frac{n(n-3)}{2}-1}\right).\end{gathered}$$

B.11. ζ is a map from $\mathrm{CSN}_n \setminus \{0\}$, and the image of the following quotient map is $\mathbb{P}\,\mathrm{CSN}_n$:

$$\begin{gathered}\zeta : \mathrm{CSN}_n \setminus \{0\} \longrightarrow \mathrm{CSN}_n \setminus \{0\}, \\ \left(t_1 \dot{S}_n^{A_1|B_1}, \cdots, t_M \dot{S}_n^{A_M|B_M}\right) \mapsto \left(t_1 \dot{S}_n^{A_1|B_1}, \cdots, t_M \dot{S}_n^{A_M|B_M}\right) / \sum_{i=1}^{M} t_i.\end{gathered}$$

B.12. Ξ is a bijection from the 0-cone $\mathrm{CSN}_n^\star$ over $\mathbb{P}\,\mathrm{CSN}_n$ to CSN_n:

$$\begin{gathered} \Xi : \mathrm{CSN}_n^\star \longrightarrow \mathrm{CSN}_n, \\ \left(k, \hat{t}_1 \dot{S}_n^{A_1|B_1}, \cdots, \hat{t}_M \dot{S}_n^{A_M|B_M}\right) \mapsto \left(k\hat{t}_1 \dot{S}_n^{A_1|B_1}, \cdots, k\hat{t}_M \dot{S}_n^{A_M|B_M}\right). \end{gathered}$$

B.13. ι is an injection from $\mathcal{F}(D_n) \subset N(D_n)_{\mathbb{Q}}$ into the *"remarkable fan"* $\mathcal{G}(D_n)$*:*

$$\begin{aligned} \iota : \mathcal{F}(D_n) &\longrightarrow \mathcal{G}(D_n), \\ \sum_{i=1}^{n-3} r_i \psi\left(\Theta_n^{A_i|B_i}\right) &\mapsto \sum_{i=1}^{M} r_i \psi\left(\Theta_n^{A_i|B_i}\right), \ M = \frac{n(n-3)}{2}. \end{aligned}$$

B.14. υ is an isomorphism between $\mathring{\mathfrak{S}}_n \subset \mathbb{P}\,\mathrm{CSN}_n$ and the set $\mathcal{R}(D_n)$:

$$\begin{gathered} \upsilon : \mathring{\mathfrak{S}}_n \longrightarrow \mathcal{R}(D_n), \\ \dot{S}_n^{A_i|B_i} \mapsto \Theta_n^{A_i|B_i}. \end{gathered}$$

B.15. $\hat{\upsilon}$ is an isomorphism between a k-dimensional orthant σ_n^k in CSN_n and a k-dimensional cone $\tau_{\mathcal{G}}^k$ of $\mathcal{G}(D_n)$ induced from $\upsilon : \mathring{\mathfrak{S}}_n \longrightarrow \mathcal{R}(D_n)$:

$$\begin{gathered} \hat{\upsilon} : \sigma_n^k \longrightarrow \tau_{\mathcal{G}}^k \\ \sum_{i=1}^{k} r_i \dot{S}^{A_i|B_i} \mapsto \sum_{i=1}^{k} r_i \psi\left(\Theta_n^{A_i|B_i}\right). \end{gathered}$$

B.16. Υ is an isomorphism between CSN_n and $\mathcal{G}(D_n)$:

$$\begin{gathered} \Upsilon : \mathrm{CSN}_n \longrightarrow \mathcal{G}(D_n), \\ \sum_{i=1}^{n(n-3)/2} r_i \dot{S}^{A_i|B_i} \mapsto \sum_{i=1}^{n(n-3)/2} r_i \psi\left(\Theta_n^{A_i|B_i}\right). \end{gathered}$$

Bibliography

[1] Dan Abramovich, Lucia Caporaso, and Sam Payne. The tropicalization of the moduli space of curves. *arXiv preprint arXiv:1212.0373*, 2012.

[2] Benjamin L Allen and Mike Steel. Subtree transfer operations and their induced metrics on evolutionary trees. *Annals of combinatorics*, 5(1):1–15, 2001.

[3] Federico Ardila, Megan Owen, and Seth Sullivant. Geodesics in cat (0) cubical complexes. *Advances in Applied Mathematics*, 48(1):142–163, 2012.

[4] John C Baez and Nina Otter. Operads and phylogenetic trees. *Theory and Applications of Categories*, 32(40):1397–1453, 2017.

[5] Louis J Billera, Susan P Holmes, and Karen Vogtmann. Geometry of the space of phylogenetic trees. *Advances in Applied Mathematics*, 27(4):733–767, 2001.

[6] Andrew J Blumberg, Prithwish Bhaumik, and Stephen G Walker. Testing to distinguish measures on metric spaces. *arXiv preprint arXiv:1802.01152*, 2018.

[7] Andrew J Blumberg, Itamar Gal, Michael A Mandell, and Matthew Pancia. Robust statistics, hypothesis testing, and confidence intervals for persistent homology on metric measure spaces. *Foundations of Computational Mathematics*, 14(4):745–789, 2014.

[8] Martin R Bridson and André Haefliger. *Metric spaces of non-positive curvature*, volume 319. Springer Science & Business Media, 2011.

[9] David Bryant and Vincent Moulton. Neighbor-net: An agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution*, 21(2):255–265, 2004.

[10] Joseph Minhow Chan, Gunnar Carlsson, and Raul Rabadan. Topology of viral evolution. *Proceedings of the National Academy of Sciences*, page 201313480, 2013.

[11] B Chen and AJ Butte. Leveraging big data to transform target selection and drug discovery. *Clinical Pharmacology & Therapeutics*, 99(3):285–297, 2016.

[12] Bin Chen, Xiao Dong, Dazhi Jiao, Huijun Wang, Qian Zhu, Ying Ding, and David J Wild. Chem2bio2rdf: a semantic framework for linking and data mining chemogenomic and systems chemical biology data. *BMC bioinformatics*, 11(1):255, 2010.

[13] Charles Darwin. On the origin of the species by natural selection. 1859.

[14] Elza C de Bruin, Nicholas McGranahan, Richard Mitter, Max Salm, David C Wedge, Lucy Yates, Mariam Jamal-Hanjani, Seema Shafi, Nirupa Murugaesu, Andrew J Rowan, et al. Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. *Science*, 346(6206):251–256, 2014.

[15] Satyan L Devadoss and Jack Morava. Diagonalizing the genome i: navigation in tree spaces. *arXiv preprint arXiv:1009.3224*, 2010.

[16] Satyan L Devadoss and Samantha Petti. A space of phylogenetic networks. *SIAM Journal on Applied Algebra and Geometry*, 1(1):683–705, 2017.

[17] W Ford Doolittle. Phylogenetic classification and the universal tree. *Science*, 284(5423):2124–2128, 1999.

[18] Andreas WM Dress and Daniel H Huson. Constructing splits graphs. *IEEE/ACM transactions on Computational Biology and Bioinformatics*, 1(3):109–115, 2004.

[19] Ambedkar Dukkipati and Aritra Sen. Tropical grassmannian and tropical linear varieties from phylogenetic trees. *arXiv preprint arXiv:1312.0752*, 2013.

[20] Nathaniel B Edelman, Paul Frandsen, Michael Miyagi, Bernardo J Clavijo, John Davey, Rebecca Dikow, Gonzalo Garcia Accinelli, Steven M Van Belleghem, Nick J Patterson, Daniel E Neafsey, et al. Genomic architecture and introgression shape a butterfly radiation. *bioRxiv*, page 466292, 2018.

[21] George F Estabrook, FR McMorris, and Christopher A Meacham. Comparison of undirected phylogenetic trees based on subtrees of four evolutionary units. *Systematic Zoology*, 34(2):193–200, 1985.

[22] William Fulton. *Introduction to toric varieties.* Number 131. Princeton University Press, 1993.

[23] Marco Gerlinger, Stuart Horswell, James Larkin, Andrew J Rowan, Max P Salm, Ignacio Varela, Rosalie Fisher, Nicholas McGranahan, Nicholas Matthews, Claudio R Santos, et al. Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nature genetics*, 46(3):225, 2014.

[24] Verne Grant. *Plant Speciation.* Columbia University Press, New York, 2nd edition, 1981.

[25] Paul Hacking, Seán Keel, and Jenia Tevelev. Stable pair, tropical, and log canonical compactifications of moduli spaces of del pezzo surfaces. *Inventiones mathematicae*, 178(1):173–227, 2009.

[26] Susan Holmes. Bootstrapping phylogenetic trees: theory and methods. *Statistical Science*, pages 241–255, 2003.

[27] Susan Holmes. Statistics for phylogenetic trees. *Theoretical population biology*, 63(1):17–32, 2003.

[28] Susan Holmes. Statistical approach to tests involving phylogenies. *Mathematics of Evolution and Phylogeny*, pages 91–120, 2005.

[29] Daniel H Huson and David Bryant. Application of phylogenetic networks in evolutionary studies. *Molecular biology and evolution*, 23(2):254–267, 2006.

[30] Daniel H Huson, Regula Rupp, and Celine Scornavacca. *Phylogenetic networks: concepts, algorithms and applications.* Cambridge University Press, 2010.

[31] Daniel H Huson and Celine Scornavacca. A survey of combinatorial methods for phylogenetic networks. *Genome biology and evolution*, 3:23–35, 2011.

[32] Guohua Jin, Luay Nakhleh, Sagi Snir, and Tamir Tuller. Maximum likelihood of phylogenetic networks. *Bioinformatics*, 22(21):2604–2611, 2006.

[33] Guohua Jin, Luay Nakhleh, Sagi Snir, and Tamir Tuller. Efficient parsimony-based methods for phylogenetic network reconstruction. *Bioinformatics*, 23(2):e123–e128, 2007.

[34] Guohua Jin, Luay Nakhleh, Sagi Snir, and Tamir Tuller. Inferring phylogenetic networks by the maximum parsimony criterion: a case study. *Molecular Biology and Evolution*, 24(1):324–337, 2007.

[35] Reza Kalhor, Kian Kalhor, Leo Mejia, Kathleen Leeper, Amanda Graveline, Prashant Mali, and George M Church. Developmental barcoding of whole mouse via homing crispr. *Science*, 361(6405):eaat9804, 2018.

[36] Mary K Kuhner and Joseph Felsenstein. A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. *Molecular biology and evolution*, 11(3):459–468, 1994.

[37] Yinbang Lin. Moduli spaces of stable pairs. *Pacific Journal of Mathematics*, 294(1):123–158, 2018.

[38] Ke Liu, Patrick A Newbury, Benjamin S Glicksberg, William ZD Zeng, Shreya Paithankar, Eran R Andrechek, and Bin Chen. Evaluating cell lines as models for metastatic breast cancer through integrative analysis of genomic data. *Nature communications*, 10(1):2138, 2019.

[39] Yusuke Matsui, Atsushi Niida, Ryutaro Uchi, Koshi Mimori, Satoru Miyano, and Teppei Shimamura. phyc: Clustering cancer evolutionary trees. *PLoS computational biology*, 13(5):e1005509, 2017.

[40] James R Munkres. *Elements of algebraic topology*. CRC Press, 2018.

[41] Peter C Nowell. The clonal evolution of tumor cell populations. *Science*, 194(4260):23–28, 1976.

[42] Megan Owen. Computing geodesic distances in tree space. *SIAM Journal on Discrete Mathematics*, 25(4):1506–1529, 2011.

[43] Megan Owen and J Scott Provan. A fast algorithm for computing geodesic distances in tree space. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, 8(1):2–13, 2011.

[44] Ral Rabadn and Andrew J. Blumberg. *Topological Data Analysis for Genomics and Evolution: Topology in Biology.* Cambridge University Press, 2019.

[45] Qingchun Ren, Steven V Sam, and Bernd Sturmfels. Tropicalization of classical moduli spaces. *Mathematics in Computer Science*, 8(2):119–145, 2014.

[46] David F Robinson and Leslie R Foulds. Comparison of phylogenetic trees. *Mathematical biosciences*, 53(1-2):131–147, 1981.

[47] Ernst Schröder. Vier combinatorische probleme. *Z. Math. Phys*, 15:361–376, 1870.

[48] Russell Schwartz and Alejandro A Schäffer. The evolution of tumour phylogenetics: principles and practice. *Nature Reviews Genetics*, 18(4):213, 2017.

[49] Peter HA Sneath. Cladistic representation of reticulate evolution. *Systematic Zoology*, 24(3):360–368, 1975.

[50] David Speyer and Bernd Sturmfels. The tropical grassmannian. *Advances in Geometry*, 4(3):389–411, 2004.

[51] Richard P Stanley. *Catalan numbers.* Cambridge University Press, 2015.

[52] Mike A Steel and David Penny. Distributions of tree comparison metricssome new results. *Systematic biology*, 42(2):126–141, 1993.

[53] Soledad Villar, Afonso S Bandeira, Andrew J Blumberg, and Rachel Ward. A polynomial-time relaxation of the gromov-hausdorff distance. *arXiv preprint arXiv:1610.05214*, 2016.

[54] Johann Wolfgang Wägele and Christoph Mayer. Visualizing differences in phylogenetic information content of alignments and distinction of three classes of long-branch effects. *BMC Evolutionary Biology*, 7(1):147, 2007.

[55] Michael S Waterman and Temple F Smith. On the similarity of dendrograms. *Journal of theoretical biology*, 73(4):789–800, 1978.

[56] John W Wenzel and Mark E Siddall. Noise. *Cladistics*, 15(1):51–64, 1999.

[57] Yingying Wu, Yiqun Liu, Richard Yen-Hsi Tsai, and Shing-Tung Yau. Investigating the role of eye movement and electrodermal activity signals in search satisfaction prediction using geometric analysis. *Journal of the Association for Information Science and Technology*, 2019.

[58] Sakellarios Zairis, Hossein Khiabanian, Andrew J Blumberg, and Raul Rabadan. Moduli spaces of phylogenetic trees describing tumor evolutionary patterns. In *International Conference on Brain Informatics and Health*, pages 528–539. Springer, 2014.

[59] Sakellarios Zairis, Hossein Khiabanian, Andrew J Blumberg, and Raul Rabadan. Genomic data analysis in tree spaces. *arXiv preprint arXiv:1607.07503*, 2016.

[60] Kaizhong Zhang and Dennis Shasha. Simple fast algorithms for the editing distance between trees and related problems. *SIAM journal on computing*, 18(6):1245–1262, 1989.