Electronic Journal of Graph Theory and Applications **9** (2) (2021), 499–505

Electronic Journal of
Graph Theory and Applications

Representing non-crossing cuts by phylogenetic trees

Thomas Lange

University of Applied Sciences Mittweida
Technikumplatz 17, 09648 Mittweida, Germany

tlange10@hs-mittweida.de

Abstract

Phylogenetic trees are representations of the evolutionary descendency of a set of species. In graph-theoretic terms, a phylogenetic tree is a partially labeled tree where unlabeled vertices have at least degree three and labels corresponds to pairwise disjoint subsets of the set of species. A cut of a graph $G = (V, E)$ is defined as bipartition $\{S, V \setminus S\}$ of the vertex set V of G. A pair of cuts $\{S, \bar{S}\}$, $\{T, \bar{T}\}$ is said to be crossing, if neither $S \cap T$, $S \cap \bar{T}$, $\bar{S} \cap T$ nor $\bar{S} \cap \bar{T}$ is empty. In this paper, we show that each set of pairwise non-crossing cuts of a graph G can be represented uniquely by a phylogenetic tree such that the set of species corresponds to the vertex set of G.

Keywords: pairwise laminar cutset, phylogenetic tree, edge-bipartition
Mathematics Subject Classification : 05C30
DOI: 10.5614/ejgta.2021.9.2.20

1. Introduction

By the evolutionary theory, existing biological species are linked by common ancestors. Studying these ancestor relations leads to phylogenetic trees as graphical representations of the postulated evolutionary relationship of a specified set of organisms. In graph-theoretic terms, a phylogenetic tree is a graph-theoretic tree together with a mapping of the set $\{1, \dots, n\}$ of labels to the vertices of the tree such that all vertices with degree less than 3 are images. Note, that some vertices may obtain several labels while other vertices (with degree at least 3) may obtain no label.

Received: 4 June 2018, Revised: 30 December 2020, Accepted: 19 March 2021.

There had been done extensive research on how to find the evolutionary most probable phylogenetic tree for a given set of species [1, 2, 3, 4, 5, 6, 10]. Among those studies, several focused on how to construct all phylogenetic trees and how to count them [2, 3, 4, 5, 6].

Given a graph $G = (V, E)$, an edge-cut of the graph can be characterized by the set of vertices on both sides of the edge-cut. Two cuts are said to be crossing, if their vertex sets intersect on both shores of the edge-cuts. Otherwise, the cuts are said to be non-crossing or laminar [7, 9]. The concept of sets of laminar cuts has been used to show certain coverings of graphs and hypergraphs.

In Section 2 of this paper bipartitions of the label set of phylogenetic trees induced by the edges of the tree are introduced. We will call those bipartitions edge-bipartitions of the phylogenetic tree. In Section 3 we will show that each phylogenetic tree is uniquely determined by its set of edge-bipartitions by giving a constructive proof. Further, some properties of the set of edge-bipartitions for phylogenetic trees resulting from each other by merging a single edge are obtained. In Section 4 we will then show that the set of edge-bipartitions of a phylogenetic tree forms a set of laminar cuts and further, that each set of laminar cuts corresponds to the set of edge-bipartitions of exactly one phylogenetic tree. In Section 5 we will summarize our results and pose some questions related to the shown combinatorial identity between phylogenetic trees and sets of laminar cuts.

2. Preliminaries

A phylogenetic tree is a tree $T = (V, E)$ together with a mapping $\phi : X \to V$ such that all all vertices with degree less than three are images of ϕ. We denote X as the label set of the phylogenetic tree, vertices in $Im(\phi)$ are called labeled vertices, vertices not in $Im(\phi)$ are called unlabeled vertices. For an edge $e \in E$, the edge-bipartition of $e = \{u, v\}$, denoted by $X_e(T)$, is the bipartition $\{X_e^1, X_e^2\}$ with $X_e^1 \cup X_e^2 = X$ such that all vertices of X_e^i, $i = 1, 2$, are in the same connected component of $T - e$. If we do not specify further, we will consider both possible enumerations for the blocks of $X_e(T)$. $X_e^u(T)$ denotes the block such that the vertices labeled by the set X_e^u are in the same connected component as u (u does not necessarily be labeled itself), $X_e^{\bar{u}}$ will denote the block of the connected component which does not contain the vertex u. The set $\mathcal{E}(T)$ denotes the set of all edge-bipartitions of T. We will omit T whenever there is no ambiguity. A cut of a graph $G = (V, E)$ is a vertex bipartition $\{S, V \setminus S\}$. Two cuts $\{T, V \setminus T\}$ and $\{S, V \setminus S\}$ are called crossing, if neither $S \cap T$, $S \cap \bar{T}$, $\bar{S} \cap T$ nor $\bar{S} \cap \bar{T}$ is empty. Otherwise they are called non-crossing or laminar. For a graph $G = (V, E)$ and an edge $e \in E$, G/e denotes the graph resulting from G by identifying both endvertices of the edge e. For phylogenetic trees, the mapping ϕ will then also identify both endvertices of e.

3. Phylogenetic trees and edge-bipartitions

In this section we will show that each phylogenetic tree can be uniquely characterized by its set of edge-bipartitions.

First, we will show that the set of edge-bipartitions of a phylogenetic tree is indeed a set and not a multiset.

Lemma 3.1. *Let $T = (V, E)$ be a phylogenetic tree and let $e, f \in E$ be two edges of T, $e \neq f$. Then $X_e \neq X_f$ holds.*

 www.ejgta.org

Proof. Since T is a tree, $T - e - f$ contains exactly three connected components. Let V_1,V_2 and V_3 denote the vertex sets of those components and let $X_i = X \cap V_i$, $i = 1, 2, 3$, be the corresponding subsets of X in those components. Without loss of generality, the bipartitions X_e and X_f have then the following representation: $X_e = \{X_1, X_2 \cup X_3\}$ and $X_f = \{X_1 \cup X_2, X_3\}$. Thus, it remains to show that the set X_2 is not empty. Consider an arbitrary vertex u on the path between the edges e and f. The vertex u clearly belongs to V_2. Now two cases are possible: If u is a labeled vertex, it holds $u \in X_2$ and thus X_2 is not empty. If u is an unlabeled vertex, it has degree at least three. Hence, $T - u$ has at least three components. The component not containing e and f is a tree consisting only of vertices of V_2. Since all leafs in T are labeled - since they do not have a degree at least three - this subtree needs to contain a labeled vertex and thus X_2 is not empty. This completes the proof that X_2 is not empty and thus that the edge-partitions X_e and X_f are not equal. $\square$

Next, we will show an interesting property concerning the edge-bipartition of an edge e and the edge-bipartitions of the other edges which have the same endpoint with e in common. This property will then provide the main argument for the unique construction of the phylogenetic tree given its set of edge-bipartitions.

Lemma 3.2. *Let $T = (V, E)$ be a phylogenetic tree, let $e = \{u, v\} \in E$ be an edge of T and let $\Gamma(u)$ denote the set of edges incident to u in T.*

- *If u is an unlabeled vertex, the set $F = \Gamma(u) \setminus \{e\}$ fulfills $X_e^u = \bigcup_{i \in F} X_i^{\bar{u}}$. Further, for each set $F' \subseteq E \setminus \{e\}$ with $X_e^u = \bigcup_{i \in F'} X_i^1$ holds $|F'| \leq |F| \Rightarrow F' = F$.*
- *If u is a labeled vertex, let U denote the label set of the vertex u. Then, for each edge $f \in E \setminus \{e\}$ it holds $X_f^u \cap X_e^{\bar{u}} \neq \emptyset$.*
  Further, for the set $F = \Gamma(u) \setminus \{e\}$ holds $X_e^u = U \cup \bigcup_{i \in F} X_i^{\bar{u}}$ and for each set $F' \subseteq E \setminus \{e\}$ with $X_e^u = U \cup \bigcup_{i \in F'} X_i^1$ holds $|F'| \leq |F| \Rightarrow F' = F$.

For unlabeled vertices this implies that there is a set of edges whose edge-bipartition have as union the set X_u^e and that the unique minimal set with this property is $\Gamma(u) \setminus \{e\}$. The first part of the statement for labeled vertices implies that there is no set $F' \subseteq E \setminus \{e\}$ with $X_e^u = \bigcup_{i \in F'} X_i^1$ since each set disjoint to $X_e^{\bar{u}}$ will miss the label set of the vertex u. The second part then implies that there exist sets where the label set of u is the only missing set in the union and that further $\Gamma(u) \setminus \{e\}$ is the unique minimal set with the desired property.

Proof. Let (V_1, E_1) and (V_2, E_2) denote the connected components of $T - e$ with $u \in V_1$.
First we will show that e is the only edge separating u and $X_e^{\bar{u}}$. For each edge $f \in E_1$, u and all vertices of V_2 are in the same connected components. Thus, those edges do not separate u and $X_e^{\bar{u}}$. For each edge $f \in E_2$ all vertices of V_1 are in the same connected component and thus $X_e^u \subseteq X_f^u$. However, since by Lemma 3.1 different edges induce different edge-partitions, $X_e^u \neq X_f^u$ and thus $X_f^u \cap X_e^{\bar{u}} \neq \emptyset$. This completes the first part of the proof which corresponds to the first statement for labeled vertices u.

www.ejgta.org

Let (V_1^f, E_1^f) and (V_2^f, E_2^f) denote the connected components for $f \in F = \Gamma(u) \setminus \{e\}$ such that $V_1^f \subseteq V_1$ for all $f \in F$. It clearly holds $\bigcup_{i \in F} V_1^i = V_1 \setminus \{u\}$. Thus, considering the corresponding labeled vertex sets, it follows that if u is unlabeled, $\bigcup_{i \in F} X_i^{\bar{u}} = X_e^u$ holds and if u is labeled, $\bigcup_{i \in F} X_i^{\bar{u}} = X_e^u \setminus U$ holds. This completes the proof of the statements concerning $F = \Gamma(u) \setminus \{e\}$.
It remains to show that F is the minimal set which allows the representation of X_e^u as the union of blocks of other edge partitions. First note for edges in E_2 both blocks contain vertices of $X_e^{\bar{u}}$. Thus, we are only concerned with edgesets $F' \subseteq E_1$.
Consider an arbitrary edge $f \in E_1 \setminus F$ and let e' denote the edge in F on the path from u to f. Let X_f^1 denote the block which does not contain vertices of $X_e^{\bar{u}}$. Then by considering the tree structure, it is obvious that $X_f^1 \subseteq X_{e'}^1$. However, by Lemma 3.1 different edges of the same tree have different edge-partitions, which implies $X_f^1 \neq X_{e'}^1$. Thus, whenever we do not choose an edge of F, we have to choose at least two edges in the corresponding subtree instead and thus F is the only set with minimal cardinality and the desired property. □

Theorem 3.1. *Let $T = (V, E)$ be a phylogenetic tree and let $\mathcal{E}$ denote the set of edge-partitions of T. Then T can be uniquely constructed from the set $\mathcal{E}$.*

Note that by 3.1 different edges have different partitions and thus $\mathcal{E}$ contains exactly one edge-partition for each edge in T. In this proof, we will use $\{u\}$ to denote the label set of a labeled vertex u. Please note that $\{u\}$ might have a cardinality greater than 1.

Proof. First note, that for each leaf u, there exists an edge $f \in E$ with $X_f = \{\{u\}, X \setminus \{u\}\}$. Further, if $u \in X$ is not a leaf, there exists no edge-partition such that $\{u\}$ is a block. Thus, we can derive the set of leafs of T from the set $\mathcal{E}$.
We now will give a procedure which uses a current root vertex u, a given edge-bipartition whose edge e will be incident to u in the resulting tree and the remaining set of edge-bipartitions to construct the subtree attached to the root vertex u by the edge e.
We will start with an arbitrary leaf u and remove its edge-partition $\{\{u\}, X \setminus \{u\}\}$ from the set $\mathcal{E}$.
Construction step: Check, if $X \setminus \{u\}$ can be represented as union of blocks in $\mathcal{E}$.
If that is possible, by Lemma 3.2 the vertex adjacent to the root vertex u, which we will denote by v, cannot be labeled and thus, we add an unlabeled vertex v and the edge $\{v, u\}$ to the constructed tree. Further, by Lemma 3.2 there exists exactly one minimal set of partitions, $\mathcal{F} \subseteq \mathcal{E}$, with this property and $\mathcal{F}$ consists of edge partitions induced by edges incident to v.
If that union representation is not possible, by Lemma 3.2 the vertex adjacent to our root vertex must be labeled. So for each vertex $v \in X \setminus \{u\}$ we check whether we can represent $X \setminus \{u, v\}$ by the union of blocks of partitions in $\mathcal{E}$. By Lemma 3.2 this will be possible for exactly one vertex v, which is the neighbor of our root vertex u so we add v and the edge $\{u, v\}$ to our constructed tree. Further, there exists a unique minimal description for $X \setminus \{u, v\}$ as union of partition-blocks. The corresponding set of partitions $\mathcal{F}$ again correspond to the edge-partitions induced by the edges incident to v.
In any case we added one edge and one vertex to our constructed tree and know the blocks of the edge-partitions of the edges incident to our newly added vertex v. So iteratively, for each of those

www.ejgta.org

blocks we check whether we can represent them as unions of other blocks in $\mathcal{E}$ and repeat the construction step where now v corresponds to the root vertex and the corresponding block in $\mathcal{F}$ is considered instead of $X \setminus \{u\}$.
Since in each construction step we add one edge to our graph and our graph contains exactly $|\mathcal{E}|$ edges, the procedure will terminate. Thus, we have given a procedure to construct T from $\mathcal{E}$ in a finite number of steps which completes the proof of the theorem. □

We have now shown that the set of edge-bipartitions $\mathcal{E}$ of a phylogenetic tree T uniquely describes the tree T. We will now demonstrate some results how modifications of the set $\mathcal{E}$ influence the structure of the corresponding tree.

Lemma 3.3. *Let T be a phylogenetic tree and let e be an edge of T. Let $\mathcal{E}(G)$ denote the set of edge-bipartitions of G for $G = T, T/e$ respectively. Then it holds*

$$\mathcal{E}(T/e) = \mathcal{E}(T) \setminus \{X_e\}.$$

Proof. Let f be an arbitrary edge of T with $f \neq e$. Then, $X_f(T) = X_f(T/e)$ since the labels of X will be in the same connected components in $T - f$ and $T/e - f$. Thus, since for each edge f besides e, the edge-bipartitions coincide. The result of the lemma follows immediately. □

Lemma 3.4. *Let T be a phylogenetic tree. Let $U \subset X$ be a label subset such that for each edge-bipartition $\{X_1, X_2\}$ holds that $U \subseteq X_1$. Let $\mathcal{E}$ denote the set of edge-bipartitions of T. Then, there exists a phylogenetic tree T' with set of edge-bipartitions $\mathcal{E}'$ such that*

$$\mathcal{E}' = \mathcal{E} \cup \{U, X \setminus U\}.$$

Proof. Note that by the consideration of both enumerations of the blocks the restriction that $U \subseteq X_1$ for each edge-bipartition $\{X_1, X_2\}$ means that for each edge f of T all labels in U are in the same connected component in $T - f$. Thus, U is a subset of the labels of one labeled vertex w (it is further easy to see that each non-empty subset of the label set of a single labeled vertex has the desired property). Let W denote the set of all labels of the vertex w. Consider the partially labeled tree T' resulting from T by relabeling the vertex w with the set $W \setminus U$ and adding a new vertex u with label set U and the edge $e = \{u, w\}$. There are two cases: If $W \setminus U$ is non-empty or $W \setminus U$ is empty and w has degree at least two in T, then T' is a phylogenetic tree. Then it is easy to see that T results from T' by contraction of the edge e which further has $X_e = \{U, X \setminus U\}$. Thus, T' is the desired phylogenetic tree by Lemma 3.3. If on the other hand $W \setminus U$ is empty and w has degree one, then let f be the edge incident to w. Since w is a leaf, it holds $X_f = \{W, X \setminus W\} = \{U, X \setminus U\}$. It follows $\mathcal{E} \cup \{U, X \setminus U\} = \mathcal{E}$ and T itself is the desired phylogenetic tree. □

4. Sets of non-crossing cuts and phylogenetic trees

In this section we will show that the set of edge-bipartitions of any phylogenetic tree forms a set of non-crossing cuts. Further, we will show that for each set of pairwise non-crossing cuts on a given ground set there exists one phylogenetic tree which has this set as its set of edge-bipartitions.

Theorem 4.1. *Let $T = (V, E)$ be a phylogenetic tree and let $e, f \in E$ be two edges of T, $e \neq f$. Then the edge-bipartitions of X_e and X_f are non-crossing.*

Proof. This result basically follows from the proof of Lemma 3.1. There we showed that $X_e = \{X_1, X_2 \cup X_3\}$ and $X_f = \{X_1 \cup X_2, X_3\}$ for suitable disjoint choices of X_1, X_2 and X_3. It follows immediately that $X_e^1 \cap X_f^2 = X_1 \cap X_3 = \emptyset$. Thus, X_e and X_f are non-crossing. □

Theorem 4.2. *Let $\mathcal{C}$ be a non-empty set of pairwise non-crossing cuts of a set X. Then there exists a phylogenetic tree T with label ground set X such that $\mathcal{C} = \mathcal{E}(T)$.*

Proof. We will use a proof by induction on the size of the set $\mathcal{C}$. Assume $\mathcal{C}$ contains exactly one element, the bipartition $\{X_1, X_2\}$. Then the phylogenetic tree T with two vertices labeled X_1 and X_2 joined by an edge has $\mathcal{C}$ as its set of edge-bipartitions.
Now let $\mathcal{C}$ be a set of cardinality at least two and assume that the theorem holds for all sets of pair-wise non-crossing cuts of X with fewer than $|\mathcal{C}|$ elements.
Let $X_S = \{S, X \setminus S\}$ be a cut in $\mathcal{C}$ such that no other cut in $\mathcal{C}$ has S as superset of one of its sets. Such a cut X_S clearly exists.
Now there are two cases possible: Either, for each other cut $X_T = \{T, X \setminus T\}$ without loss of generality holds $S \subseteq T$, or there exists some cut $X_T = \{T, X \setminus T\}$ such that $S \cap T \neq \emptyset$ and $S \cap (X \setminus T) \neq \emptyset$.
We will assume the latter, and we will show that then X_S and X_T are crossing. We already know that $S \cap T \neq \emptyset$ and $S \cap (X \setminus T) \neq \emptyset$ holds in this case. Further, by the definition of X_S holds $T \not\subset S$ and $(X \setminus T) \not\subset S$. Thus, $(X \setminus S) \cap T \neq \emptyset$ and $(X \setminus S) \cap (X \setminus T) \neq \emptyset$ and it follows that X_S and X_T must be crossing, a contradiction to the definition of the set $\mathcal{C}$.
Thus, for all other cuts $X_T = \{T, X \setminus T\} \in \mathcal{C}$ holds $S \subseteq T$. We will now consider the set $\mathcal{C}'$ obtained by removing the cut X_S from $\mathcal{C}$. Since $\mathcal{C}'$ has fewer elements than $\mathcal{C}$, by induction there exists a phylogenetic tree T' which has $\mathcal{C}'$ as its set of edge-bipartitions. Further, for all edge-partitions of $X_T = \{T, X \setminus T\}$ of $\mathcal{C}'$ holds $S \subseteq T$. By Lemma 3.4, there exists a tree which has $\mathcal{C}' \cup \{S, X \setminus S\} = \mathcal{C}$ as its set of edge-bipartitions which completes the induction. □

5. Conclusion and open problems

By introducing the concept of edge-bipartitions of phylogenetic trees we could show that there exists a one-to-one correspondence between sets of pairwise non-crossing cuts and phylogenetic trees. While this combinatorial identity is interesting in itself, it poses several new questions which may lead to useful applications:

1. Is there a combinatorial relation between sets of crossing cuts and phylogenetic trees?
2. Can we use phylogenetic trees to represent all minimum cuts of a graph?
3. Which graph classes have a small set of pairwise non-crossing cuts and thus yield a phylogenetic tree with relatively few vertices?
4. Can we use graph cuts for a compact representation of a phylogenetic tree arising from evolutionary problems?

There are other interesting applications related to phylogenetic trees. For example did Lucet, Carlier and Manouvrier [8] obtain phylogenetic trees when considering the splitting classes for the two-edge connected reliability. The concept of edge-bipartitions can be easily used to proof that those splitting classes are minimal, i.e. that there are no two classes which describe the same connectivity case – a result missing in [8] which we will present in a forthcoming paper.

Acknowledgement

This research was supported by an ESF grant.

References

[1] R. Alberich, G. Cardona, F. Rosselló, and G. Valiente, An algebraic metric for phylogenetic trees, *Appl. Math. Lett.* (2009), 1320–1324.

[2] J. Felsenstein, The number of evolutionary trees, *Systematic Biology* **27** (1) (1978), 27–33.

[3] W.M. Fitch and E. Margoliash, Construction of phylogenetic trees, *Science* **155** (3760) (1967), 279–284.

[4] C. Flight, How many stemmata?, *Manuscripta* **34** (2) (1990), 122–128.

[5] L.R. Foulds and R.W. Robinson, Enumeration of binary phylogenetic trees, *Combinatorial Mathematics VIII* (1981), 187–202.

[6] L.R. Foulds and R.W. Robinson, Enumerating phylogenetic trees with multiple labels, *Discrete Math.* **72** (1-3) (1988), 129–139.

[7] L. Lovasz, On two minimax theorems in graph, *J. Combin. Theory Ser. B* **21** (2) (1976), 96–103.

[8] C. Lucet, J-F. Manouvrier, and J. Carlier, Evaluating network reliability and 2-edge-connected reliability in linear time for bounded pathwidth graphs, *Algorithmica* **27** (2000), 316–336.

[9] I.P. McWhirter and D.H. Younger, Strong covering of a bipartite graph, *J. Lond. Math. Soc.* s2-3 (1) (1971), 86–90.

[10] D.F. Robinson and L.R. Foulds, Comparison of phylogenetic trees, *Mathematical Biosciences* **53** (1-2) (1981), 131–147.