OAW
Austrian Academy
of Sciences

JMU
JOHANNES KEPLER
UNIVERSITY LINZ

Five Equivalent Ways to Describe a Phylogenetic Tree

Jiayue Qi   Josef Schicho

A–4040 LINZ, ALTENBERGERSTRASSE 69, AUSTRIA

Five Equivalent Ways to Describe
a Phylogenetic Tree

Jiayue Qi [*] and Josef Schicho [†]

November 23, 2020

Abstract

A phylogenetic tree is a tree with a fixed set of leaves N that has no vertices of degree two. In this paper, we axiomatically define four other discrete structures on N; each of the structures is an equivalent representation of a phylogenetic tree.

Introduction

Trees with labeled leaves and without vertices of degree two are used in phylogenetics to describe evolutionary relations between species. These trees are called phylogenetic trees (see [1]). They arise in two kinds, rooted and unrooted. In this paper, we study unrooted phylogenetic trees from a combinatorial point of view. In mathematics, various ways of representing and analyzing phylogenetic trees have been given. In [2], M. A. Ragan represented phylogenetic trees by matrices. In [3], L. J. Billera et al. defined a continuous space that has a decomposition into disjoint subsets, each corresponding to a phylogenetic tree. The compatibility of unrooted phylogenetic trees – whether or not each tree in a given set can be constructed by deleting leaves and contracting edges of a single larger tree – has been studied in [4, 6].

A different context where phylogenetic trees arise is the moduli space of n-pointed stable curves of genus zero [5]. Similar as in [3], this moduli space has a natural decomposition into disjoint subsets. The points in each subset correspond to stable curves that have the same number of irreducible components, the same intersection numbers of irreducible components, and the same incidence relations of marked points and irreducible components. Since the intersection number of two irreducible curves can only be zero or one, the intersection numbers are determined by the "dual graph", which is defined as follows: each vertex corresponds to irreducible component; two vertices are joined by an edge if and only if the corresponding components intersect. For

---

[*]Doctoral Program "Computational Mathematics" W1214, Johannes Kepler University Linz.

[†]Research Institute for Symbolic Computation, Johannes Kepler University Linz.

each component C and for each marked point contained in C, we attach a new edge with a leaf (a new vertex of degree one) to the vertex corresponding to C. In this way, we also get a phylogenetic tree.

In a joint research together with H. Hauser, we are currently studying the combinatorics of this moduli space. As a byproduct which is interesting on its own, we observed that there are several other discrete structures which can be viewed as equivalent descriptions of an unrooted phylogenetic tree: a collection of partitions, a set of cuts (where a cut is a partition which contains two parts), an equivalence relation on the three-element-subsets, and a quaternary relation which we call "crossing relation" because of its connection with the cross product of four points on the projective line. For all these structures, there are specific axioms which specify which collections of partitions/sets of cuts/equivalence relations/quaternary relations are indeed equivalent to unrooted phylogenetic trees. In this paper, we present these four axiom systems and we prove that each of the four discrete structures is an equivalent description of an unrooted phylogenetic tree. As a potential application, we note that the question of compatibility could equivalently posed with any of the four other structures; this could lead to a new approach to decide compatibility.

Here is an informal taxonomy of the four equivalent representations of phylogenetic trees. The collections of partitions and the sets of cuts are *macroscopic* pictures, in the sense that they are composed of elements of bigger scales. In contrast, the equivalence relations on three-element-subsets and the crossing relations are composed of smaller-scale elements; they are *microscopic* pictures: the crossing relations are just quaternary relations on the set of leaves, and the equivalence relations on three-element subsets can be considered as 6-ary relations on the set of leaves. The partitions and the cuts focus on the vertices of the phylogenetic tree, in particular the non-leaves, while the cuts and the crossing relations focus on its edges.

1 Structures and Axioms

In this section, we axiomatically define five discrete structures on a fixed finite set N of cardinality at least 3: phylogenetic trees, collections of partitions, sets of cuts, crossing relations, and equivalence relations of triples. We will also introduce some functions converting one of these structures to another, and use them to construct examples. Other (more involved) conversion functions will be introduced in Section 2.

1.1 Trees

Recall that an unrooted tree is an undirected graph that is connected and has no cycles.

Definition 1.1. A *phylogenetic tree* with leaf set N is an unrooted tree (V, E) without vertex of degree 2 such that $N \subset V$ is the set of leaves. We say that two phylogenetic trees (V_1, E_1) and (V_2, E_2) with leaf set V are isomorphic if
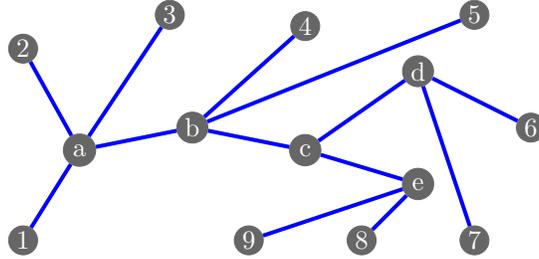
Figure 1: This is a phylogenetic tree with leaf set $N = \{1, \ldots, 9\}$. The set of internal nodes is $V \setminus N = \{a, b, c, d, e\}$.

and only if there is a graph isomorphism that restricts to the identity on the subset of leaves.

The set $\mathcal{T}_N$ is defined as the set of all isomorphism classes of phylogenetic trees with leaf set N.

An example of a phylogenetic tree with leaf set $N = \{1, \ldots, 9\}$ can be seen in Figure 1.

1.2 Sets of Partitions

Recall that a partition of N is a set of disjoint and non-empty subset of N such that their union is N.

Definition 1.2. A collection/set of partitions of N is *phylogenetic* if it fulfills the following axioms:

(P1) Every partition has at least 3 subsets; we also call these subsets the *parts*.

(P2) Every one-element subset of N is a part of some partition.

(P3) Every subset of N is a part of at most one partition.

(P4) For every part $A \subset N$ of cardinality bigger than one, the complement $N \setminus A$ is also a part (necessarily of a different partition, by Axiom (P1)).

We denote as $\mathcal{P}_N$ the set of all phylogenetic collections of partitions of N. The set $\overline{\mathcal{P}_N}$ is the set of all sets of collections of partitions of N.

Example 1.3. For $N = \{1, \ldots, 9\}$, let $P = \{p_a, p_b, p_c, p_d, p_e\}$ be the collection of the partitions

$$p_a = \{\{1\}, \{2\}, \{3\}, \{4, 5, 6, 7, 8, 9\}\},$$

$$p_b = \{\{1, 2, 3\}, \{4\}, \{5\}, \{6, 7, 8, 9\}\},$$

$$p_c = \{\{1, 2, 3, 4, 5\}, \{6, 7\}, \{8, 9\}\},$$

$$p_d = \{\{1, 2, 3, 4, 5, 8, 9\}, \{6\}, \{7\}\},$$

3

$$p_e = \{\{1, 2, 3, 4, 5, 6, 7\}, \{8\}, \{9\}\}.$$

It can be checked that the axioms (P1),(P2),(P3),(P4) are fulfilled. Therefore the collection P is phylogenetic.

The above example can be constructed from the tree in Figure 1 in a systematic way, which is described in the following definition.

Definition 1.4. For every phylogenetic tree $T = (V, E)$ with leaf set N, we define a collection of partitions which is in bijection with the set $V \setminus N$ of non-leaves, as follows.

For each non-leaf vertex v, and for each edge e incident to v, we have a subset of leaves containing the leaves i such that the unique path from v to i begins with e. For each non-leaf vertex v, these subsets are the parts of a partition of N. The collection of partitions of N is denoted by P_T. The function $\mathcal{T}_N \to \overline{\mathcal{P}_N}$ mapping the class of T to P_T is denoted by t_{TP} and we call it the transformation from trees to partition collections.

Proposition 1.5. *For every phylogenetic tree $T = (V, E)$ with leaf set N, the collection P_T is phylogenetic.*

Proof. Every non-leaf has at least 3 edges, hence (P1) holds. Every leaf has a unique neighbor which must be a non-leaf, otherwise the tree would only have two vertices (which violates our assumption on the cardinality of N); this implies (P2). Uniqueness of the neighbor also implies (P3) for the special case of cardinality-one parts.

Let A be a part of P_T such that $2 \leq |A| \leq |N| - 2$. Then there are non-leaves a, b and an edge $e = \{a, b\} \in E$ such that A is equal to the set of leaves i such that the unique path from a to i contains e. And then $N \setminus A$ is the set of leaves j such that the unique path from b to j contains e, hence $N \setminus A$ is part of the partition corresponding to the non-leaf b. This shows (P4).

It remains to prove Axiom (P3). Suppose that A belongs to two distinct partitions of P_T, corresponding to non-leafs a and a' respectively. Let $e = \{a, b\}$ and $e' = \{a', b'\}$ be the two edges corresponding to A in the partitions corresponding to a and a' respectively. There is a unique path $p_{a,a'}$ between a and a'. In the sequel, we do case distinctions on whether e and e' belongs to path $p_{a,a'}$ or not.

1. Neither e nor e' belongs to $p_{a,a'}$. In this case, we remove the edge e, and we obtain two components T_b and T_a, where T_b contains the elements of A and T_a contains the elements of $N \setminus A$. Now we remove e' in T_a, and we further obtain two components $T_{b'}$ and $T_{a'}$, where $T_{b'}$ contains the elements from A. However, $T_{b'}$ and T_b are distinct components, they cannot both contain the elements from A. This is a contradiction.

2. Both e and e' belong to $p_{a,a'}$. In this case, we can argue analogously, by interchanging the roles of a with b, and the roles of a' with b'. We remove edge e, obtaining components T_a which contains elements in $N \setminus A$, and T_b which contains elements in A. Similarly, there must be a unique path from

b to b' connecting edge e and e'. Hence, e' is in T_b. Now, in component T_b, we remove the edge e', obtaining components $T_{a'}$ and $T_{b'}$, where $T_{b'}$ is the component containing elements in $N \setminus A$. However, $T_{b'}$ and T_a are distinct components, hence cannot both contain elements in $N \setminus A$. This is a contradiction.

3. Only one of e and e' belongs to $p_{a,a'}$. Without loss of generality, assume that e' belongs to path $p_{a,a'}$ and e does not. We remove edge e from the tree T, obtaining T_b and T_a, where the set of leaves in T_b intersected with N is A. We remove edge e' from the tree T, obtaining $T_{b'}$ and $T_{a'}$, where the set of leaves in $T_{b'}$ intersected with N is A. Because T is phylogenetic, there is an edge e'' not contained in the path $p_{a,a'}$, but incident with some vertex on this path. Following this edge, we eventually arrive at some leaf l. Then we have that l is in $T_{b'} \cap N$ but not in $T_b \cap N$. This is a contradiction.

Hence, Axiom (P3) holds. $\qquad\square$

1.3   Sets of Cuts

In this structure, we are particularly interested in partitions that violate (P1).

Definition 1.6. A *cut* of N is a partition of N into two subsets A, B of cardinality larger than one. The subsets A, B are called *clusters*. We denote such a cut as $(A \mid B) = (B \mid A)$, and omit the curly brackets when A and B are given by the enumeration of elements. Axiom for a set C of cuts of N to be *phylogenetic* is as follows. And we denote as $cl(C)$ the set of all clusters of C.

(C) For any two cuts $(A_1 \mid B_1)$, $(A_2 \mid B_2)$ in C, at least one of the following four sets is empty: $A_1 \cap A_2$, $A_1 \cap B_2$, $B_1 \cap A_2$, $B_1 \cap B_2$.

Denote as $\mathcal{C}_N$ the set of all phylogenetic sets of cuts of N. The set $\overline{\mathcal{C}_N}$ is the set of all sets of cuts of N.

Remark 1.7. One can check that actually we can omit "at least" in the above statement, since when it holds, it cannot happen that two of those four sets are both empty.

Example 1.8. For $N = \{1, \ldots, 9\}$, let $C = \{c_x, c_y, c_z, c_w\}$ be the set of cuts

$$c_x = (1, 2, 3 \mid 4, 5, 6, 7, 8, 9),$$

$$c_y = (1, 2, 3, 4, 5 \mid 6, 7, 8, 9),$$

$$c_z = (1, 2, 3, 4, 5, 8, 9 \mid 6, 7),$$

$$c_w = (1, 2, 3, 4, 5, 6, 7 \mid 8, 9).$$

It can be checked that axiom (C) is fulfilled, hence C is phylogenetic.

Note that the clusters in these cuts are exactly the parts that appear in some partition in Example 1.3 that have cardinality bigger than one. Every cut

corresponds to an internal edge of the phylogenetic tree, i.e., an edge connecting two non-leafs. The clusters are just the set of leaves of the two connected components which arise when the corresponding edge is removed.

Definition 1.9. Let T be a phylogenetic tree with leaf set N. Then C_T is the set of cuts of N that corresponds to internal edges of T, with clusters being the two sets of leaves of the two components that arise when the corresponding edge is removed. The function $\mathcal{T}_N \to \overline{\mathcal{C}_N}$, $[T] \mapsto C_T$ is denoted by t_{TC} and we call it the transformation from trees to sets of cuts.

Let P be a phylogenetic collection of partitions. The set of cuts whose clusters are exactly the parts of cardinality at least 2 is denoted by C_P. The function $\mathcal{P}_N \to \overline{\mathcal{C}_N}$, $P \mapsto C_P$ is denoted by t_{PC} and we call it the transformation from partition collections to sets of cuts.

Proposition 1.10. *For every tree T, we have $C_{P_T} = C_T$; in other words, we have $t_{PC} \circ t_{TP} = t_{TC}$.*

Proof. Straightforward. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

1.4 Crossing Relations

Definition 1.11. A *crossing relation* is a set X of a pair of cardinality-two subsets of N. We write its element as $(i, j \mid k, l)$ - such that if $(i, j \mid k, l) \in X$, then the information that i, j, k, l are pairwise distinct is contained. And we call it a *cross (of X)*, since we can interchange i, j or k, l, or two sets $\{i, j\}$ with $\{k, l\}$ without changing the element.

Axioms for a crossing relation X to be *phylogenetic* are as follows.

(X1) If $(i, j \mid k, l)$, then not $(i, k \mid j, l)$, i.e., $(i, k \mid j, l) \notin X$.

(X2) If $(i, j \mid k, l)$ and $(i, j \mid k, m)$ and $l \neq m$, then $(i, j \mid l, m)$.

(X3) If $(i, j \mid k, l)$ and m is distinct from i, j, k, l, then $(i, j \mid k, m)$ or $(i, m \mid k, l)$.
Note that this "or" here means at least one should hold and it may happen that both hold.

Denote as $\mathcal{X}_N$ the set of all phylogenetic crossing relations on N. Denote as $\overline{\mathcal{X}_N}$ the set of all crossing relations of N.

Example 1.12. Let $N := \{1, \ldots, 9\}$. We define a crossing relation as follows. For any i, j, k, l that are pairwise distinct, the relation $(i, j \mid k, l)$ holds if and only if one of the following statements is true:

$\quad \triangleright\ i, j \in \{1, 2, 3\}$ and $k, l \in \{4, 5, 6, 7, 8, 9\}$ (45 crosses);

$\quad \triangleright\ i, j \in \{1, 2, 3, 4, 5\}$ and $k, l \in \{6, 7, 8, 9\}$ (60 crosses);

$\quad \triangleright\ i, j \in \{1, 2, 3, 4, 5, 8, 9\}$ and $\{k, l\} = \{6, 7\}$ (21 crosses);

$\quad \triangleright\ i, j \in \{1, 2, 3, 4, 5, 6, 7\}$ and $\{k, l\} = \{8, 9\}$ (21 crosses).

(Its relation with Example 1.8 is apparent.) In total, this crossing relation consists of 108 crosses. We will see later that this crossing relation is phylogenetic.

The following definition is a generalization of the construction in Example 1.12.

Definition 1.13. For every set C of cuts, we define a crossing relation X_C as follows: (i, j, k, l) is in X_C if and only if C contains a cut $(A \mid B)$ such that $i, j \in A$ and $k, l \in B$. The function $\mathcal{C}_N \to \overline{\mathcal{X}_N}$, $C \mapsto X_C$ is denoted by t_{CX}, we call it the transformation from sets of cuts to crossing relations.

Remark 1.14. The moduli space of n-pointed stable curves of genus zero has a natural decomposition into strata that correspond to phylogenetic trees with leaf set of cardinality n.

For any such stratum T, the crossing relation is then exactly the set of $(i, j \mid k, l)$ such that the cross ratio of the four marked points p_i, p_j, p_k, p_l has value 1. As we will see, it is possible to transform a phylogenetic crossing relation to a phylogenetic tree. In the context of moduli spaces, this is equivalent to saying that we can recover the dual graph of the n-pointed graph from the values of its cross ratios.

1.5 Equivalences of Triples

A *triple* in N is a 3-element subset of N. We denote the set of triples in N by $\binom{N}{3}$. A set $S \subset \binom{N}{3}$ of triples is called *diverse* if it is non-empty and it fulfills the following two axioms:

(D1) If $\{i, j, k\} \in S$, and $l \in N$, then S also contains one of the triples $\{i, j, l\}$, $\{i, k, l\}$, or $\{j, k, l\}$.

(D2) Let $a, b, c, x, y, z \in N$. If S contains the triples $\{a, x, y\}$, $\{b, y, z\}$, and $\{c, x, z\}$, then it also contains $\{x, y, z\}$.

We say that an equivalence relation on $\binom{N}{3}$ is *phylogenetic* if and only if the following axiom is fulfilled:

(E0) Each class of the equivalence relation is diverse.

Denote as $\mathcal{E}_N$ the set of all phylogenetic equivalence relations on the triples of N. Denote as $\overline{\mathcal{E}_N}$ the set of all equivalence relations on the triples of N.

Example 1.15. Let $N = \{1, 2, 3, 4, 5\}$. We define an equivalence relation with three classes as follows:

$$\{1, 2, 3\} \sim \{1, 2, 4\} \sim \{1, 2, 5\},$$

$$\{1, 4, 5\} \sim \{2, 4, 5\} \sim \{3, 4, 5\},$$

$$\{1, 3, 4\} \sim \{1, 3, 5\} \sim \{2, 3, 4\} \sim \{2, 3, 5\}.$$

It can be checked that the axioms (D1) and (D2) are fulfilled in each class, hence the equivalence is phylogenetic.

In order to construct interesting equivalence relations of triples, we need a lemma on trees.

Lemma 1.16. *Let $T = (V, E)$ be a tree, and let $i, j, k \in N$ be pairwise distinct leaves. Then there is a unique vertex $v \in V \setminus N$ such that the three paths from v to i, j, and k are edge-disjoint.*

Proof. Let π_{ij} be the unique path connecting i and j, and let π_{ik} be the unique path connecting i and k. The common edges of π_{ij} and π_{ik} also form a path, which connects i to some non-leaf, and this is the vertex v we are looking for. Indeed, the edges that are in π_{ij} but not in π_{ik} connect j with v, and the edges that are in π_{ik} but not in π_{ij} connect v with k. The property of v which is claimed in the lemma implies that the common edges of π_{ij} and π_{ik} form a path from i to v, and this implies uniqueness. $\square$

Definition 1.17. Let $T = (V, E) \in \mathcal{T}_N$ be a phylogenetic tree. We define an equivalence relation $\sim_T$ on $\binom{N}{3}$ as follows: $\{i, j, k\} \sim_T \{l, m, n\}$ holds if and only if the unique non-leaf v such that the three paths from v to i, j, and k are edge-disjoint is equal to the unique non-leaf w such that the three paths from w to l, m, and n are edge-disjoint. The function $\mathcal{T}_N \to \overline{\mathcal{E}_N}$, $[T] \mapsto \sim_T$ is denoted by t_{TE}, we call it the transformation from trees to equivalence relations.

Example 1.18. Let $N = \{1, \ldots, 9\}$, and let $T = (V, E)$ be the phylogenetic tree in Figure 1. Then the equivalence relation $\sim_T$ in $\binom{N}{3}$ has five equivalence classes $E_a, \ldots, E_b$, corresponding to the five non-leaves in $V \setminus N$:

  ▷ The class E_a consists of all triples $\{i, j, k\}$ such that ($i, j \in \{1, 2, 3\}$ and $k \in \{4, 5, 6, 7, 8, 9\}$) or ($i = 1$, $j = 2$, and $k = 3$). These are 18+1=19 triples.

  ▷ The class E_b consists of all triples $\{i, j, k\}$ such that ($i, j \in \{1, 2, 3\}$ and $j \in \{4, 5\}$ and $k \in \{6, 7, 8, 9\}$), or ($i = 4$ and $j = 5$ and $k \in \{1, 2, 3, 6, 7, 8, 9\}$). These are $24 + 7 = 31$ triples.

  ▷ The class E_c consists of all triples $\{i, j, k\}$ such that $i \in \{1, 2, 3, 4, 5\}$, $j \in \{6, 7\}$ and $k \in \{8, 9\}$. These are 20 triples.

  ▷ The class E_d consists of all triples $\{i, 6, 7\}$ such that $i \in \{1, 2, 3, 4, 5, 8, 9\}$. These are 7 triples.

  ▷ The class E_e consists of all triples $\{i, 8, 9\}$ such that $i \in \{1, 2, 3, 4, 5, 6, 7\}$. These are 7 triples.

Note that $19 + 31 + 20 + 7 + 7 = 84 = \binom{9}{3}$, which indicates that we did not make a mistake – every triple occurs in exactly one class.

In the next section, we will introduce the transformation between equivalences of triples and crossing relations.

$$\mathcal{T}_N$$

$$\overline{\mathcal{E}_N} \qquad\qquad \mathcal{P}_N$$

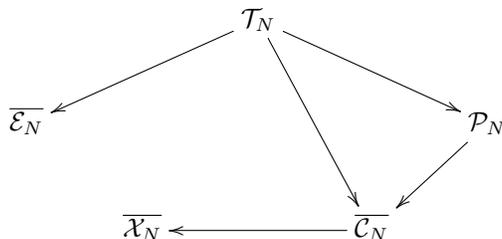$$\overline{\mathcal{X}_N} \longleftarrow \overline{\mathcal{C}_N}$$

Figure 2: This diagram shows the conversion maps between different types of structures that have been defined in Section 1. We also have seen that the triangle on the right part is commutative.

2 Conversions

In this section, we prove that the five structures introduced in Section 1 are equivalent. In Section 1, we already introduced the maps shown in Figure 2, and we have seen that the triangle contained in this diagram is commutative. We still have to show that the images of t_{TC}, t_{CX}, t_{TE} are phylogenetic, we have to construct more conversion maps so that the diagram has directed paths between any two vertices, and we have to show that the enlarged diagram commutes.

2.1 Trees and Partitions

Definition 2.1. For every set $P = \{p_1, \ldots, p_m\}$ of partitions of N, we define the graph G_P as follows. The vertex set is $N \cup P$. Two vertices in $p, q \in P$ are connected by an edge if and only if there is a cut $(A \mid B)$ such that $A \in p$ and $B \in q$. A vertex $p \in P$ and a vertex $i \in N$ are connected if $\{i\} \in p$. There is no edge connecting two vertices in N.

In the following, we will show that G_P is a phylogenetic tree whenever P is phylogenetic, and that the construction $P \to G_P$ is the inverse of t_{TP}.

Theorem 2.2. *Assume that P is a phylogenetic set of partitions of N. Then G_P is a phylogenetic tree.*

Proof (H. Hauser). Let $i \in N$ and $p \in P$. We claim that there is a path connecting i and p. Let A be the part in the partition p that contains i. If $A = \{i\}$, then there is an edge connecting i and p.

If A has cardinality bigger than 1, then there is a unique partition q containing $N \setminus A$. It also has a unique part B that contains i, which must be a strict subset of A. By induction on the cardinality of A, there is a path connecting i and q. This shows the existence of a path connecting i and p.

It follows that the graph G_P is connected. In order to show that G_P is a tree, it suffices to show that it has no cycle. The vertices of such a cycle would have to be in P, because the vertices in N have degree 1.

Let $(p_1, \ldots, p_k, p_{k+1} = p_1)$ be a cycle. For $r = 1, \ldots, k$, there is a unique part $A_r \in p_r$ that contains i. For the edge $e = p_1 p_2$, one part of its corresponding cut, say (I, J) must contain i, and it must be either A_1 or A_2. If it is A_1, then we have $A_{r+1} \subsetneq A_r$ for $r = 1, \cdots, k$ because of Axiom (P3). If it is A_2, then we have that $A_r \subsetneq A_{r+1}$ for $r = 1, \cdots, k$. Both cases lead to $A_1 \subsetneq A_1$, which is a contradiction.

The degree of any vertex in P is equal to the number of its parts, which is at least three. Therefore, the tree G_P is phylogenetic. $\qquad\square$

If T is a phylogenetic tree, then it is straightforward to check that G_{P_T} is isomorphic to T. Also, if P is a phylogenetic set of partitions, then $P_{G_P} = P$. Hence the construction $P \to G_P$ provides the inverse to $t_{TP} : \mathcal{T}_N \to \mathcal{P}_N$.

2.2 Trees and Cuts

Proposition 2.3. *For every phylogenetic tree $T = (V, E)$ with leaf set $N \subset V$, the set C_T of cuts is phylogenetic.*

Proof. Let $(A_1 \mid B_1)$ and $(A_2 \mid B_2)$ be two arbitrary cuts in C_T, and let e_1 and e_2 be their corresponding edges. If we remove both edges from the graph, then we get at most three components. correspondingly, we obtain three leaf sets. Each of the four sets $A_1 \cap A_2$, $A_1 \cap B_2$, $A_2 \cap B_1$, $A_2 \cap B_2$ equals to one of these leaf sets, if not empty. Also, note that these four sets are pairwise disjoint. Therefore, at least one of these four sets must be empty. Since the two cuts were chosen arbitrarily, it follows that Axiom (C) is fulfilled, and C_T is phylogenetic. $\qquad\square$

For the construction of transformation t_{CT} from cuts to trees, recall the following concept: if $(P, \leq)$ is a finite partially ordered set, then the *Hasse diagram* of $(P, \leq)$ is the directed graph with vertex set P. And there is an edge from vertex a to vertex b if and only if ($a \leq b$ and for all c such that $a \leq c \leq b$, we have $a = c$ or $b = c$).

We call a set/subset with exactly one element a *singleton*.

Definition 2.4. Let C be a phylogenetic set of cuts. Let $c = (A \mid B)$ be a cut in C. Let V_A be the set of all clusters or singletons that are subsets of A (including A itself). Let $G_A = (V_A, E_A)$ be the Hasse diagram of V_A ordered by inclusion. Let V_B be the set of all clusters or singletons that are subsets of B (including B itself). Let $G_B = (V_B, E_B)$ be the Hasse diagram of V_B ordered by inclusion.

Let $G_{C,c} = (V_{C,c}, E_{C,c})$ be the undirected graph with $V_{C,c} := V_A \cup V_B$, and $E_{C,c}$ is equal to the union of E_A and E_B – forgetting the direction – plus one extra edge connecting A and B. We call $G_{C,c}$ the *cut graph of c*.

Example 2.5. Let $N = \{1, \ldots, 9\}$. Let C be the set of cuts in Example (1.8). Recall that $c_y = (A \mid B)$ where $A = \{1, 2, 3, 4, 5\}$ and $B = \{6, 7, 8, 9\}$. Figure 3 shows the Hasse diagrams G_A and G_B and the graph G_{C,c_y}.
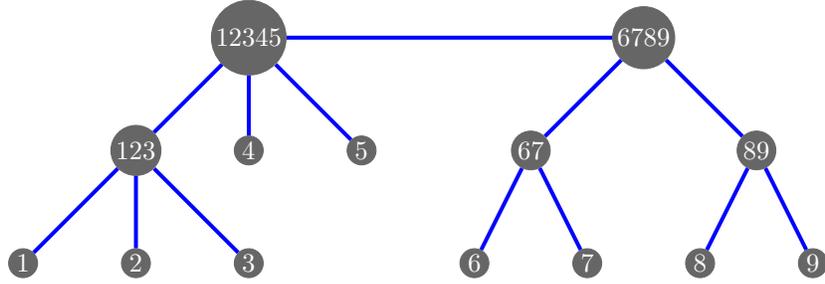
Figure 3: For the set of cuts in Example 1.8, this figure shows the two Hasse diagrams of the partial orders of clusters that are subsets of $A = \{1, 2, 3, 4, 5\}$ (on the left) and $B = \{6, 7, 8, 9\}$ (on the right), respectively; and the graph $G_{C_{c_y}}$ which is defined as the union of these two Hasse diagrams plus one extra edge. We omit the directions of the edges in the Hasse diagram here; the convention is that they are always upward.

Lemma 2.6. *Let C be a phylogenetic set of cuts of N, and let $c = (A \mid B) \in C$. Then $G_{C,c}$ is a phylogenetic tree with leaf set $\{\{i\} \mid i \in N\}$.*

Proof. Since the partially ordered sets V_A and V_B have a largest element, the two Hasse diagrams are connected. The extra edge connects the two Hasse diagrams, therefore $G_{C,c}$ is connected.

Let $v, w \in V_A$. By Axiom (C), the set of clusters u such that $v \leq u \leq w$ is totally ordered by inclusion. Therefore, the Hasse diagram G_A has no cycle. The same holds for the Hasse diagram G_B. Hence both Hasse diagrams are trees. $G_{C,c}$ is obtained by connecting two trees with one extra edge, and so $G_{C,c}$ is also a tree. Its leaves are the minimal elements of the two partial orders, which are exactly the singletons of N.

Assume, for the sake of contradiction, that $G_{C,c}$ has a vertex of degree two. Then there is a cluster D of C such that the partially ordered set V_D has a "second largest element" D', i.e. an element which is largest in the subset of elements not equal to D. Then $D \setminus D'$ is not empty, hence there is a singleton $\{a\}$ such that $\{a\} \subseteq D$ but $\{a\} \not\subseteq D'$, which is a contradiction. $\square$

Proposition 2.7. *Let $T = (V, E)$ be a phylogenetic tree with leaf set $N \subset V$. Let $e = \{u, v\} \in E$ be an internal edge and $c = (A \mid B) \in C_T$ be its corresponding cut. Then the phylogenetic tree $G_{C_T, c} = (V_G, E_G)$ is isomorphic to T.*

Proof. We obtain two components T_u and T_v after removing e from T. The set of leaves of T_u is $A \cup \{u\}$, and the set of leaves of T_v is $B \cup \{v\}$. We define a map $f : V \to V_G$. Let $w \in V$.

1. If w is a leaf of T, then $f(w) := \{w\}$.

2. If w is an internal vertex of T contained in T_u, then let e' denote the first edge on the unique path in T from w to v. Let $(A' \mid B')$ be the

11

corresponding cut, and assume without loss of generality that $A' \subset A$. We set $f(w) := A'$.

3. Analogously, if w is an internal vertex T contained in T_v, then let e' denote the first edge on the unique path in T from w to u. Let $(A' \mid B')$ be the corresponding cut, and assume without loss of generality that $B' \subset B$. We set $f(w) := B'$.

Injectivity of f is a consequence of the fact that P_T satisfies the axiom (P3). We claim that f is also surjective. Let x be a vertex of V_G. If $x = \{i\}$ is a singleton, then $f(i) = x$. If $x = A_1$ is a cluster contained in A, then let e_1 be the edge corresponding to the cut $(A_1 \mid N \setminus A_1)$. Then f maps one of the two vertices of e_1 to x. Similarly, we can find a preimage if $x = A_1$ is a cluster contained in B. Therefore, f provides a bijection between V and V_G.

We claim that f is a graph isomorphism. Let $v_1, v_2 \in V$. If v_1, v_2 are both leaves, then $\{v_1, v_2\}$ is not an edge of T and $\{f(v_1), f(v_2)\}$ is not an edge of E_G.

Assume $v_1 \in N$ and $v_2 \notin N$. Then $\{v_1, v_2\}$ is an edge of T if and only if v_2 is the unique vertex adjacent to v_1, and this is true if and only if $f(v_2)$ is the unique minimal cluster contained in A and containing v_1, and this is true if and only if $\{f(v_1), f(v_2)\}$ is an edge in E_G. The case $v_1 \in B$ is treated analogously.

If v_1 and v_2 are non-leaves of T contained in T_u. Assume that $e' = \{v_1, v_2\}$ is an edge of T. Let $(A' \mid B')$ be the corresponding cut; without loss of generality, assume $A' \subset A$ and $A' \cup \{v_1\}$ is the set of leaves of one of the two components that we get when we remove e' from T. Let $\{v_2, v_3\}$ be the first edge on the path from v_2 to u and let $(A'' \mid B'')$ be its corresponding cut such that $A'' \subset A$. Then we see that $f(v_1) = A'$ and $f(v_2) = A''$. Also we know that $A' \subset A'' \subset A$ and there is no other cluster of C_T in between A'' and A' with respect to inclusion. This implies that $\{f(v_1), f(v_2)\}$ is an edge in E_G. Conversely, if $\{f(v_1), f(v_2)\}$ is an edge in E_G, then two edges corresponding to the cuts $(f(v_1) \mid N \setminus f(v_1))$ and $(f(v_2) \mid N \setminus f(v_2))$ have to equal, and the corresponding edge is the edge $\{v_1, v_2\}$ in T. – The case where both v_1 and v_2 are non-leaves of T contained in T_v is similar.

If $v_1 \in T_u$ and $v_2 \in T_v$, then $\{v_1, v_2\}$ is an edge of T if and only if $v_1 = u$ and $v_2 = v$, and this is true if and only if $\{f(v_1), f(v_2)\}$ is an edge in E_G. $\quad\square$

Corollary 2.8. *Let T be a phylogenetic tree. For any two cuts $c_1, c_2 \in C_T$, the cut graphs G_{C_T, c_1} and G_{C_T, c_2} and T are all isomorphic.*

Proof. Immediate consequence of Proposition 2.7. $\quad\square$

Lemma 2.9. *Let C be a phylogenetic set of cuts of N. Let $c = (A \mid B) \in C$ be a cut. Then C is equal to $C_{G_{C,c}}$.*

Proof. Denote the last added edge in the construction of $G_{C,c}$ as $e = \{v, u\}$ and assume without loss of generality that A is the leaf set of component T_v when we remove edge e from $G_{C,c}$. Let $c' = (A' \mid B')$ be an arbitrary cut in C. If $c' = c$, then it is equal to the cut in $C_{G_{C,c}}$ corresponding to the edge e, hence $c' \in C_{G_{C,c}}$.

Assume $c' \neq c$. Because C is phylogenetic, we know that exactly one of the following four statements $A' \subset A$, $A' \subset B$, $B' \subset A$, $B' \subset B$ is true. Without loss of generality, assume that $A' \subset A$. Then $A' \in V(A)$. Let w be the first vertex on the unique path from A' to u. Then we see that the corresponding cut for edge $\{A', w\}$ is $(A' \mid B')$. Hence $(A' \mid B') \in C_{G_{C,c}}$. Because c' was chosen arbitrarily, we conclude that $C \subset C_{G_{C,c}}$.

Now, take any cut $c' = (A' \mid B') \in C_{G_{C,c}}$. If c' is the cut corresponding to the edge e, then $c = \{A \mid B\}$, which implies $c \in C$. Assume c' corresponds to some edge $e' = \{v', u'\}$ in T_v. Without loss of generality, we may suppose that u' is on the unique path from v' to u. Then $c' = (v' \mid N \backslash v')$. Since v' is a cluster of C, we obtain that $c' \in C$. If c' corresponds to some edge in T_u, we proceed analogously. Therefore we get $C_{G_{C,c}} \subset C$ and consequently $C = C_{G_{C,c}}$. $\square$

For any phylogenetic set of cuts C, we can choose a cut $c \in C$. The class of $G_{C,c}$ does not depend on the choice of c, so this construction provides an inverse to $t_{TC} : \mathcal{T}_N \to \mathcal{C}_N$.

2.3   Cuts and Crossings

Proposition 2.10. *For every phylogenetic set C of cuts, the crossing relation X_C is phylogenetic.*

Proof. Assume that i, j, k, l, m are pairwise distinct (but otherwise arbitrary) elements of N. Assume, for the sake of contradiction, that $(i, j \mid k, l)$ and $(i, k \mid j, l)$ are both in X_C. Then there is a cut $(A_1 \mid B_1)$ such that $i, j \in A_1$ and $k, l \in B_1$, and another cut $(A_2 \mid B_2)$ such that $i, k \in A_2$ and $j, l \in B_2$. Then all four sets $A_1 \cap A_2$, $A_1 \cap B_2$, $B_1 \cap A_2$, and $B_2 \cap B_2$ are not empty. This contradicts Axiom (C). Hence the assumption must have been wrong, which proves that Axiom (X1) is fulfilled.

Now assume that $(i, j \mid k, l)$ and $(i, j \mid l, m)$ are both in X_C. By Axiom (C), the set of all clusters that contain both i and j and that do not contain l is totally ordered by inclusion. Let A be the smallest such cluster. Then A does not contain k and does not contain m. Then $(A \mid N \setminus A)$ is a cut with i, j on the left side and k, m in the right side. Hence $(i, j \mid k, m)$ is in X_C, and it follows that Axiom (X2) is fulfilled.

Now assume that $(i, j \mid k, l)$ is in X_C. Then there is a cut $(A \mid B)$ such that $i, j \in A$ and $k, l \in B$. If $m \in A$, then $(i, m \mid k, l)$ is in X_C, and if $m \in B$, then $(i, j \mid l, m)$ is in X_C. It follows that Axiom (X3) is fulfilled, and that X_C is phylogenetic. $\square$

The following definition is only needed for the proof of Lemma 2.13.

Definition 2.11. A *partial cut* of N is a cut of some subset of N. Fix a phylogenetic crossing relation X. We say that a cut or a partial cut $(A \mid B)$ is *compatible* with X if and only if for any distinct $i, j \in A$ and distinct $k, l \in B$, there is a crossing relation $(i, j \mid k, l) \in X$.

Example 2.12. If $(i, j \mid k, l) \in X$, then the cross itself, considered as a partial cut, is compatible with X.

Lemma 2.13. *Let X be a phylogenetic crossing relation on N. Then, for any $(i, j \mid k, l) \in X$. Then there exists a cut $(A \mid B)$ compatible with X such that $i, j \in A$ and $k, l \in B$.*

Proof. We prove that for any n such that $4 \leq n \leq |N|$, there is a partial cut $(A \mid B)$ compatible with X such that $i, j \in A$ and $k, l \in B$, and $|A| + |B| = n$. We proceed by induction on n. For $n = 4$, the statement is trivially true.

Assume $5 \leq n \leq |N|$. By induction hypothesis, there exists a compatible partial cut $(C \mid D)$ such that $i, j \in C$, $k, l \in D$, and $|C| + |D| = n - 1$. Let $m \in N \setminus (C \cup D)$. We claim that either $(C \cup \{m\} \mid D)$ or $(C \mid D \cup \{m\})$ is compatible with X. Assume, for the sake of contradiction, that this claim is wrong. Then there exist $a, b, p \in C$ and $c, r, s \in D$ such that $a \neq b$, $r \neq s$, and the relations $(a, b \mid c, m)$ and $(p, m \mid r, s)$ do not hold. We may also assume $a \neq p$ and $r \neq c$.

Since $(C \mid D)$ is compatible, it follows that $(a, p \mid r, s)$ holds. By Axiom (X3), it follows that $(a, p \mid r, m)$ or $(p, m \mid r, s)$ holds – but we have that $(p, m \mid r, s)$ does not hold, hence we have $(a, p \mid r, m)$. If $b = p$, then $(a, b \mid r, m)$ holds. If $b \neq p$, then we use $(b, p \mid r, s)$ and Axiom (X3) and get $(b, p \mid r, m)$, since $(p, m \mid r, s)$ does not hold. Then, from $(a, p \mid r, m)$ and $(b, p \mid r, m)$, we obtain $(a, b \mid r, m)$ by Axiom (X2). By the compatibility of $(C \mid D)$, we get $(a, b \mid r, c)$. By Axiom (X2), from $(a, b \mid r, c)$ and $(a, b \mid r, m)$, we get $(a, b \mid c, m)$. This contradicts the assumption. $\square$

We can now the define a transformation t_{XC} from sets of cuts to crossing relations.

Definition 2.14. For any phylogenetic crossing relation X on N, we define $t_{XC}(X)$ as the set of all cuts that are compatible with X and denote it as C_X.

Proposition 2.15. *For any phylogenetic crossing relation X, the set C_X of compatible cuts is phylogenetic.*

Proof. Assume, for the sake of contradiction, that C_X does not fulfill Axiom (C). Then there exists cuts $(A_1 \mid B_1)$, $(A_2 \mid B_2)$ in C_X and four elements $i \in A_1 \cap A_2$, $j \in A_1 \cap B_2$, $k \in B_1 \cap A_2$, and $l \in B_1 \cap B_2$. Because $(A_1 \mid B_1)$ is compatible, we have $(i, j \mid k, l) \in X$. Because $(A_2 \mid B_2)$ is compatible, we have $(i, k \mid j, l) \in X$. This contradicts Axiom (X1). $\square$

Theorem 2.16. *The two sets $\mathcal{X}_N$ and $\mathcal{C}_N$ are in bijection: function $t_{XC} : \mathcal{X}_N \to \mathcal{C}_N$ is the inverse of function $t_{CX} : \mathcal{C}_N \to \mathcal{X}_N$.*

Proof. Let $X \in \mathcal{X}_N$ and $i, j, k, l \in N$ pairwise distinct. If $(i, j \mid k, l)$ is in X, then Lemma 2.13 implies that there is a cut $(A \mid B)$ in C_X such that $i, j \in A$ and $k, l \in B$. Therefore (i, j, k, l) is also in X_{C_X}. Conversely, if $(i, j \mid k, l)$ is in X_{C_X}, then there is a cut $(A \mid B)$ in C_X such that $i, j \in A$ and $k, l \in B$. Since $(A \mid B)$ is compatible with X, it follows that $(i, j \mid k, l)$ is in X.

Let $C \in \mathcal{C}_N$ be a phylogenetic set of cuts. Let $(A \mid B)$ be a cut. If $(A \mid B)$ is in C, then all quadruples $(i, j \mid k, l)$ with $i, j \in A$ and $k, l \in B$ are in X_C. Hence $(A \mid B)$ is also in C_{X_C}. Conversely, assume that $(A \mid B)$ is in C_{X_C}. We pick elements $a \in A$ and $b \in B$. Let α be the set of clusters of C that contain a but not b, and let β be the set of clusters of C that contain b but not a. By Axiom (C), both sets α and β are totally ordered by set inclusion. For any $i \in A \setminus \{a\}$ and $j \in B \setminus \{b\}$, the quadruple $(a, i \mid b, j)$ is in X_C because $(A \mid B)$ is compatible with X_C. Then, there must exist a cut $(A' \mid B') \in C$ such that $a, i \in A'$ and $b, j \in B'$. Consequently, we have a cluster $A' \in \alpha$ for every element $i \in A \setminus \{a\}$, and therefore the largest cluster of α is a superset of A. Similarly, we can show that the largest cluster of β is a superset of B.

Let A'' be the smallest cluster of α that is still a superset of A. We claim that $A'' = A$; if this claim is true, then $(A \mid B)$ would be in C, which would finish the proof of the converse and of the whole theorem.

To prove the claim, we assume, for the sake of contradiction, that there is an element $c \in A'' \setminus A = A'' \cap B$. Let α' be the set of clusters of C that contain a but neither b nor c. This set is also totally ordered by set inclusion. For any choice of elements $k \in A \setminus \{a\}$ and $l \in B \setminus \{b\}$, the quadruple $(a, k \mid b, l)$ is in X_C. Hence there exists a cut in $(A''' \mid B''')$ of C such that $a, k \in A'''$ and $b, l \in B'''$. In particular, A''' is in α'. Since we can vary k, it follows that the largest cluster of α' is a superset of A. Hence there is a superset of A in α that does not contain c, which is a contradiction to the fact that the smallest cluster of α that contains A, namely A'', does contain c. $\qquad\square$

2.4   Partitions and Equivalences

In order to prepare for the conversion between partitions and equivalences, we prove a result which could be considered as a kind of converse of Lemma . Let us say that a partition *separates* a triple $\{a, b, c\} \in \binom{N}{3}$ if and only if a, b, and c are in three pairwise distinct parts of the partition.

Theorem 2.17. *Let P be a collection of partitions of N satisfying (P1) such that for every triple in $\binom{N}{3}$, there is a unique partition separating it. Then P is phylogenetic.*

Remark 2.18. The converse is also true: if a collection of partition is phylogenetic, then it is equal to P_T for some phylogenetic tree, by Theorem 2.2. By Lemma 1.16, it follows that every triple is separated by a unique partition in P_T.

In order to prove Theorem 2.17, we need the following proposition.

Lemma 2.19. *Let P be a collection of partitions of N satisfying Axiom (P1) such that for every triple in $\binom{N}{3}$, there is a unique partition separating it. Let p_1, p_2 be two distinct partitions. Then there is a unique pair of parts $A_1 \in p_1$ and $A_2 \in p_2$ such that $A_1 \cup A_2 = N$.*

Moreover, if $B_1 \in p_1$ is any part of p_1 distinct from A_1, and $B_2 \in p_2$ is any part of p_2 distinct from A_2, then $B_1 \subset A_2$, $B_2 \subset A_1$, and $B_1 \cap B_2 = \emptyset$.

Proof. For $i = 1, 2$, let $a_i, b_i, c_i \in N$ be elements from three different parts of p_i. The partition separating $\{a_i, b_i, c_i\}$ is unique, therefore at least two of a_1, b_1, c_1 must be in the same part of p_2; without loss of generality, we may assume that a_1 and b_1 are in the same part. We choose A_2 to be this part. Similarly, we may assume that a_2 and b_2 are in the same part of p_1, and we choose A_1 to be this part.

Suppose that $A_1 \cup A_2 \subsetneq N$. Take any $x \in N \setminus (A_1 \cup A_2)$. Assume, without loss of generality, that x is not in the same part with a_1 in p_1 – otherwise, we exchange a_1 and b_1. Analogously we may assume that x is not in the same part with a_2 in p_2. Then we see that the triple $\{x, a_1, a_2\}$ is separated by both partitions p_1 and p_2. We have our contradiction. It follows that $A_1 \cup A_2 = N$.

The second statement is a consequence of $A_1 \cup A_2 = N$ and the fact that the any part of p_i different from A_i is a subset of $N \setminus A_i$, for $i = 1, 2$. □

Remark 2.20. As a consequence of Lemma 2.19, the set of parts of any set of P fulfills an axiom that is similar to the cluster axiom (C): any two parts, whether they are in the same partition or not, are either contained one in the other, or disjoint, or their union is N.

Proof of Theorem 2.17. We already know that P satisfies Axiom (P1), by assumption.

Axiom (P2): let $a \in N$ be arbitrary. Let $b \in B$ such that $b \neq a$. By Remark 2.20, the set of all parts of any partition that contain a but not b is totally ordered by inclusion. Let A be the smallest such part. We claim that $A = \{a\}$. Assume, for the sake of contradiction, that A contains a second element $c \neq a$. Then the triple $\{a, b, c\}$ cannot be separated by any partition, contradicting the assumption.

Axiom (P3): for the sake of contradiction, assume that A is a part with $|A| \geq 2$ that shows up in two distinct partitions $p_1, p_2 \in P$. By Lemma 2.19, there exists $B \in p_2$ such that $A \cup B = N$. This implies that $|p_2| = 2$, which violates Axiom (P1).

Axiom (P4): Suppose, for the sake of contradiction, that (P4) is not fulfilled. Let A be a part of some partition of p of P, with cardinality at least 2, such that $N \setminus A$ is not a part of any partition. Let a and b be two distinct elements of A. Let $\mathcal{X}$ be the set of all parts that are supersets of $N \setminus A$ and do not contain b.

To show that $\mathcal{X}$ is not empty, we pick an element $e \in N \setminus A$. There must be a partition q separating the triple $\{a, b, e\}$. By Lemma 2.19, there are parts $F \in p$ and $E \in q$ such that $E \cup F = N$. The part $A \in p$ has non-empty intersection with at least two parts of q, namely with the part containing a and with the part containing b. Neither part can be a superset of A. Then, by Remark 2.20, both are subsets of A. This implies $F = A$ and therefore $E \cup A = N$. Then we get $e \in E$ and $b \notin E$. So, $E \in \mathcal{X}$.

By Remark 2.20, $\mathcal{X}$ is totally ordered by inclusion. Let C be the smallest element of $\mathcal{X}$. Since $N \setminus A$ is not a part, there exists an element d in $C \setminus (N \setminus A) = A \cap C$. Now we repeat the argument that we used above with the triple $\{d, b, e\}$.
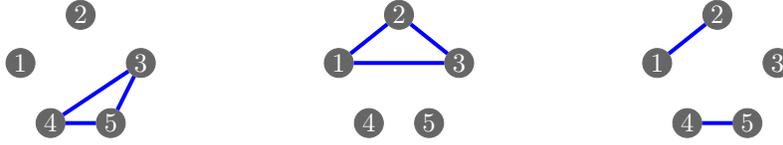
Figure 4: These are the graph G_S defined by the three equivalence classes in Example 1.15. Each of the graphs is a disconnected union of complete graphs, in accordance to Lemma 2.25.

Note that d and b are two distinct elements in A, since d is in some part of $\mathcal{X}$ while b is not. Hence there must be a partition q' separating the triple $\{d, b, e\}$. By Lemma 2.19, there are parts $F' \in p$ and $E' \in q'$ such that $F' \cup E' = N$. Then, with the analogous reasoning, we obtain that the parts containing d and the part containing b in q' are both subsets of A. Hence we have that $F' = A$ and $E' \cup A = N$. Also we observe that $b \notin E'$. Therefore, $E' \in \mathcal{X}$ and $d \notin E'$.

Since $\mathcal{X}$ is totally ordered by inclusion and C is the smallest element of $\mathcal{X}$, we have that $C \subset E'$. This implies $d \in E'$, which is a contradiction.

$\square$

Remark 2.21. It is a fun question to ask what happens to if we replace "triples" by "quadruples", "quintuples" etc. The second author conjectures that there are almost no collections of partitions such that each partition has four parts, and every quadruple is separated by a unique partition. More precisely, any such collection has only a single partition, where every part is a singleton.

In order to convert triples to partitions, we also need one more definition.

Definition 2.22. For any partition p of N, let S_p be the set of all triples that are separated by p.

For any set S of triples, let G_S be the graph with vertex set N, and an edge between $i, j \in N$ if and only if no triples of S contain both i and j. Let p_S be the partition of N defined by the connected components of G_S.

Example 2.23. In Figure 4 we see the graphs G_S when S is one of the three equivalence classes of triples in Example 1.15.

Example 2.24. Let $p = p_a$ be as in Example 1.3. Then S_{p_a} is exactly the equivalence class of triples E_a in Example 1.18. Moreover, p_{E_a} is the partition p_a.

Lemma 2.25. *Let S be a diverse set of triples. Then G_S is a disconnected union of complete graphs.*

Proof. Let $i, j, k \in N$ be three distinct vertices of G_S. Suppose that $\{i, j\}$, $\{j, k\}$ are edges of G_S, but $\{i, k\}$ is not an edge. Then there exists $l \in N \setminus \{i, j, k\}$ such that $\{i, k, l\} \in S$. By Axiom (D1), one of the triples $\{i, j, l\}$, $\{j, k, l\}$, or $\{i, j, k\}$ is in S. This violates the fact that $\{i, j\}$ and $\{j, k\}$ are edges. This shows that

any two vertices in the same connected component of G_U are connected by an edge; the graph is a disconnected union of complete graphs. $\square$

Lemma 2.26. *For any partition p with at least three parts, we have that S_p is diverse and $p_{S_p} = p$.*

For any diverse set S of triples, the partition p_S has at least three parts, and we have $S_{p_S} = S$.

Proof. Let p be a partition with at least three parts. Then we know that S_p is non-empty. Assume that the triple $\{i, j, k\}$ is in S_p, which means that i, j, and k are in three distinct parts. A fourth element $l \in N$ can at most be in one of this three parts, hence p separates l and two other elements out of i, j, and k. Therefore S_p satisfies (D1).

Let $a, b, c, x, y, z \in N$. Assume that p separates the triples $\{a, x, y\}$, $\{b, y, z\}$, and $\{c, x, z\}$. Then x, y, and z are in pairwise distinct parts, so p also separates $\{x, y, z\}$. Therefore S_p satisfies (D2), hence S_p is diverse. Moreover, i, j are in the same part of p if and only if no triples in S_p contain both i and j if and only if i and j are in the same component of G_{S_p} if and only if i and j are in the same part of p_{S_p}. Hence $p_{S_p} = p$.

Let S be any diverse set of triples. If a triple $\{i, j, k\}$ is in S, then none of $\{i, j\}$, $\{i, k\}$ or $\{j, k\}$ is an edge of G_S. By Lemma 2.25, we obtain that i, j, k are in pairwise distinct components in G_S. Hence, i, j, k are in pairwise distinct parts of p_S. Therefore, $\{i, j, k\} \in S_{p_S}$. For the other direction, let $\{i, j, k\}$ be any triple in S_{p_S}. This indicates that i, j, k are in three pairwise distinct parts in p_S, i.e., i, j, k are in three pairwise distinct components of the graph G_S. Therefore, none of $\{i, j\}$, $\{i, k\}$, $\{j, k\}$ is an edge of G_S. Hence, S contains triples $\{i, j, a\}$, $\{i, k, b\}$ and $\{j, k, c\}$ for some $a, b, c \in N$. By Axiom (D2), $\{i, j, k\} \in S$. Hence $S_{p_S} = S$.

$\square$

Now we define the conversion from equivalences on triples to collections of partitions. For any phylogenetic equivalence relation E of triples in N, we define P_E as the collection of all partitions p_U for any equivalence class U of E. The function from $\mathcal{E}_N$ to $\overline{\mathcal{P}_N}$ that maps E to P_E is denoted by t_{EP}.

Lemma 2.27. *Let E be a phylogenetic equivalence relation of triples in N. Then P_E is a phylogenetic set of partitions.*

Proof. Every equivalence class U contains at least one triple. This triple is separated by p_U, and it follows that p_U must have at least three parts. Therefore P_E satisfies Axiom (P1).

Every triple $\tau = \{i, j, k\} \in \binom{N}{3}$ is in a unique equivalence class U, and p_U separates τ by Lemma 2.26. Moreover, if V is any equivalence class such that p_V separates τ, then $\tau \in S_{p_V} = V$, and therefore $U = V$. This implies that every triple is separated by a unique partition in P_E. By Theorem 2.17, P_E is phylogenetic. $\square$
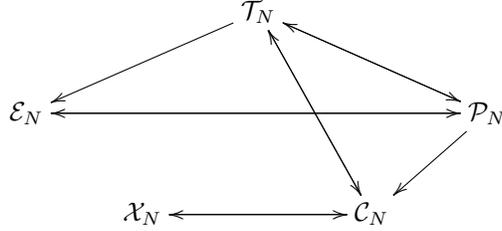
Figure 5: This diagram shows all conversion maps between different types of structures that have been defined in this paper. We also have seen that the triangles are commutative.

For any phylogenetic collection P of partitions, each triple $\tau \in \binom{N}{3}$ is separated by a unique partition in P, by Remark 2.18. We define E_P as follows: Two triples τ_1 and τ_2 are equivalent if and only if the unique partition separating τ_1 is the same as the unique partition separating τ_2. The function from $\mathcal{P}_N$ to $\mathcal{E}_N$ that maps P to E_P is denoted by t_{PE}. It is straightforward to see that $t_{TE} = t_{PE} \circ t_{TP}$.

Lemma 2.28. *Let P be a phylogenetic set of partitions of N. Then E_P is a phylogenetic equivalence relation of triples in $\binom{N}{3}$.*

Proof. For any partition p in P, the set of triples S_p is diverse by Lemma 2.26; hence Axiom (E0) is fulfilled. □

Theorem 2.29. *The two sets $\mathcal{P}_N$ and $\mathcal{E}_N$ are in bijection: function t_{EP} : $\mathcal{E}_N \to \mathcal{P}_N$, $E \mapsto P_E$ is the inverse of function $t_{PE} : \mathcal{P}_N \to \mathcal{E}_N$, $P \mapsto E_P$.*

Proof. For any $E \in \mathcal{E}_N$, any class U of E is diverse. Then, by Lemma 2.26 we have that $S_{p_U} = U$. We see that $\{S_{p_U}\}_{U \in E}$ is exactly E_{P_E} – where the foot index $U \in E$ means that U is a class of E. Hence, we have $E_{P_E} = E$.

For any $P \in \mathcal{P}_N$, each partition $p \in P$ has at least three parts. By Lemma 2.26, we know that $p_{S_p} = p$. Since $\{p_{S_p}\}_{p \in P}$ is exactly the partition collection P_{E_P}, we obtain that $P_{E_P} = P$. □

In Figure 5, we display the diagram consisting of all conversion algorithms in this paper.

Acknowledgement

thank Antonio Jiménez-Pastor for the discussion on the conversion from a phylogenetic crossing relation to a phylogenetic equivalence relation on the triples.

References

[1] A.D. Scott and D.A. Baum. Phylogenetic Tree. *Encyclopedia of Evolutionary Biology*, 2016.

[2] M. A. Ragan. Phylogenetic inference based on matrix representation of trees. *Molecular Phylogenetics and Evolution, Volume 1, Issue 1*, March 1992, Pages 53-58.

[3] L. J. Billera, S. P. Holmes and K. Vogtmann. Geometry of the Space of Phylogenetic Trees. *Advances in Applied Mathematics* **27**, 733-767 (2001).

[4] D. Bryant and J. Lagergren. Compatibility of unrooted phylogenetic trees is FPT. *Theoretical Computer Science, Volume 351, Issue 3*, 28 February 2006, Pages 296-302.

[5] F. F. Knudsen. The projectivity of the moduli space of stable curves II. The stacks $M_{g,n}$. *Math. Scand.* **52** 161-199 (1983).

[6] S. Vakati and D. Fernández-Baca. Graph triangulations and the compatibility of unrooted phylogenetic trees. *Applied Mathematics Letters, Volume 24, Issue 5*, May 2011, Pages 719-723.

Technical Reports of the Doctoral Program
"Computational Mathematics"

2020

2020-01   N. Smoot: *A Single-Variable Proof of the Omega SPT Congruence Family Over Powers of 5* Feb 2020. Eds.: P. Paule, S. Radu

2020-02   A. Schafelner, P.S. Vassilevski: *Numerical Results for Adaptive (Negative Norm) Constrained First Order System Least Squares Formulations* March 2020. Eds.: U. Langer, V. Pillwein

2020-03   U. Langer, A. Schafelner: *Adaptive space-time finite element methods for non-autonomous parabolic problems with distributional sources* March 2020. Eds.: B. Jüttler, V. Pillwein

2020-04   A. Giust, B. Jüttler, A. Mantzaflaris: *Local (T)HB-spline projectors via restricted hierarchical spline fitting* March 2020. Eds.: U. Langer, V. Pillwein

2020-05   K. Banerjee, M. Ghosh Dastidar: *Hook Type Tableaux and Partition Identities* June 2020. Eds.: P. Paule, S. Radu

2020-06   A. Bostan, F. Chyzak, A. Jiménez-Pastor, P. Lairez: *The Sage Package* comb_walks *for Walks in the Quarter Plane* June 2020. Eds.: M. Kauers, V. Pillwein

2020-07   A. Meddah: *A stochastic multiscale mathematical model for low grade Glioma spread* June 2020. Eds.: E. Buckwar, V. Pillwein

2020-08   M. Ouafoudi: *A Mathematical Description for Taste Perception Using Stochastic Leaky Integrate-and-Fire Model* June 2020. Eds.: E. Buckwar, V. Pillwein

2020-09   A. Bostan, A. Jiménez-Pastor: *On the exponential generating function of labelled trees* July 2020. Eds.: M. Kauers, V. Pillwein

2020-10   J, Forcan, J. Qi: *How fast can Dominator win in the Maker-Breaker domination game?* July 2020. Eds.: V. Pillwein, J. Schicho

2020-11   J. Qi: *A calculus for monomials in Chow group of zero cycles in the moduli space of stable curves* Sept 2020. Eds.: P. Paule, J. Schicho

2020-12   J. Qi, J. Schicho: *Five Equivalent Ways to Describe a Phylogenetic Tree* November 2020. Eds.: M. Kauers, S. Radu

Doctoral Program

"Computational Mathematics"

Director:

Assoc. Prof. Dr. Veronika Pillwein
Research Institute for Symbolic Computation

Deputy Director:

Prof. Dr. Bert Jüttler
Institute of Applied Geometry

Address:

Johannes Kepler University Linz
Doctoral Program "Computational Mathematics"
Altenbergerstr. 69
A-4040 Linz
Austria
Tel.: ++43 732-2468-6840

E-Mail:

office@dk-compmath.jku.at

Homepage:

http://www.dk-compmath.jku.at