Article:
Seemann, CR and Hellmuth, M orcid.org/0000-0002-1620-5508 (2018) The matroid structure of representative triple sets and triple-closure computation. European Journal of Combinatorics, 70. pp. 384-407. ISSN 0195-6698

https://doi.org/10.1016/j.ejc.2018.02.013

The Matroid Structure of Representative Triple Sets and Triple-Closure Computation

Marc Hellmuth and Carsten R. Seemann

Dpt. of Mathematics and Computer Science, University of Greifswald, Walther- Rathenau-Strasse 47, D-17487 Greifswald, Germany

Saarland University, Center for Bioinformatics, Building E 2.1, P.O. Box 151150, D-66041 Saarbrücken, Germany

Email: mhellmuth@mailbox.org

Abstract

The closure $\mathrm{cl}(R)$ of a consistent set R of triples (rooted binary trees on three leaves) provides essential information about tree-like relations that are shown by any supertree that displays all triples in R. In this contribution, we are concerned with representative triple sets, that is, subsets R' of R with $\mathrm{cl}(R') = \mathrm{cl}(R)$. In this case, R' still contains all information on the tree structure implied by R, although R' might be significantly smaller. We show that representative triple sets that are minimal w.r.t. inclusion form the basis of a matroid. This in turn implies that minimal representative triple sets also have minimum cardinality. In particular, the matroid structure can be used to show that minimum representative triple sets can be computed in polynomial time with a simple greedy approach. For a given triple set R that "identifies" a tree, we provide an exact value for the cardinality of its minimum representative triple sets. In addition, we utilize the latter results to provide a novel and efficient method to compute the closure $\mathrm{cl}(R)$ of a consistent triple set R that improves the time complexity $\mathcal{O}(|R||L_R|^4)$ of the currently fastest known method proposed by Bryant and Steel (1995). In particular, if a minimum representative triple set for R is given, it can be shown that the time complexity to compute $\mathrm{cl}(R)$ can be improved by a factor up to $|R||L_R|$. As it turns out, collections of quartets (unrooted binary trees on four leaves) do not provide a matroid structure, in general.

Keywords: Rooted Triple, Closure, Matroid, Ahograph, BUILD, Greedy, Phylogeny, Quartet

1 Introduction

Inference of phylogenetic relationships between genes or species based on genomic sequence information is one of the main issues in phylogenomics [52]. The evolutionary history of genes and species is usually represented as a tree. One of the possible building blocks for the reconstruction of the histories of both, genes and species, are provided by triples (rooted binary trees on three leaves) [8, 16, 24, 28, 30, 32, 38, 42, 44, 56, 61]. Such triples can be obtained directly from sequence data and are combined to a "supertree" that provides then the information of the history of the respective genes or species [9–12, 21, 23, 29, 31, 39, 41, 43]. In this contribution, we consider *consistent* sets R of triples, that is, all triples of R fit into a common supertree, which enforces further tree-like relations to hold [6, 13, 25]. This allows one to define a *closure operation* $\mathrm{cl}(R)$ for R that comprises all triples that are displayed by every proper supertree for R. The closure of sets of rooted or unrooted trees has been extensively studied in the last decades [4–6, 13, 15, 25, 34] and has various applications in phylogenomics [19, 20, 31, 35, 46, 47, 53, 55, 62].

Here, we are particularly interested in the computation of the closure and *representative* sets R' for consistent triple sets R, that is, subsets R' of R that satisfy $\mathrm{cl}(R') = \mathrm{cl}(R)$. Such representative sets R' are of particular interest, since on the one hand, they can reduce the space complexity to store all information on the tree-like relationships that is also provided by R and, on the other hand, will significantly improve the time complexity to compute the closure, as we shall see later. Natural optimization problems within this context aim at finding representative sets R' that are minimal w.r.t. inclusion or have minimum size among all representative subsets

of R. Grünewald, Steel and Swenson [25] established important results to the latter problems. In particular, they characterized minimal representative triple sets $R' \subseteq R$ for the case that R "identifies" a given tree T and gave lower bounds $B(T)$ on their cardinalities. Moreover, Mike Steel showed that all minimal "tree-defining" sets of rooted triples must have the same size [58]. However, for an arbitrary consistent triple set R it is still unclear whether the (decision version of the) problem of finding a representative subset $R' \subseteq R$ of minimum size is NP-complete or polynomial-time solvable.

In this contribution, we show that minimum representative subset $R' \subseteq R$ can be computed in polynomial time. To this end, we show that minimal representative sets $R' \subseteq R$ form the basis of the matroid $(R, \mathbb{F}_R)$ [40, 50]. Since all basis elements of a matroid have the same size and since minimum representative sets are minimal, it turns out that minimum representative sets can be computed with a simple greedy algorithm. We emphasize that there is a clear difference between the closure operator $\mathrm{cl}(R)$ for rooted triple sets R and the respective matroid closure operator, although $\mathrm{cl}(R)$ is used to define the matroid $(R, \mathbb{F}_R)$, see [5] or Section 4 for further details. We exploit the techniques we used to prove the matroid structure and provide a novel algorithm to compute the closure $\mathrm{cl}(R)$ of a consistent set R of triples. Let L_R denote the set of leaves on which R is defined on. If R is large sized, that is, $|R| = \Theta(L_R^3)$, then our algorithm has the same asymptotic time complexity as the method proposed by Bryant and Steel which runs in $\mathcal{O}(|R||L_R|^4) \subseteq \mathcal{O}(|R|^5)$ time [6]. However, our algorithm has a time complexity of $\mathcal{O}(|R|^2|L_R|) \subseteq \mathcal{O}(|R|^3)$ and thus, significantly improves the computational effort for moderately sized input triple sets R. Further runtime improvements (up to a factor of $|R||L_R|$) can be achieved whenever minimum representative subset $R' \subseteq R$ are used as input triple set. It should be noted that Bryant and Steel established this algorithm in order to show that $\mathrm{cl}(R)$ can be computed in polynomial-time rather than to be efficient. Nevertheless, they supposed that "a far more efficient algorithm could be found". However, over the last two decades no such algorithm appeared in the literature. We wish to point out that the theory of matroids has touched phylogenetics also in many other contexts, see e.g. [2, 3, 17, 22, 27, 48, 49, 51, 59].

This contribution is organized as follows: In Section 2, we present the basic and relevant concepts used in this paper. In particular, we review important results for closure operations on rooted triple sets established by Bryant and Steel [5, 6]. A key property that will play a major role in this paper is provided by the graph representation of triple sets (Ahograph) and its connected components. In Section 3, we are concerned with structural properties of representative subsets $R' \subseteq R$ that are closely related to the structure of the Ahograph. The latter results will be used in Section 4 to show that minimal representative sets $R' \subseteq R$ (and its subsets) form a matroid $(R, \mathbb{F}_R)$. In Section 5, we present a novel method to compute the closure $\mathrm{cl}(R)$. Finally, we discuss in Section 6 further results. We give sufficient conditions that are quite useful to check whether an arbitrary triple is contained in all minimal representative sets and if R is already minimal. Moreover, we review and generalize some of the results established for triple sets R that "identify" or "define" a tree. In addition, we address the problem of finding minimal representative sets $Q' \subseteq Q$ of a collection Q of quartets (unrooted binary tree on four leaves). As it turns out, such sets do not provide a matroid structure. We conclude with a short discussion about the established results and open problems in Section 7.

2 Preliminaries

We consider undirected graphs $G = (V, E)$ with non-empty vertex set V and edge set E. A graph $G = (V, E)$ is *connected* if for any two vertices $x, y \in V$ there is a sequence of vertices $(x, v_1, \ldots, v_n, y)$, called *walk*, such that the edges $(x, v_1), (v_n, y)$ and (v_i, v_{i+1}), $1 \leq i \leq n-1$ are contained in E. A walk $(x, v_1, \ldots, v_n, y)$ in which all vertices are pairwise distinct is called a *path* and denoted by P_{xy}. A *cycle* is a walk $(x, v_1, \ldots, v_n, x)$ for which $n \geq 2$ and $(x, v_1, \ldots, v_n)$ is a path. A graph $H = (W, F)$ is a *subgraph* of $G = (V, E)$, in symbols $H \subseteq G$, if $W \subseteq V$ and $F \subseteq E$. The subgraph $H = (W, F)$ is an *induced* subgraph of $G = (V, E)$, if $x, y \in W$ and $(x, y) \in E$ implies $(x, y) \in F$. If $H = (W, F)$ is an induced subgraph of G we write $\langle W \rangle_G$ or simply $\langle W \rangle$ if there is no risk of confusion. A *connected component* of a graph $G = (V, E)$ is a subset $W \subseteq V$ such that $\langle W \rangle_G$ is connected and maximal w.r.t. inclusion.

A *tree* $T = (V, E)$ is a connected graph that does not contain cycles. The *leaf set* $L \subseteq V$ of T comprises all vertices that have degree 1. The vertices that are contained in $V^0 := V \setminus L$ are called *inner* vertices. The set of inner edges E^0 contains all edges $(x, y) \in E$ for which $x, y \in V^0$. A *rooted tree* $T = (V, E)$ is a tree with one distinguished inner vertex $\rho_T \in V$ called *root* of T. If every inner vertex of an unrooted tree has degree 3, the tree is called *binary*. A rooted tree is called *binary* if the degree of each inner vertex $v \neq \rho_T$ is 3 and the degree of the

root ρ_T is 2. In what follows, we consider rooted trees $T = (V, E)$ such that all inner vertices that are distinct from the root have degree at least three. For every vertex $v \in V$ we denote by $C(v)$ the leaf set of the subtree of T rooted at v and put $\mathcal{C}(T) = \bigcup_{v \in V}\{C(v)\}$, called the *hierarchy of* T. We say that a rooted tree T' *refines* T, in symbols $T \leq T'$, if $\mathcal{C}(T) \subseteq \mathcal{C}(T')$.

A *triple* $\mathtt{ab|c}$ is a binary rooted tree T on three leaves a, b and c such that the path from a to b does not intersect the path from c to the root ρ_T. A rooted tree T with leaf set L *displays* a triple $\mathtt{ab|c}$, if $a, b, c \in L$ and the path from a to b does not intersect the path from c to the root ρ_T. Note, that no distinction is made between $\mathtt{ab|c}$ and $\mathtt{ba|c}$. The set of all triples that are displayed by the rooted tree T is denoted by $\mathcal{R}(T)$. An arbitrary collection R of triples is called *triple set*. A triple set R is *consistent* if there is a rooted tree T such that $R \subseteq \mathcal{R}(T)$. In the latter case, we say that T *displays* R. The set $L_R := \cup_{\mathtt{ab|c} \in R}\{a, b, c\}$ is the union of the leaf set of each triple in R. A triple set R *identifies* a rooted tree T with leaf set L_R, if T displays R and any other tree that displays R refines T. A triple set R *defines* a rooted tree T with leaf set L_R, if T is the unique tree (up to isomorphism) that displays R.

There is a polynomial-time algorithm, which is customarily referred to as `BUILD` [57, 60], that was established by Aho, Sagiv, Szymanski, and Ullman [1]. `BUILD` either constructs a rooted tree T that displays R or recognizes that R is inconsistent [1]. The runtime of `BUILD` is $\mathcal{O}(|L_R||R|)$ [57]. Further practical implementations and improvements have been discussed in [14, 33, 37, 54]. `BUILD` is a top-down, recursive algorithm [1, 6] that uses an auxiliary graph that is also known as *Ahograph* [36], *clustering graph* [57] or *cluster graph* [18]. We will use the term "Ahograph". This graph is used to represent the structure of a collection of triples: For a given triple set R and an arbitrary subset $\mathcal{L} \subseteq L_R$, the Ahograph $[R, \mathcal{L}]$ has vertex set $\mathcal{L}$ and two vertices $a, b \in \mathcal{L}$ are linked by an edge, if there is a triple $\mathtt{ab|c} \in R$ with $c \in \mathcal{L}$. Based on connectedness properties of the graph $[R, \mathcal{L}]$ for particular subsets $\mathcal{L} \subseteq L_R$, the algorithm `BUILD` determines whether R is consistent or not. In particular, this algorithm makes use of the following well-known theorem.

Theorem 2.1 ([1, 6])**.** *A set of triples R is consistent if and only if for each subset $\mathcal{L} \subseteq L_R$ with $|\mathcal{L}| > 1$ the graph $[R, \mathcal{L}]$ is disconnected.*

Since we will use the Ahograph and its key features as a frequent tool in upcoming proofs, we now summarize some of its basic properties.

Lemma 2.2 ([6])**.** *If R' is a subset of the triple set R and $L' \subseteq L \subseteq L_R$, then $[R', L']$ is a subgraph of $[R, L]$.*

Lemma 2.3. *Let R be a triple set and $\mathcal{L} \subseteq L_R$. Assume that $A \subseteq \mathcal{L}_A \subseteq \mathcal{L}$ and $B \subseteq \mathcal{L}_B \subseteq \mathcal{L}$ such that the induced subgraphs $\langle A\rangle_{[R,\mathcal{L}_A]}$ and $\langle B\rangle_{[R,\mathcal{L}_B]}$ in $[R, \mathcal{L}_A]$ and $[R, \mathcal{L}_B]$, respectively, are connected. If $A \cap B \neq \emptyset$, then $\langle A \cup B\rangle_G$ is connected in G, where $G = [R, \mathcal{L}_A \cup \mathcal{L}_B]$ or $G = [R, \mathcal{L}]$.*

Proof. Let $A \subseteq \mathcal{L}_A \subseteq \mathcal{L}$, $B \subseteq \mathcal{L}_B \subseteq \mathcal{L}$ and assume that the induced subgraphs $\langle A\rangle$ of $[R, \mathcal{L}_A]$ and $\langle B\rangle$ of $[R, \mathcal{L}_B]$ are connected. By Lemma 2.2, $[R, \mathcal{L}_A]$ and $[R, \mathcal{L}_B]$ are subgraphs of $[R, \mathcal{L}_A \cup \mathcal{L}_B]$. Let $x \in A \cap B$. Thus, every vertex $y \in A \cup B$ is reachable from x by a walk in $[R, \mathcal{L}_A \cup \mathcal{L}_B]$. Hence, any two vertices $y, y' \in A \cup B$ are reachable by a walk (over x) in $[R, \mathcal{L}_A \cup \mathcal{L}_B]$ and therefore, $\langle A \cup B\rangle_{[R,\mathcal{L}_A \cup \mathcal{L}_B]}$ is a connected subgraph in $[R, \mathcal{L}_A \cup \mathcal{L}_B]$. Since $\mathcal{L}_A, \mathcal{L}_B \subseteq \mathcal{L}$ we can apply Lemma 2.2 and conclude that $[R, \mathcal{L}_A \cup \mathcal{L}_B]$ is a subgraph of $[R, \mathcal{L}]$ from what the statement follows. ■

The requirement that a set R of triples is consistent, and thus, that there is a tree displaying all triples, allows to infer new triples from the trees that display R and to define a *closure operation* for R. Let $\mathrm{span}(R)$ be the set of all rooted trees with leaf set L_R that display R. The closure of a consistent triple set R is defined as

$$\mathrm{cl}(R) = \bigcap_{T \in \mathrm{span}(R)} \mathcal{R}(T).$$

Hence, a triple r is contained in the closure $\mathrm{cl}(R)$ if all trees that display R also display r. This operation satisfies the usual three properties of a closure operator [6], namely:

- $R \subseteq \mathrm{cl}(R)$,
- $\mathrm{cl}(\mathrm{cl}(R)) = \mathrm{cl}(R)$, and
- if $R' \subseteq R$, then $\mathrm{cl}(R') \subseteq \mathrm{cl}(R)$.

There is a simple polynomial time algorithm to compute the closure that is based on the following lemmas.

Lemma 2.4 ([6, Prop. 9(1)]). *Let R be a consistent triple set. If $\mathrm{cl}(R)$ does not contain any triples with leaves $\{a,b,c\}$, then $\mathrm{cl}(R) \cup \{\mathtt{ab|c}\}$, $\mathrm{cl}(R) \cup \{\mathtt{ac|b}\}$ and $\mathrm{cl}(R) \cup \{\mathtt{bc|a}\}$ are all consistent.*

Lemma 2.5. *Let R be consistent. For all $\{a,b,c\} \in \binom{L_R}{3}$ exactly one of $R \cup \{\mathtt{ab|c}\}$, $R \cup \{\mathtt{ac|b}\}$ and $R \cup \{\mathtt{bc|a}\}$ is consistent (say $R \cup \{\mathtt{ab|c}\}$) if and only if $\mathtt{ab|c} \in \mathrm{cl}(R)$.*

Proof. Assume that only $R \cup \{\mathtt{ab|c}\}$ is consistent while $R \cup \{\mathtt{ac|b}\}$ and $R \cup \{\mathtt{bc|a}\}$ are not. Since the latter two sets are not consistent, there is no tree that displays R and, in addition, $\mathtt{ac|b}$, resp., $\mathtt{bc|a}$. Thus, $\mathtt{ac|b}, \mathtt{bc|a} \notin \mathrm{cl}(R)$. Assume for contradiction that additionally $\mathtt{ab|c} \notin \mathrm{cl}(R)$. Hence, $\mathrm{cl}(R)$ does not contain any triples with the leaves $\{a,b,c\}$. Lemma 2.4 implies that $\mathrm{cl}(R) \cup \{\mathtt{ac|b}\}$ is consistent. However, this implies that there is a tree T that display all triples of $\mathrm{cl}(R)$ and the triple $\mathtt{ac|b}$. Since $R \subseteq \mathrm{cl}(R)$ this tree T displays $R \cup \{\mathtt{ac|b}\}$; a contradiction.

Conversely, let $\mathtt{ab|c} \in \mathrm{cl}(R)$. Thus, every tree that displays R must also display $\mathtt{ab|c}$. Therefore, any tree that displays R does not display $\mathtt{ac|b}$ and $\mathtt{bc|a}$. Hence, there is no tree that displays R and in addition, $\mathtt{ac|b}$ (resp. $\mathtt{bc|a}$), which implies that $R \cup \{\mathtt{ac|b}\}$ and $R \cup \{\mathtt{bc|a}\}$ are not consistent. ■

Based on the latter result, the closure of a given consistent set R can be computed in $\mathcal{O}(|R||L_R|^4)$ time [6] as follows: For any three distinct leaves $a,b,c \in L_R$ test whether exactly one of the sets $R \cup \{\mathtt{ab|c}\}$, $R \cup \{\mathtt{ac|b}\}$, $R \cup \{\mathtt{bc|a}\}$ is consistent (e.g. with the $\mathcal{O}(|L_R||R|)$-time algorithm `BUILD`), and if so, add the respective triple to the closure $\mathrm{cl}(R)$ of R. A further characterization of the closure by means of the Ahograph is given by Bryant [5, Cor. 3.9].

Theorem 2.6. *For a consistent triple set R we have $\mathtt{ab|c} \in \mathrm{cl}(R)$ if and only if there is a subset $\mathcal{L} \subseteq L_R$ such that the Ahograph $[R,\mathcal{L}]$ has exactly two connected components, one containing a and b and the other containing c.*

We complete this section with a last result for later reference.

Lemma 2.7. *Let R be consistent and $R' \subseteq R$. Then $R' \subseteq \mathrm{cl}(R \setminus R')$ if and only if $\mathrm{cl}(R \setminus R') = \mathrm{cl}(R)$. In particular, if $\mathrm{cl}(R \setminus R') = \mathrm{cl}(R)$, then $\mathrm{cl}(R \setminus \{r\}) = \mathrm{cl}(R)$ for any triple $r \in R'$.*

Proof. If $R' \subseteq \mathrm{cl}(R \setminus R')$, then clearly $\mathrm{cl}(R \setminus R') = \mathrm{cl}(R \setminus R') \cup R'$. Therefore, $\mathrm{cl}(R \setminus R') = \mathrm{cl}(\mathrm{cl}(R \setminus R')) = \mathrm{cl}(\mathrm{cl}(R \setminus R') \cup R')$. Theorem 3.1(8) in [5] states that $\mathrm{cl}(\mathrm{cl}(A) \cup B) = \mathrm{cl}(A \cup B)$. Hence, $\mathrm{cl}(\mathrm{cl}(R \setminus R') \cup R') = \mathrm{cl}((R \setminus R') \cup R') = \mathrm{cl}(R)$. Conversely, if $\mathrm{cl}(R \setminus R') = \mathrm{cl}(R)$, then $R' \subseteq R \subseteq \mathrm{cl}(R)$ implies that $R' \subseteq \mathrm{cl}(R \setminus R')$.

Now let $R' \subseteq R$, $r \in R'$ and assume that $\mathrm{cl}(R \setminus R') = \mathrm{cl}(R)$. Since $R \setminus R' \subseteq R \setminus \{r\}$, we have $\mathrm{cl}(R) = \mathrm{cl}(R \setminus R') \subseteq \mathrm{cl}(R \setminus \{r\}) \subseteq \mathrm{cl}(R)$. Thus, $\mathrm{cl}(R \setminus \{r\}) = \mathrm{cl}(R)$. ■

3 Representative Triple Sets

The closure $\mathrm{cl}(R)$ provides all information of further triples that are implied by a consistent triple set R. Nevertheless, there might be subsets $R' \subseteq R$ that provide the same information, that is, $\mathrm{cl}(R') = \mathrm{cl}(R)$. See Figure 1 for an example.

Definition 3.1. *Let R be a consistent triple set. A set $R' \subseteq R$ is* representative *for R if $\mathrm{cl}(R) = \mathrm{cl}(R')$. The set $\mathfrak{sc}(R)$[1] comprises all representative triple sets of R. Moreover, we put*

$$\min(\mathfrak{sc}(R)) := \{R' \in \mathfrak{sc}(R) \colon R' \text{ is minimal w.r.t. inclusion } \}$$

and

$$\mathrm{MIN}(\mathfrak{sc}(R)) := \{R' \in \mathfrak{sc}(R) \colon |R'| \leq |R''| \text{ for any } R'' \in \mathfrak{sc}(R)\}.$$

It is easy to see that $\mathrm{MIN}(\mathfrak{sc}(R)) \subseteq \min(\mathfrak{sc}(R))$. As we shall see later, even $\mathrm{MIN}(\mathfrak{sc}(R)) = \min(\mathfrak{sc}(R))$ is satisfied. In order to investigate the sets $\mathrm{MIN}(\mathfrak{sc}(R))$ and $\min(\mathfrak{sc}(R))$ in more detail, we utilize the Ahograph and, in particular, Theorem 2.6. Note, Theorem 2.6 implies that $\mathtt{ab|c} \in \mathrm{cl}(R)$ if and only if there is a subset $\mathcal{L} \subseteq L_R$ such that $[R,\mathcal{L}]$ has exactly two connected components A and B, one containing a,b and the other c. These two connected components will play a major role in the proof for matroid properties. Since there might be several subsets $\mathcal{L}$ of L_R that satisfy the properties of Theorem 2.6 for a given triple $\mathtt{ab|c}$, we collect the respective connected components A and B in the set $\mathfrak{L}_{\mathtt{ab|c}}(R)$.

[1] $\mathfrak{sc}$ stands for "same closure"

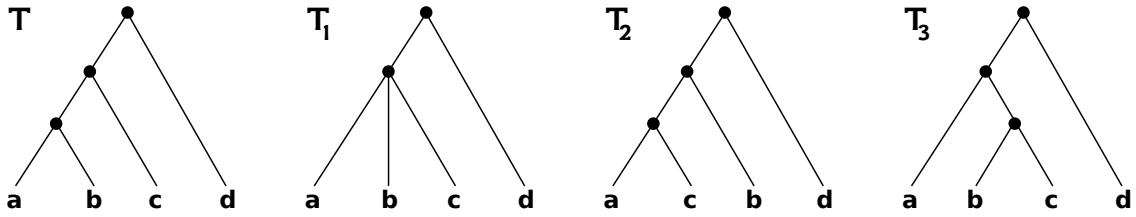

Figure 1: Given the set $R = \{\mathtt{ab}|\mathtt{c}, \mathtt{ac}|\mathtt{d}, \mathtt{bc}|\mathtt{d}\}$, there is only one tree T that displays R (shown left). Thus, $\mathrm{cl}(R) = \mathcal{R}(T) = \{\mathtt{ab}|\mathtt{c}, \mathtt{ac}|\mathtt{d}, \mathtt{bc}|\mathtt{d}, \mathtt{ab}|\mathtt{d}\}$. The subsets $R_1 = \{\mathtt{ab}|\mathtt{c}, \mathtt{ac}|\mathtt{d}\}$ and $R_2 = \{\mathtt{ab}|\mathtt{c}, \mathtt{bc}|\mathtt{d}\}$ are representative triple sets for R. In particular, both R_1 and R_2 are minimal and have minimum size. However, not all subsets of R with size two are representative. By way of example consider $R_3 = \{\mathtt{ac}|\mathtt{d}, \mathtt{bc}|\mathtt{d}\}$. Although T displays R_3, there are three further trees T_1, T_2 and T_3 that display R_3 as well. Thus, $\mathrm{cl}(R_3) = \mathcal{R}(T) \cap \mathcal{R}(T_1) \cap \mathcal{R}(T_2) \cap \mathcal{R}(T_3) = \{\mathtt{ac}|\mathtt{d}, \mathtt{bc}|\mathtt{d}, \mathtt{ab}|\mathtt{d}\} \neq \mathrm{cl}(R)$.

Definition 3.2. *Let R be a consistent triple set and $\mathtt{ab}|\mathtt{c}$ a triple with $a, b, c \in L_R$. The set*

$$\mathfrak{L}_{\mathtt{ab}|\mathtt{c}}(R)$$

comprises all sets $\{A, B\}$ for which $A, B \subseteq L_R$ and $[R, A \cup B]$ has exactly two connected components A and B, one containing a and b and the other containing c.

We emphasize that we do not assume that $\mathtt{ab}|\mathtt{c} \in R$ in Definition 3.2. The following lemma is an immediate consequence of Definition 3.2 and Theorem 2.6.

Lemma 3.3. *Let R be a consistent triple set. Then,*

$$\mathfrak{L}_{\mathtt{ab}|\mathtt{c}}(R) \neq \emptyset \text{ if and only if } \mathtt{ab}|\mathtt{c} \in \mathrm{cl}(R).$$

In what follows, we show that elements $\{A^\star, B^\star\} \in \mathfrak{L}_{\mathtt{ab}|\mathtt{c}}(R)$ with $|A^\star \cup B^\star| \geq |A \cup B|$ for all $\{A, B\} \in \mathfrak{L}_{\mathtt{ab}|\mathtt{c}}(R)$ are unique in $\mathfrak{L}_{\mathtt{ab}|\mathtt{c}}(R)$ and that A must be a subset of $A^\star$ (resp. $B^\star$) while B is a subset of $B^\star$ (resp. $A^\star$). In other words, the Ahograph $[R, A \cup B]$ must be a subgraph of $[R, A^\star \cup B^\star]$, where one of the two connected components of $[R, A \cup B]$ is entirely contained in $A^\star$ and the other in $B^\star$. To this end, we start with the following lemma.

Lemma 3.4. *Let R be a consistent triple set and $\mathtt{ab}|\mathtt{c}, \mathtt{a'b'}|\mathtt{c'} \in \mathrm{cl}(R)$. Assume that $\{A, B\} \in \mathfrak{L}_{\mathtt{ab}|\mathtt{c}}(R)$ and $\{A', B'\} \in \mathfrak{L}_{\mathtt{a'b'}|\mathtt{c'}}(R)$. If $A \cap A' \neq \emptyset$ and $B \cap B' \neq \emptyset$, then $\{A \cup A', B \cup B'\} \in \mathfrak{L}_{\mathtt{ab}|\mathtt{c}}(R) \cap \mathfrak{L}_{\mathtt{a'b'}|\mathtt{c'}}(R)$.*

Proof. Let R be consistent and $\mathtt{ab}|\mathtt{c}, \mathtt{a'b'}|\mathtt{c'} \in \mathrm{cl}(R)$. By Lemma 3.3, there are $\{A, B\} \in \mathfrak{L}_{\mathtt{ab}|\mathtt{c}}(R)$ and $\{A', B'\} \in \mathfrak{L}_{\mathtt{a'b'}|\mathtt{c'}}(R)$. Assume that $A \cap A' \neq \emptyset$ and $B \cap B' \neq \emptyset$ and let $\mathcal{L}' = A \cup A' \cup B \cup B'$.

By Lemma 2.2, both $[R, A \cup B]$ and $[R, A' \cup B']$ are subgraphs of $[R, \mathcal{L}']$. Moreover, by definition, $[R, A \cup B]$ and $[R, A' \cup B']$ have two connected components A, B, resp., A', B'. Furthermore, since $A \cap A' \neq \emptyset$ and $B \cap B' \neq \emptyset$ we can apply Lemma 2.3 and conclude that the induced subgraphs $\langle A \cup A' \rangle_{[R,\mathcal{L}']}$ and $\langle B \cup B' \rangle_{[R,\mathcal{L}']}$ form connected subgraphs in $[R, \mathcal{L}']$. Moreover, since R is consistent, Theorem 2.1 implies that $[R, \mathcal{L}']$ cannot be connected. Hence, $A \cup A'$ and $B \cup B'$ must be the connected components in $[R, \mathcal{L}']$, still one containing a and b (resp. a', b') and the other c (resp. c') and therefore, $\{A \cup A', B \cup B'\} \in \mathfrak{L}_{\mathtt{ab}|\mathtt{c}}(R)$ (resp. $\{A \cup A', B \cup B'\} \in \mathfrak{L}_{\mathtt{a'b'}|\mathtt{c'}}(R)$). ■

Lemma 3.5. *Let R be a consistent triple set with $\mathtt{ab}|\mathtt{c} \in \mathrm{cl}(R)$. Let $\{A^\star, B^\star\} \in \mathfrak{L}_{\mathtt{ab}|\mathtt{c}}(R)$ such that $|A^\star \cup B^\star| \geq |A \cup B|$ for all $\{A, B\} \in \mathfrak{L}_{\mathtt{ab}|\mathtt{c}}(R)$. Then, either $A \subseteq A^\star$ and $B \subseteq B^\star$ or $B \subseteq A^\star$ and $A \subseteq B^\star$.*

Moreover, the element $\{A^\star, B^\star\}$ in $\mathfrak{L}_{\mathtt{ab}|\mathtt{c}}(R)$ with $|A^\star \cup B^\star| \geq |A \cup B|$ for all $\{A, B\} \in \mathfrak{L}_{\mathtt{ab}|\mathtt{c}}(R)$ is unique.

Proof. Let R be a consistent triple set and $\mathtt{ab}|\mathtt{c} \in \mathrm{cl}(R)$. By Lemma 3.3, the set $\mathfrak{L}_{\mathtt{ab}|\mathtt{c}}(R)$ is not empty, and thus, there is an element $\{A^\star, B^\star\} \in \mathfrak{L}_{\mathtt{ab}|\mathtt{c}}(R)$ such that $|A^\star \cup B^\star| \geq |A \cup B|$ for all $\{A, B\} \in \mathfrak{L}_{\mathtt{ab}|\mathtt{c}}(R)$. W.l.o.g. assume that $a, b \in A$ and $c \in B$ for some $\{A, B\} \in \mathfrak{L}_{\mathtt{ab}|\mathtt{c}}(R)$. There are two cases, either $a, b \in A^\star$ and $c \in B^\star$ or, $c \in A^\star$ and $a, b \in B^\star$. Let us first assume that $a, b \in A^\star$ and $c \in B^\star$. Thus, $A \cap A^\star \neq \emptyset$ and $B \cap B^\star \neq \emptyset$. Lemma 3.4 implies that $\{A \cup A^\star, B \cup B^\star\} \in \mathfrak{L}_{\mathtt{ab}|\mathtt{c}}(R)$ and, by choice of $A^\star$ and $B^\star$, $|A^\star \cup B^\star| \geq |\mathcal{L}'|$ where $\mathcal{L}' = A^\star \cup A \cup B^\star \cup B$.

We continue to show that $A \subseteq A^\star$ and $B \subseteq B^\star$. Since $\{A \cup A^\star, B \cup B^\star\} \in \mathfrak{L}_{\mathtt{ab|c}}(R)$ we can conclude that $(A \cup A^\star) \cap (B \cup B^\star) = \emptyset$. Now, assume for contradiction that $A \not\subseteq A^\star$. Thus, there is an $x \in A \setminus (A^\star \cup B^\star)$ and therefore, $A^\star \cup B^\star \subsetneq (A \cup A^\star) \cup (B \cup B^\star)$. Hence, $|(A \cup A^\star) \cup (B \cup B^\star)| > |A^\star \cup B^\star|$; a contradiction. Analogously, $B \subseteq B^\star$.

The latter arguments immediately imply that for any $\{A_1^\star, B_1^\star\}, \{A_2^\star, B_2^\star\} \in \mathfrak{L}_{\mathtt{ab|c}}(R)$ with $|A_1^\star \cup B_1^\star| = |A_2^\star \cup B_2^\star| \geq |A \cup B|$ for all $\{A, B\} \in \mathfrak{L}_{\mathtt{ab|c}}(R)$ it must hold $\{A_1^\star, B_1^\star\} = \{A_2^\star, B_2^\star\}$. ■

For our results it will be convenient to explicitly name the unique element $\{A^\star, B^\star\}$ that has maximum cardinality $|A^\star \cup B^\star|$ in $\mathfrak{L}_{\mathtt{ab|c}}(R)$ as defined next.

Definition 3.6. *Let R be a consistent triple set with $\mathtt{ab|c} \in \mathrm{cl}(R)$. Then,*

$$\ell^\star_{\mathtt{ab|c}}(R)$$

denotes the unique element $\{A^\star, B^\star\} \in \mathfrak{L}_{\mathtt{ab|c}}(R)$ for which $|A^\star \cup B^\star| \geq |A \cup B|$ for all $\{A, B\} \in \mathfrak{L}_{\mathtt{ab|c}}(R)$.

Moreover, for a subset $R' \subseteq \mathrm{cl}(R)$ we set

$$\mathfrak{L}^\star_{R'}(R) \coloneqq \bigcup_{\mathtt{ab|c} \in R'} \{\ell^\star_{\mathtt{ab|c}}(R)\}.$$

It is easy to verify that $|\mathfrak{L}^\star_R(R)| \leq |R|$. In what follows, we will show that for any consistent set R the sets $\mathfrak{L}^\star_R(R), \mathfrak{L}^\star_{\mathrm{cl}(R)}(R)$ and $\mathfrak{L}^\star_{\mathrm{cl}(R)}(\mathrm{cl}(R))$ are identical. This in turn is used to show that $\mathfrak{L}^\star_R(R) = \mathfrak{L}^\star_{R'}(R')$ whenever $\mathrm{cl}(R') = \mathrm{cl}(R)$ for some $R' \subseteq R$. In particular, the elements in $\mathfrak{L}^\star_R(R)$ and $\mathfrak{L}^\star_{R'}(R')$ are identical w.r.t. to a given triple $\mathtt{ab|c} \in \mathrm{cl}(R)$, that is, $\ell^\star_{\mathtt{ab|c}}(R) = \ell^\star_{\mathtt{ab|c}}(R')$ for any $\mathtt{ab|c} \in \mathrm{cl}(R)$. Hence, if $\ell^\star_{\mathtt{ab|c}}(R) = \{A, B\}$ and $\ell^\star_{\mathtt{ab|c}}(R') = \{A', B'\}$, then $[R, A \cup B]$ and $[R', A' \cup B']$ have the same two connected components. Note, the latter does not imply that the Ahographs $[R, A \cup B]$ and $[R', A' \cup B']$ are isomorphic. We refer to Figure 2 for an illustrative example. To establish these results we provide first the following lemma.

Lemma 3.7. *Let R be a consistent triple set. Assume that there are distinct $\{A', B'\}, \{A, B\} \in \mathfrak{L}^\star_R(R)$. If $A' \cap (A \cup B) \neq \emptyset$ and $B' \cap (A \cup B) \neq \emptyset$, then $A' \cup B' \subseteq A$ or $A' \cup B' \subseteq B$.*

Proof. If $\{A, B\}, \{A', B'\}$ are distinct elements of $\mathfrak{L}^\star_R(R)$, then there are distinct triples $\mathtt{ab|c}$ and $\mathtt{a'b'|c'}$ in R such that $\ell^\star_{\mathtt{ab|c}}(R) = \{A, B\}$ and $\ell^\star_{\mathtt{a'b'|c'}}(R) = \{A', B'\}$. Assume that $A' \cap (A \cup B) \neq \emptyset$ and $B' \cap (A \cup B) \neq \emptyset$.

First consider the case $A' \cap A \neq \emptyset$ *and* $A' \cap B \neq \emptyset$. Lemma 2.3 implies that the induced subgraph $\langle A' \cup A \cup B \rangle$ of $[R, A \cup B \cup A' \cup B']$ is connected. Since $B' \cap (A \cup B) \neq \emptyset$ and B' is a connected component in $[R, A' \cup B']$, we can apply Lemma 2.2 and 2.3 and conclude that $[R, A \cup B \cup A' \cup B']$ is a connected graph; a contradiction to Theorem 2.6. Hence, the case $A' \cap A \neq \emptyset$ *and* $A' \cap B \neq \emptyset$ cannot occur. Similarly, $B' \cap A \neq \emptyset$ *and* $B' \cap B \neq \emptyset$ is not possible. Thus, we have either $A' \cap A \neq \emptyset$ or $A' \cap B \neq \emptyset$ as well as, either $B' \cap A \neq \emptyset$ or $B' \cap B \neq \emptyset$.

First assume that $A' \cap A \neq \emptyset$ and $B' \cap B \neq \emptyset$. By Lemma 3.4, $\{A \cup A', B \cup B'\} \in \mathfrak{L}_{\mathtt{ab|c}}(R) \cap \mathfrak{L}_{\mathtt{a'b'|c'}}(R)$. Thus, by Lemma 3.5 we have, on the one hand, $A \cup A' \subseteq A$ and $B \cup B' \subseteq B$ and, one the other hand, $A \cup A' \subseteq A'$ and $B \cup B' \subseteq B'$. Hence, $A = A'$ and $B = B'$; a contradiction since we assumed that $\{A, B\}$ and $\{A', B'\}$ are distinct. Thus, the case $A' \cap A \neq \emptyset$ and $B' \cap B \neq \emptyset$ cannot occur. Similarly, the case $B' \cap A \neq \emptyset$ and $A' \cap B \neq \emptyset$ is impossible.

Therefore, we are left with two exclusive cases: (1) $A' \cap A \neq \emptyset$ and $B' \cap A \neq \emptyset$ or (2) $A' \cap B \neq \emptyset$ and $B' \cap B \neq \emptyset$. Let us assume case (1) $A' \cap A \neq \emptyset$ and $B' \cap A \neq \emptyset$. Repeated application of Lemma 2.3 shows that the induced subgraph $\langle A \cup A' \cup B' \rangle$ of $[R, A \cup B \cup A' \cup B']$ is connected. Since R is consistent, Theorem 2.6 implies that the graph $[R, A \cup B \cup A' \cup B']$ must be disconnected. Hence, $[R, A \cup B \cup A' \cup B']$ has as connected components $A \cup A' \cup B'$ and B. Therefore, $\{A \cup A' \cup B', B\} \in \mathfrak{L}_{\mathtt{ab|c}}(R)$. Now it must hold that $A' \cup B' \subseteq A$ as otherwise $|A \cup A' \cup B' \cup B| > |A \cup B| = \ell^\star_{\mathtt{ab|c}}(R)$ would yield a contradiction. In case (2) it is shown analogously that $A' \cup B' \subseteq B$. ■

Lemma 3.7 immediately implies the following

Corollary 3.8. *Let R be a consistent triple set. Assume that there are distinct $\{A, B\}, \{A', B'\} \in \mathfrak{L}^\star_R(R)$. If $A \cap A' \neq \emptyset$, then $B \cap B' = \emptyset$.*

Proof. Assume that $A \cap A' \neq \emptyset$ and $B \cap B' \neq \emptyset$. Hence, $A' \cap (A \cup B) \neq \emptyset$ and $B' \cap (A \cup B) \neq \emptyset$. Lemma 3.7 implies that $A' \cup B' \subseteq A$ or $A' \cup B' \subseteq B$. W.l.o.g. assume that $A' \cup B' \subseteq A$. Analogously, Lemma 3.7 implies that $A \cup B \subseteq A'$ or $A \cup B \subseteq B'$. Now, $A' \cup B' \subseteq A$ and $A \cup B \subseteq A'$ would imply that $B' \subseteq A'$; a contradiction. Furthermore, if $A \cup B \subseteq B'$, then $A' \cup B' \subseteq A$ would imply that $B \subseteq A$; again a contradiction. ■

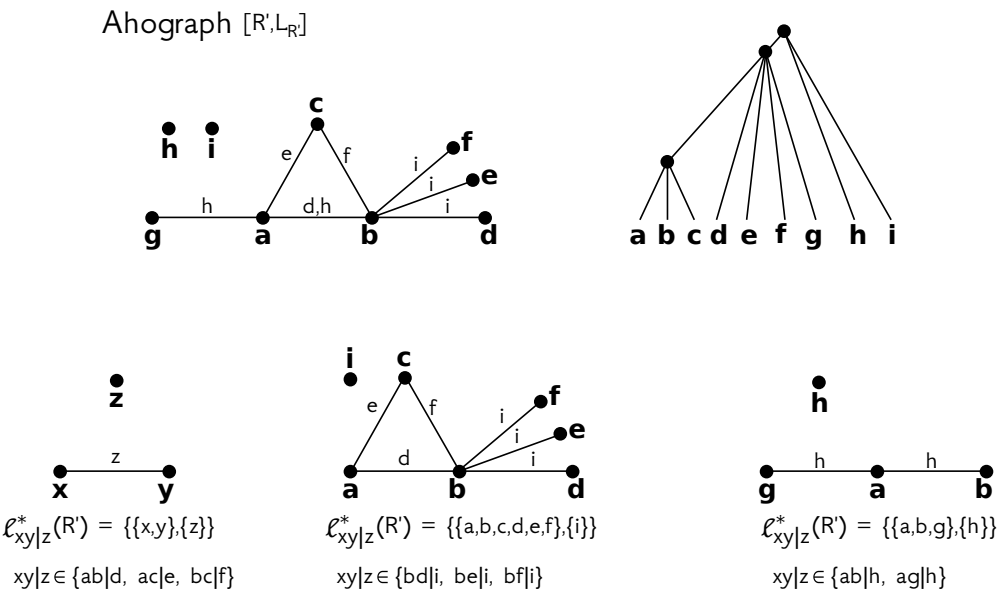

Figure 2: Consider the triple set $R = \{\mathtt{ab|d}, \mathtt{ab|h}, \mathtt{ac|e}, \mathtt{ag|h}, \mathtt{bc|f}, \mathtt{bc|i}, \mathtt{bd|i}, \mathtt{be|i}, \mathtt{bf|i}, \mathtt{bg|h}\}$ and let $R' = R \setminus \{\mathtt{bc|i}, \mathtt{bg|h}\}$. In this example, $L_{R'} = L_R$. Clearly, any tree that display R' and thus, $\{\mathtt{bc|f}, \mathtt{bf|i}\}$, resp., $\{\mathtt{ab|h}, \mathtt{ag|h}\}$, must also display $\mathtt{bc|i}$, resp., $\mathtt{bg|h}$ [13]. Thus, $\{\mathtt{bc|i}, \mathtt{bg|h}\} \subseteq \mathrm{cl}(R') = \mathrm{cl}(R \setminus \{\mathtt{bc|i}, \mathtt{bg|h}\})$. Lemma 2.7 implies that $\mathrm{cl}(R') = \mathrm{cl}(R)$ and therefore, $R' \in \mathfrak{sc}(R)$. The Ahograph $[R', L_{R'}]$ and a tree T that displays R' is shown on the top of this figure. In $[R', L_{R'}]$ each edge (x, y) is labeled with z that corresponds to the respective triple $\mathtt{xy|z}$ that supports the edge (x, y).
Since $R \subseteq \mathrm{cl}(R) = \mathrm{cl}(R')$, the tree T also displays R. The respective maximal elements $\ell^\star_{\mathtt{xy|z}}(R') = \{A, B\} \in \mathfrak{L}^\star_{R'}(R')$ with corresponding Ahographs $[R', A \cup B]$ are depicted below. It is easy to verify that each triple $\mathtt{xy|z} \in R'$ is a bridge in the respective Ahograph $[R', A \cup B]$ and hence, R' is minimal (cf. Lemma 4.6). By Theorem 4.8, R' has also minimum cardinality. Note, $\mathrm{cl}(R') = \mathrm{cl}(R)$ and Theorem 3.9 imply that $\mathfrak{L}^\star_{R'}(R') = \mathfrak{L}^\star_{\mathrm{cl}(R)}(\mathrm{cl}(R)) = \mathfrak{L}^\star_R(R)$. In this example, $\mathfrak{L}^\star_R(R) = \big\{\{\{a, b\}, \{d\}\}, \{\{a, c\}, \{e\}\}, \{\{b, c\}, \{f\}\}, \{\{a, b, c, d, e, f\}, \{i\}\}, \{\{a, b, g\}, \{h\}\}\big\}$. In order to determine $\mathrm{cl}(R)$ it suffices to add for each $\{A, B\} \in \mathfrak{L}^\star_R(R)$ all triples $\mathtt{xy|z}$ with $x, y \in A, z \in B$ or $z \in A, x, y \in B$ to $\mathrm{cl}(R)$ (cf. Thm. 5.1). Finally, application of Theorem 6.4 shows that R' does not identify T, since $9 = B(T) > |R'| = 8$ and thus, $\mathrm{cl}(R') \neq \mathcal{R}(T)$. Moreover, since $\mathrm{cl}(R) = \mathrm{cl}(R')$ neither R identifies T.

Theorem 3.9. *For any consistent triple set R it holds that*

$$\mathfrak{L}^\star_R(R) = \mathfrak{L}^\star_{\mathrm{cl}(R)}(R) = \mathfrak{L}^\star_{\mathrm{cl}(R)}(\mathrm{cl}(R)).$$

Proof. We start with showing $\mathfrak{L}^\star_R(R) = \mathfrak{L}^\star_{\mathrm{cl}(R)}(R)$. Since $R \subseteq \mathrm{cl}(R)$, we also have $\mathfrak{L}^\star_R(R) \subseteq \mathfrak{L}^\star_{\mathrm{cl}(R)}(R)$.

To see that $\mathfrak{L}^\star_{\mathrm{cl}(R)}(R) \subseteq \mathfrak{L}^\star_R(R)$, let $\{A, B\} \in \mathfrak{L}^\star_{\mathrm{cl}(R)}(R)$. Hence, there is a triple $\mathtt{ab|c} \in \mathrm{cl}(R)$ with $\ell^\star_{\mathtt{ab|c}}(R) = \{A, B\} \in \mathfrak{L}_{\mathtt{ab|c}}(R)$. By definition, $[R, A \cup B]$ has two connected components and at least one contains 2 or more vertices. First, assume for contradiction that there is no triple $\mathtt{a'b'|c'} \in R$ with $\{A, B\} \in \mathfrak{L}_{\mathtt{a'b'|c'}}(R)$. Since $|A| > 1$ or $|B| > 1$ and A, B are connected components in $[R, A \cup B]$, all edges (x, y) within $\langle A\rangle_{[R,A\cup B]}$ and $\langle B\rangle_{[R,A\cup B]}$ are, therefore, provided by triples $\mathtt{xy|z} \in R$ such that either $x, y, z \in A$ or $x, y, z \in B$. But then $[R, A]$ or $[R, B]$ is connected; contradicting Theorem 2.1. Thus, there must be a triple $\mathtt{a'b'|c'} \in R$ with $\{A, B\} \in \mathfrak{L}_{\mathtt{a'b'|c'}}(R)$. We continue with showing that $\ell^\star_{\mathtt{a'b'|c'}}(R) = \{A, B\}$. If this was not the case, then there is an $\ell^\star_{\mathtt{a'b'|c'}}(R) = \{A^\star, B^\star\}$ with $|A^\star \cup B^\star| > |A \cup B|$. Since $\ell^\star_{\mathtt{ab|c}}(R) = \{A, B\} \in \mathfrak{L}_{\mathtt{ab|c}}(R)$, one of A and B is containing a, b and the other c. Moreover, Lemma 3.5 implies that $A \subseteq A^\star$ and $B \subseteq B^\star$ or $B \subseteq A^\star$ and $A \subseteq B^\star$. Taken the latter two arguments together, one of $A^\star$ and $B^\star$ is containing a, b and the other c. Therefore, $\{A^\star, B^\star\} \in \mathfrak{L}_{\mathtt{ab|c}}(R)$. However, since $|A^\star \cup B^\star| > |A \cup B|$, we have $\ell^\star_{\mathtt{ab|c}}(R) \neq \{A, B\}$; a contradiction to the assumption $\ell^\star_{\mathtt{ab|c}}(R) = \{A, B\}$. Therefore, $\ell^\star_{\mathtt{a'b'|c'}}(R) = \{A, B\}$ and thus, $\{A, B\} \in \mathfrak{L}^\star_R(R)$. Hence, $\mathfrak{L}^\star_R(R) = \mathfrak{L}^\star_{\mathrm{cl}(R)}(R)$.

Now we show that $\mathfrak{L}^\star_{\mathrm{cl}(R)}(R) = \mathfrak{L}^\star_{\mathrm{cl}(R)}(\mathrm{cl}(R))$. Let $\{A, B\} = \ell^\star_{\mathtt{ab|c}}(R) \in \mathfrak{L}^\star_{\mathrm{cl}(R)}(R)$. Since $R \subseteq \mathrm{cl}(R)$ and by Lemma 2.2, $[R, A \cup B]$ is a subgraph of $[\mathrm{cl}(R), A \cup B]$. Since $\{A, B\} = \ell^\star_{\mathtt{ab|c}}(R)$, the graph $[R, A \cup B]$ has exactly two connected components A and B. Thus, $[\mathrm{cl}(R), A \cup B]$ has at most two connected components. Still, the induced subgraphs $\langle A\rangle$ and $\langle B\rangle$ of $[\mathrm{cl}(R), A \cup B]$ are connected. However, since $\mathrm{cl}(R)$ is consistent we can apply Theorem 2.1 and conclude that

$[\mathrm{cl}(R), A \cup B]$ must be disconnected, and thus, has as connected components A and B. Therefore, $\{A, B\} \in \mathfrak{L}_{\mathtt{ab|c}}(\mathrm{cl}(R))$. Assume now for contradiction that $\ell^\star_{\mathtt{ab|c}}(\mathrm{cl}(R)) = \{A^\star, B^\star\} \neq \{A, B\}$. By Lemma 3.5, $|A^\star \cup B^\star| > |A \cup B|$ and either $A \subseteq A^\star$ and $B \subseteq B^\star$ or $B \subseteq A^\star$ and $A \subseteq B^\star$. W.l.o.g. assume that $A \subseteq A^\star$ and $B \subseteq B^\star$ and $a, b \in A$, $c \in B$. Hence, there is a vertex $d \in (A^\star \cup B^\star) \setminus (A \cup B)$. If $d \in A^\star$, then Theorem 2.6 and $a \in A^\star$, $c \in B^\star$ imply that $\mathtt{ad|c} \in \mathrm{cl}(\mathrm{cl}(R)) = \mathrm{cl}(R)$. Again, Theorem 2.6 implies that there is a subset $\mathcal{L} \subseteq L_R$ such that $[R, \mathcal{L}]$ has exactly two connected components A', B' with $a, d \in A'$ and $c \in B'$. Thus, $\{A', B'\} \in \mathfrak{L}_{\mathtt{ad|c}}(R)$. Recap that $\{A, B\} \in \mathfrak{L}_{\mathtt{ab|c}}(R)$. Since $A \cap A' \neq \emptyset$ and $B \cap B' \neq \emptyset$ we can apply Lemma 3.4 and conclude that $\{A \cup A', B \cup B'\} \in \mathfrak{L}_{\mathtt{ab|c}}(R)$. However, since $d \in A' \setminus A$ it holds that $|A \cup A' \cup B \cup B'| > |A \cup B|$; a contradiction to $\{A, B\} = \ell^\star_{\mathtt{ab|c}}(R)$. By similar arguments one derives a contradiction if $d \in B^\star$. Thus, $\ell^\star_{\mathtt{ab|c}}(\mathrm{cl}(R)) = \{A^\star, B^\star\} = \{A, B\}$ and therefore, $\{A, B\} \in \mathfrak{L}^\star_{\mathrm{cl}(R)}(\mathrm{cl}(R))$. Hence, $\mathfrak{L}^\star_{\mathrm{cl}(R)}(R) \subseteq \mathfrak{L}^\star_{\mathrm{cl}(R)}(\mathrm{cl}(R))$.

Finally, let $\{A, B\} = \ell^\star_{\mathtt{ab|c}}(\mathrm{cl}(R)) \in \mathfrak{L}^\star_{\mathrm{cl}(R)}(\mathrm{cl}(R))$. By Lemma 3.3 and since $\mathtt{ab|c} \in \mathrm{cl}(R)$, we have $\mathfrak{L}_{\mathtt{ab|c}}(R) \neq \emptyset$. Thus, there is a maximal element $\ell^\star_{\mathtt{ab|c}}(R) = \{A^\star, B^\star\} \in \mathfrak{L}^\star_{\mathrm{cl}(R)}(R) \subseteq \mathfrak{L}^\star_{\mathrm{cl}(R)}(\mathrm{cl}(R))$. Assume that $\{A, B\}$ and $\{A^\star, B^\star\}$ are distinct. Since both $\{A, B\}$ and $\{A^\star, B^\star\}$ are contained in $\mathfrak{L}^\star_{\mathrm{cl}(R)}(\mathrm{cl}(R))$ and $A \cap (A^\star \cup B^\star) \neq \emptyset$, $B \cap (A^\star \cup B^\star) \neq \emptyset$ we can apply Lemma 3.7 and conclude that $A \cup B \subseteq A^\star$ or $A \cup B \subseteq B^\star$. If $A \cup B \subseteq A^\star$, then $a, b, c \in A^\star$; a contradiction, since one of $A^\star$ and $B^\star$ contains a, b and the other c. Analogously, $A \cup B \subseteq B^\star$ cannot occur. Hence, $\{A, B\}$ and $\{A^\star, B^\star\}$ must be equal and therefore, $\ell^\star_{\mathtt{ab|c}}(R) = \ell^\star_{\mathtt{ab|c}}(\mathrm{cl}(R)) = \{A, B\} \in \mathfrak{L}^\star_{\mathrm{cl}(R)}(R)$. Thus, $\mathfrak{L}^\star_{\mathrm{cl}(R)}(R) = \mathfrak{L}^\star_{\mathrm{cl}(R)}(\mathrm{cl}(R))$. ■

Theorem 3.10. *Let R be a consistent triple set. If $R' \in \mathfrak{sc}(R)$, then $\mathfrak{L}^\star_R(R) = \mathfrak{L}^\star_{R'}(R')$. In particular, for every $\mathtt{ab|c} \in cl(R)$ and $R' \in \mathfrak{sc}(R)$, it holds that $\ell^\star_{\mathtt{ab|c}}(R') = \ell^\star_{\mathtt{ab|c}}(R)$.*

Proof. Let $R' \in \mathfrak{sc}(R)$ and thus, $\mathrm{cl}(R') = \mathrm{cl}(R)$. Therefore,

$$\mathfrak{L}^\star_R(R) \overset{\text{Thm. 3.9}}{=} \mathfrak{L}^\star_{\mathrm{cl}(R)}(\mathrm{cl}(R)) = \mathfrak{L}^\star_{\mathrm{cl}(R')}(\mathrm{cl}(R')) \overset{\text{Thm. 3.9}}{=} \mathfrak{L}^\star_{R'}(R').$$

Now let $\ell^\star_{\mathtt{ab|c}}(R) = \{A^\star, B^\star\} \in \mathfrak{L}^\star_{\mathrm{cl}(R)}(R)$. Note, Theorem 3.9 implies that $\mathfrak{L}^\star_{\mathrm{cl}(R)}(R) = \mathfrak{L}^\star_R(R)$ and therefore, $\{A^\star, B^\star\} \in \mathfrak{L}^\star_R(R) = \mathfrak{L}^\star_{R'}(R')$. Since $\mathtt{ab|c} \in \mathrm{cl}(R) = \mathrm{cl}(R')$ we can apply Lemma 3.3 and conclude that $\mathfrak{L}_{\mathtt{ab|c}}(R') \neq \emptyset$. Let $\ell^\star_{\mathtt{ab|c}}(R') = \{A, B\} \in \mathfrak{L}^\star_{\mathrm{cl}(R')}(R') \overset{\text{Thm. 3.9}}{=} \mathfrak{L}^\star_{R'}(R')$. Thus, both $\{A^\star, B^\star\}$ and $\{A, B\}$ are contained in $\mathfrak{L}^\star_{R'}(R')$ and $A \cap (A^\star \cup B^\star) \neq \emptyset$, $B \cap (A^\star \cup B^\star) \neq \emptyset$. Now we can argue analogously as in the last part of the proof of Theorem 3.3 to conclude that $\{A, B\} = \{A^\star, B^\star\}$ which implies that $\ell^\star_{\mathtt{ab|c}}(R') = \ell^\star_{\mathtt{ab|c}}(R)$. ■

4 The Matroid Structure of Minimal and Minimum Representative Triple Sets

By definition, $R' \in \min(\mathfrak{sc}(R))$ if and only if there is no subset $R'' \subsetneq R'$ with $\mathrm{cl}(R'') = \mathrm{cl}(R)$. Furthermore, since any minimum representative triple set is, in particular, minimal, we have $\mathrm{MIN}(\mathfrak{sc}(R)) \subseteq \min(\mathfrak{sc}(R))$. The computation of a minimal representative set R' of R can be done in combination with the $\mathcal{O}(|R||L_R|^4)$ method to compute the closure [6] in polynomial time as follows: Set $R' = R$ and as long as there is a triple $r \in \mathrm{cl}(R' \setminus r)$ remove r from R'. By Lemma 2.7, removal of r from R' still preserves $\mathrm{cl}(R) = \mathrm{cl}(R')$. However, the computational complexity of finding a minimum representative set R' of R is still an open problem. We show that one can determine minimum representative sets in polynomial time. To this end, we give the following

Definition 4.1. *A* matroid *is an ordered pair $(E, \mathbb{F}_E)$ consisting of a finite set E and a collection $\mathbb{F}_E$ of subsets of E having the following three properties:*

(I1) $\emptyset \in \mathbb{F}_E$*;*

(I2) If $I \in \mathbb{F}_E$ and $I' \subseteq I$, then $I' \in \mathbb{F}_E$;

(I3) If $I_1, I_2 \in \mathbb{F}_E$ and $|I_1| < |I_2|$, then there is an element $x \in I_2 \setminus I_1$ such that $I_1 \cup \{x\} \in \mathbb{F}_E$.

The elements in $\mathbb{F}_E$ are called *independent* in $(E, \mathbb{F}_E)$. Maximal independent elements of a matroid are called a basis of $(E, \mathbb{F}_E)$. Every matroid $(E, \mathbb{F}_E)$ is determined by its collection of its bases. We refer the reader to [40, 50] for more detailed background on matroid theory.

In what follows, we show that $\min(\mathfrak{sc}(R))$ forms the collection of bases of a matroid. In this case, $\mathrm{MIN}(\mathfrak{sc}(R)) = \min(\mathfrak{sc}(R))$ since all basis elements of a matroid have the same cardinality [40, 50]. A useful characterization is given by the next result.

Lemma 4.2 ([50, Cor. 1.2.5]). *Let $\mathcal{B}$ be a collection of subsets of E. Then $\mathcal{B}$ is the collection of bases of a matroid $(E, \mathbb{F}_E)$ if and only if it has the following properties:*

(B1) $\mathcal{B} \neq \emptyset$;

(B2) *If $B_1, B_2 \in \mathcal{B}$ and $x \in B_1 \setminus B_2$, then there is an element $y \in B_2 \setminus B_1$ such $(B_1 \setminus \{x\}) \cup \{y\} \in \mathcal{B}$.*

Definition 4.3. *In what follows, $(R, \mathbb{F}_R)$ denotes the ordered pair where*

1. *R is a consistent triple set and*
2. *$\mathbb{F}_R = \{R'' \subseteq R' \colon R' \in \min(\mathfrak{sc}(R))\}$ is the collection of all subsets of the minimal representative sets of R.*

It is easy to see that $(R, \mathbb{F}_R)$ is an independent system, that is, it satisfies Conditions (I1) and (I2). Moreover, the collection of bases of $(R, \mathbb{F}_R)$ is the set $\min(\mathfrak{sc}(R))$. We will utilize Lemma 4.2 to show $\mathcal{B} = \min(\mathfrak{sc}(R))$ satisfies (B1) and (B2). To this end, we give the notion of "bridges" in the Ahograph, that is, triples $\mathtt{ab|c}$ for which the Ahograph $[R \setminus \{\mathtt{ab|c}\}, \mathcal{L}]$ has more connected components than $[R, \mathcal{L}]$. As it turns out, elements $R' \in \min(\mathfrak{sc}(R))$ are characterized by the bridge-property of triples $\mathtt{ab|c} \in R'$. We first give the following result.

Lemma 4.4. *Let R be a consistent triple set and $R' \in \mathfrak{sc}(R)$. Then $R' \in \min(\mathfrak{sc}(R))$ if and only if $\mathrm{cl}(R') \neq \mathrm{cl}(R' \setminus \{r\})$ for all $r \in R'$.*

Proof. Let R be a consistent triple set and $R' \in \mathfrak{sc}(R)$ and thus, $\mathrm{cl}(R') = \mathrm{cl}(R)$. Clearly, if $\mathrm{cl}(R) = cl(R') = \mathrm{cl}(R' \setminus \{r\})$ for any $r \in R'$, then $R' \notin \min(\mathfrak{sc}(R))$. Conversely, if $R' \notin \min(\mathfrak{sc}(R))$, then there is a subset $R'' \subsetneq R'$ with $\mathrm{cl}(R'') = \mathrm{cl}(R)$. Since $R' \in \mathfrak{sc}(R)$, it also holds that $\mathrm{cl}(R') = \mathrm{cl}(R) = \mathrm{cl}(R'')$. Let $r \in R' \setminus R''$. Since $R'' \subseteq R' \setminus \{r\} \subsetneq R'$, we have $\mathrm{cl}(R'') \subseteq \mathrm{cl}(R' \setminus \{r\}) \subseteq \mathrm{cl}(R') = \mathrm{cl}(R'')$ and therefore, $\mathrm{cl}(R' \setminus \{r\}) = \mathrm{cl}(R')$. ■

Definition 4.5. *Let R be a consistent triple set, $\mathtt{ab|c} \in R$ and $\mathcal{L} \subseteq L_R$ such that $a, b, c \in \mathcal{L}$. The triple $\mathtt{ab|c}$ is called* bridge *in $[R, \mathcal{L}]$ if a, b are in different connected components of $[R \setminus \{\mathtt{ab|c}\}, \mathcal{L}]$.*

Lemma 4.6. *Let R be a consistent triple set and $R' \in \mathfrak{sc}(R)$. Then, $R' \in \min(\mathfrak{sc}(R))$ if and only if every $\mathtt{ab|c} \in R'$ is a bridge in $[R', A \cup B]$ with $\{A, B\} = \ell^\star_{\mathtt{ab|c}}(R')$. In particular, $[R' \setminus \{\mathtt{ab|c}\}, A \cup B]$ must have three connected components α, β, γ with $a \in \alpha$, $b \in \beta$ and $c \in \gamma$, that is, either $A = \alpha \cup \beta$ and $B = \gamma$ or $B = \alpha \cup \beta$ and $A = \gamma$.*

Proof. Let $R' \in \min(\mathfrak{sc}(R))$, $\mathtt{ab|c} \in R'$ and $\ell^\star_{\mathtt{ab|c}}(R') = \{A, B\}$. By definition, $[R', A \cup B]$ has exactly two connected components, one containing a, b and the other c. Assume for contradiction that $\mathtt{ab|c}$ is not a bridge in $[R', A \cup B]$. Thus, a and b are still connected by a walk in $[R' \setminus \{\mathtt{ab|c}\}, A \cup B]$. Note, by Lemma 2.2 the Ahograph $[R' \setminus \{\mathtt{ab|c}\}, A \cup B]$ is a subgraph of $[R', A \cup B]$ that differs from $[R', A \cup B]$ only by the edge (a, b). Therefore, $[R' \setminus \{\mathtt{ab|c}\}, A \cup B]$ still consists of the two connected components A and B, one containing a, b and the other c. Theorem 2.6 implies that $\{\mathtt{ab|c}\} \in \mathrm{cl}(R' \setminus \{\mathtt{ab|c}\})$. Lemma 2.7 implies that $\mathrm{cl}(R' \setminus \{\mathtt{ab|c}\}) = \mathrm{cl}(R') = \mathrm{cl}(R)$; a contradiction to $R' \in \min(\mathfrak{sc}(R))$.

Conversely, assume that $R' \notin \min(\mathfrak{sc}(R))$. Thus, there is some triple $\mathtt{ab|c} \in R'$ such that $\mathrm{cl}(R' \setminus \{\mathtt{ab|c}\}) = \mathrm{cl}(R)$. Since $R' \in \mathfrak{sc}(R)$, we can apply Theorem 3.10 and conclude that $\ell^\star_{\mathtt{ab|c}}(R' \setminus \{\mathtt{ab|c}\}) = \ell^\star_{\mathtt{ab|c}}(R') = \{A, B\}$. Thus, $[R' \setminus \{\mathtt{ab|c}\}, A \cup B]$ has two connected components A and B. Therefore, $\mathtt{ab|c}$ is not a bridge in $[R', A \cup B]$.

For the last statement, observe that $R' \in \min(\mathfrak{sc}(R))$ and $\{A, B\} = \ell^\star_{\mathtt{ab|c}}(R')$ implies that the graph $[R', A \cup B]$ has exactly two connected components, one containing a, b (say A) and the other (B) contains c. Since $\mathtt{ab|c} \in R'$ is a bridge in $[R', A \cup B]$, a and b are in distinct connected components α and β of $[R' \setminus \{\mathtt{ab|c}\}, A \cup B]$, respectively. However, since only the edge (a, b) has been removed from $[R', A \cup B]$ to obtain $[R' \setminus \{\mathtt{ab|c}\}, A \cup B]$ it is clear that the set A decomposes into these connected components α, β, i.e., $A = \alpha \cup \beta$. Besides the edge (a, b) no other edge has been removed or added to $[R' \setminus \{\mathtt{ab|c}\}, A \cup B]$ and thus, $B = \gamma$ is still a connected component in $[R' \setminus \{\mathtt{ab|c}\}, A \cup B]$ with $c \in \gamma$. ■

We are now in the position to show that is a matroid.

Theorem 4.7. *If R is a consistent triple set, then $(R, \mathbb{F}_R)$ is a matroid.*

Proof. In order to show that $(R, \mathbb{F}_R)$ is a matroid, we show that its collection of bases $\mathcal{B} = \min(\mathfrak{sc}(R))$ satisfies the Conditions (B1) and (B2) of Lemma 4.2. Recall that $(R, \mathbb{F}_R)$ is an independent system with collection of bases $\mathcal{B} = \min(\mathfrak{sc}(R))$ and $\min(\mathfrak{sc}(R)) \neq \emptyset$. Thus,

Condition (B1) is trivially satisfied. The proof of Condition (B2) consists of several steps (Claim 1 - 5).

We fix the notion as follows: We assume that $R_1, R_2 \in \mathcal{B}$, $\mathtt{ab|c} \in R_1 \setminus R_2$ and $\ell^\star_{\mathtt{ab|c}}(R_1) = \{A, B\} \in \mathfrak{L}^\star_{R_1}(R_1)$. Moreover, we will frequently make use of $\mathfrak{L}^\star_{R_1}(R_1) = \mathfrak{L}^\star_R(R) = \mathfrak{L}^\star_{R_2}(R_2)$, which is because of $\mathrm{cl}(R_1) = \mathrm{cl}(R) = \mathrm{cl}(R_2)$ and Theorem 3.10. Furthermore, Lemma 4.6 implies that $\mathtt{ab|c}$ is a bridge in $[R_1, A \cup B]$ and that $[R_1 \setminus \{\mathtt{ab|c}\}, A \cup B]$ decomposes into the connected components α, β, γ with $a \in \alpha$, $b \in \beta$ and $c \in \gamma$. W.l.o.g. we will assume that $a, b \in A, c \in B$ and thus, $A = \alpha \cup \beta$ and $B = \gamma$.

Claim 1: *There exists a triple* $\mathtt{a'b'|c'} \in R_2$ *with* $a' \in \alpha$, $b' \in \beta$ *and* $c' \in \gamma$.

Proof of Claim 1. We begin by showing that there is a triple $\mathtt{a'b'|c'} \in R_2$ such that $a' \in \alpha$, $b' \in \beta$ and $c' \in A \cup B$ and then show that $c' \in \gamma$.

Assume for contradiction that there is no triple $\mathtt{a'b'|c'} \in R_2$ such that $a' \in \alpha$, $b' \in \beta$ and $c' \in A \cup B$. Hence, there is no edge (x, y) in $[R_2, A \cup B]$ for any $x \in \alpha$ and $y \in \beta$, that is, $\langle A \rangle$ is disconnected in $[R_2, A \cup B]$. But then $\{A, B\} \notin \mathfrak{L}^\star_{R_2}(R_2)$; contradicting $\mathfrak{L}^\star_{R_1}(R_1) = \mathfrak{L}^\star_{R_2}(R_2)$. Thus there is a triple $\mathtt{a'b'|c'} \in R_2$ such that $a' \in \alpha$, $b' \in \beta$ and $c' \in A \cup B$.

We continue to show that $c' \in \gamma$. Assume for contradiction that $c' \notin \gamma$. Since $\gamma = B$, we have $c' \in A$. Let $\ell^\star_{\mathtt{a'b'|c'}}(R_2) = \{A', B'\} \in \mathfrak{L}^\star_{R_2}(R_2)$. Note, since $\mathfrak{L}^\star_{R_1}(R_1) = \mathfrak{L}^\star_{R_2}(R_2)$ we also have $\{A', B'\} \in \mathfrak{L}^\star_{R_1}(R_1)$ and hence, the graph $[R_1, A' \cup B']$ has the two connected components A' and B', one containing a', b' and the other c'. Furthermore, since $a', b', c' \in A$ we have $A \cap A' \neq \emptyset$ and $A \cap B' \neq \emptyset$. Hence, $A' \cap (A \cup B) \neq \emptyset$ and $B' \cap (A \cup B) \neq \emptyset$, and we can apply Lemma 3.7 to conclude that $A' \cup B' \subseteq A$. Therefore, Lemma 2.2 implies that $[R_1, A' \cup B'] \subseteq [R_1, A \cup B]$. In particular, both $\langle A' \rangle \subseteq \langle A \rangle$ and $\langle B' \rangle \subseteq \langle A \rangle$ are connected subgraphs in $[R_1, A \cup B]$. Since $c \notin A$ we have $c \notin A' \cup B'$ and thus, $[R_1 \setminus \{\mathtt{ab|c}\}, A' \cup B'\}] = [R_1, A' \cup B']$. Hence, $\langle A' \rangle$ and $\langle B' \rangle$ remain connected subgraphs in $[R_1 \setminus \{\mathtt{ab|c}\}, A \cup B\}]$. Since $\ell^\star_{\mathtt{a'b'|c'}}(R_2) = \{A', B'\}$ it holds that either $a', b' \in A'$ and $c' \in B'$ or $a', b' \in B'$ and $c' \in A'$. Assume that $a', b' \in A'$. By choice of $\mathtt{a'b'|c'} \in R_2$ we have $a' \in \alpha$ and $b' \in \beta$. Since A', α and β induce a connected subgraph in $[R_1 \setminus \{\mathtt{ab|c}\}, A \cup B\}]$, respectively, and since $a' \in A' \cap \alpha$ and $b' \in A' \cap \beta$, the induced subgraph $\langle A' \cup \alpha \cup \beta \rangle$ is connected in $[R_1 \setminus \{\mathtt{ab|c}\}, A \cup B\}]$. However, since $a \in \alpha$ and $b \in \beta$, the triple $\mathtt{ab|c}$ is not a bridge in $[R_1, A \cup B]$; a contradiction to Lemma 4.6. By analogous arguments one obtains a contradiction if $a', b' \in B'$. Therefore, there is a triple $\mathtt{a'b'|c'} \in R_2$ such that $a' \in \alpha$, $b' \in \beta$ and $c' \in \gamma$.

– End Proof Claim 1 –

In what follows, let $\mathtt{a'b'|c'} \in R_2$ be chosen such that $a' \in \alpha$, $b' \in \beta$ and $c' \in \gamma$.

Claim 2: *It holds that* $\mathtt{a'b'|c'} \in R_2 \setminus R_1$.

Proof of Claim 2. Recall that $\mathtt{ab|c} \in R_1 \setminus R_2$ and thus, the triples $\mathtt{ab|c}$ and $\mathtt{a'b'|c'}$ must be distinct. Assume for contradiction that $\mathtt{a'b'|c'} \in R_1$. In this case, one can easily verify that there are either two edges (a, b) and (a', b') in $[R_1, A \cup B]$ connecting α and β or, if $(a, b) = (a', b')$, then the edge (a, b) is supported by two triples. In either case, $\mathtt{ab|c}$ is not a bridge in $[R_1, A \cup B]$; a contradiction to Lemma 4.6.

– End Proof Claim 2 –

In what follows, we set $R_{\mathrm{new}} \coloneqq (R_1 \setminus \{\mathtt{ab|c}\}) \cup \{\mathtt{a'b'|c'}\}$.

Claim 3: *It holds that* $R_{\mathrm{new}} \in \mathfrak{sc}(R)$.

Proof of Claim 3. Clearly, $R_{\mathrm{new}} \subseteq R$. Hence, in order to show that $R_{\mathrm{new}} \in \mathfrak{sc}(R)$ it remains to show that $\mathrm{cl}(R_{\mathrm{new}}) = \mathrm{cl}(R)$. To this end, recap that $[R_1 \setminus \{\mathtt{ab|c}\}, A \cup B]$ has the connected components α, β, γ with $a, a' \in \alpha$, $b, b' \in \beta$ and $c, c' \in \gamma$. Moreover, Lemma 2.2 implies that $[R_1 \setminus \{\mathtt{ab|c}\}, A \cup B]$ is a subgraph of $[R_{\mathrm{new}}, A \cup B]$ and thus, $\langle \alpha \rangle$ and $\langle \beta \rangle$ remain connected subgraphs in $[R_{\mathrm{new}}, A \cup B]$. However, since $a', b', c' \in A \cup B$ and $\mathtt{a'b'|c'} \in R_{\mathrm{new}}$ we have an additional edge in $[R_{\mathrm{new}}, A \cup B]$ that connects $\langle \alpha \rangle$ and $\langle \beta \rangle$ by the edge (a', b'). Hence, $A = \alpha \cup \beta$ induces a connected subgraph in $[R_{\mathrm{new}}, A \cup B]$, while $\langle B \rangle = \langle \gamma \rangle$ remains unchanged and thus still provides a connected component in $[R_{\mathrm{new}}, A \cup B]$. In summary, $[R_{\mathrm{new}}, A \cup B]$ has two connected components, where $a, b \in A$ and $c \in B$. Theorem 2.6 implies that $\mathtt{ab|c} \in \mathrm{cl}(R_{\mathrm{new}})$. Application of Lemma 2.7 yields $\mathrm{cl}(R_{\mathrm{new}}) = \mathrm{cl}(R_1 \setminus \{\mathtt{ab|c}\}) \cup \{\mathtt{a'b'|c'}\}) = \mathrm{cl}(R_1 \cup \{\mathtt{a'b'|c'}\})$. Moreover, it holds that $\mathrm{cl}(R) = \mathrm{cl}(R_1) \subseteq \mathrm{cl}(R_1 \cup \{\mathtt{a'b'|c'}\}) = \mathrm{cl}(R_{\mathrm{new}}) \subseteq \mathrm{cl}(R)$ and therefore, $\mathrm{cl}(R) = \mathrm{cl}(R_{\mathrm{new}})$. Thus, $R_{\mathrm{new}} \in \mathfrak{sc}(R)$.

– End Proof Claim 3 –

In what follows, we want to show that all triples $\mathtt{xy|z} \in R_{\text{new}}$ are bridges in $[R_{\text{new}}, A'' \cup B'']$ where $\ell^\star_{\mathtt{xy|z}}(R_{\text{new}}) = \{A'', B''\}$ (see Claim 5). In this case, Lemma 4.6 would imply that $R_{\text{new}} \in \min(\mathfrak{sc}(R))$. To this end, however, we first need to prove Claim 4.

Claim 4: *Assume there is a triple* $\mathtt{xy|z} \in R_{\text{new}}$ *which is not a bridge in* $[R_{\text{new}}, A'' \cup B'']$ *where* $\ell^\star_{\mathtt{xy|z}}(R_{\text{new}}) = \{A'', B''\}$. *Then,* $\mathtt{xy|z} \neq \mathtt{a'b'|c'}$; $a', b', c' \in A'' \cup B''$; x *and* y *are connected by a path in in* $[R_{\text{new}} \setminus \{\mathtt{xy|z}\}, A'' \cup B'']$ *and every path* P_{xy} *in* $[R_{\text{new}} \setminus \{\mathtt{xy|z}\}, A'' \cup B'']$ *contains the edge* (a', b'); *and* $\{A'', B''\} \neq \{A, B\}$.

Proof of Claim 4. Assume that $\mathtt{xy|z} \in R_{\text{new}}$ is not a bridge in $[R_{\text{new}}, A'' \cup B'']$. First, we show that $\mathtt{xy|z} \neq \mathtt{a'b'|c'}$. Assume for contradiction that $\mathtt{xy|z} = \mathtt{a'b'|c'}$. Now, we show that in this case $\{A'', B''\} = \{A, B\}$. Note, since $\ell^\star_{\mathtt{a'b'|c'}}(R_{\text{new}}) = \{A'', B''\}$ and $R_{\text{new}} \in \mathfrak{sc}(R)$, we can apply Theorem 3.10 and conclude that $\{A'', B''\} = \ell^\star_{\mathtt{a'b'|c'}}(R)$. Since $\ell^\star_{\mathtt{ab|c}}(R) = \{A, B\}$, we have $\{A, B\}, \{A'', B''\} \in \mathfrak{L}^\star_R(R)$. Moreover, since by construction $a, b, a', b' \in A$ and $c' \in B$, we have $A \cap (A'' \cup B'') \neq \emptyset$ and $B \cap (A'' \cup B'') \neq \emptyset$. Hence, we can argue analogously as in the last part of the proof of Theorem 3.3 to conclude that $\{A, B\} = \{A'', B''\}$. Now, since $\mathtt{xy|z} = \mathtt{a'b'|c'}$, the triple $\mathtt{a'b'|c'}$ is not a bridge in $[R_{\text{new}}, A'' \cup B'']$. Thus, there is a path $P_{a'b'}$ in $[R_{\text{new}} \setminus \{\mathtt{a'b'|c'}\}, A'' \cup B''] = [R_{\text{new}} \setminus \{\mathtt{a'b'|c'}\}, A \cup B] = [R_1 \setminus \{\mathtt{ab|c}\}, A \cup B]$. However, this implies that $P_{a'b'}$ connects $a' \in \alpha$ and $b' \in \beta$ in $[R_1 \setminus \{\mathtt{ab|c}\}, A \cup B]$ and therefore, $\mathtt{ab|c}$ is not a bridge in $[R_1, A \cup B]$; a contradiction to $R_1 \in \min(\mathfrak{sc}(R))$ and Lemma 4.6. Hence $\mathtt{xy|z} \neq \mathtt{a'b'|c'}$.

We continue to show that every path P_{xy} in $[R_{\text{new}} \setminus \{\mathtt{xy|z}\}, A'' \cup B'']$ contains the edge (a', b'). Since $\mathtt{xy|z} \in R_{\text{new}}$ is not a bridge in $[R_{\text{new}}, A'' \cup B'']$ there must be a path P_{xy} in $[R_{\text{new}} \setminus \{\mathtt{xy|z}\}, A'' \cup B'']$. Assume for contradiction that P_{xy} does not contain the edge (a', b'). Hence, P_{xy} still connects x and y in $[R_{\text{new}} \setminus \{\mathtt{xy|z}, \mathtt{a'b'|c'}\}, A'' \cup B'']$. Since $R_{\text{new}} \setminus \{\mathtt{xy|z}, \mathtt{a'b'|c'}\} \subseteq R_1 \setminus \{\mathtt{xy|z}\}$ and by Lemma 2.2, the graph $[R_{\text{new}} \setminus \{\mathtt{xy|z}, \mathtt{a'b'|c'}\}, A'' \cup B'']$ is a subgraph of $[R_1 \setminus \{\mathtt{xy|z}\}, A'' \cup B'']$. Therefore, the path P_{xy} connects x and y in $[R_1 \setminus \{\mathtt{xy|z}\}, A'' \cup B'']$. Note, Theorem 3.10 implies that $\ell^\star_{\mathtt{xy|z}}(R_1) = \{A'', B''\}$. Since $\mathtt{a'b'|c'} \neq \mathtt{xy|z} \in R_{\text{new}}$, we have $\mathtt{xy|z} \in R_1$. But then $\mathtt{xy|z}$ is not a bridge in $[R_1, A'' \cup B'']$; a contradiction to Lemma 4.6.

The latter, in particular, implies that $a', b' \in A'' \cup B''$. Now, assume for contradiction that $c' \notin A'' \cup B''$. Hence, $[R_{\text{new}} \setminus \{\mathtt{xy|z}, \mathtt{a'b'|c'}\}, A'' \cup B''] = [R_{\text{new}} \setminus \{\mathtt{xy|z}\}, A'' \cup B'']$. By the preceding arguments, every path P_{xy} in $[R_{\text{new}} \setminus \{\mathtt{xy|z}\}, A'' \cup B'']$ contains the edge (a', b'). Again, since $[R_{\text{new}} \setminus \{\mathtt{xy|z}, \mathtt{a'b'|c'}\}, A'' \cup B'']$ is a subgraph of $[R_1 \setminus \{\mathtt{xy|z}\}, A'' \cup B'']$ this path is also contained in $[R_1 \setminus \{\mathtt{xy|z}\}, A'' \cup B'']$ and the triple $\mathtt{xy|z}$ is not a bridge in $[R_1, A'' \cup B'']$; a contradiction to Lemma 4.6 and $\ell^\star_{\mathtt{xy|z}}(R_1) = \{A'', B''\}$. Thus, $c' \in A'' \cup B''$

Finally, we show that $\{A'', B''\} \neq \{A, B\}$. Assume for contradiction that $\{A'', B''\} = \{A, B\}$. W.l.o.g. let $A'' = A = \alpha \cup \beta$ and $B'' = B = \gamma$. First, we show that neither $x \in \gamma$ nor $y \in \gamma$. Assume w.l.o.g. that $x \in \gamma$. Note the path P_{xy} with edge (a', b') in $[R_{\text{new}} \setminus \{\mathtt{xy|z}\}, A'' \cup B'']$ is also contained $[R_{\text{new}}, A'' \cup B'']$. However, since $a', b' \in A''$ and $x \in \gamma = B''$ this path P_{xy} connects the two connected components A'', B'' in $[R_{\text{new}}, A'' \cup B'']$; a contradiction.

We continue to show that neither $x, y \in \alpha$ nor $x, y \in \beta$. Assume for contradiction that $x, y \in \alpha$. Since $[R_{\text{new}} \setminus \{\mathtt{xy|z}\}, A'' \cup B'']$ contains a path P_{xy} with edge (a', b'), and $x, y \in \alpha$ there must be a second edge (a'', b'') distinct from (a', b') in P_{xy} where $a'' \in \alpha, b'' \in \beta$. Since $R_{\text{new}} \setminus \{\mathtt{xy|z}\} = (R_1 \setminus \{\mathtt{ab|c}, \mathtt{xy|z}\}) \cup \{\mathtt{a'b'|c'}\}$, $\{A'', B''\} = \{A, B\}$ and removal of $\{\mathtt{a'b'|c'}\}$ would still preserve the edge (a'', b''), this edge (a'', b'') must also be contained in $[R_1 \setminus \{\mathtt{ab|c}, \mathtt{xy|z}\}, A \cup B]$. Since $[R_1 \setminus \{\mathtt{ab|c}, \mathtt{xy|z}\}, A \cup B]$ is a subgraph of $[R_1 \setminus \{\mathtt{ab|c}\}, A \cup B]$, the latter graph contains the edge (a'', b'') that connects the components α and β. But then $\mathtt{ab|c}$ is not a bridge in $[R_1, A \cup B]$; a contradiction to $R_1 \in \min(\mathfrak{sc}(R))$ and Lemma 4.6. Hence, x and y cannot be both in α, and by similar arguments, not both in β.

Thus, there are only two cases left: $x \in \alpha$ and $y \in \beta$, or $y \in \alpha$ and $x \in \beta$. Assume w.l.o.g. that $x \in \alpha$ and $y \in \beta$. Since $\mathtt{xy|z} \in R_1 \setminus \{\mathtt{ab|c}\}$, there must be the edge (x, y) in $[R_1 \setminus \{\mathtt{ab|c}\}, A \cup B]$, in which case α and β form a connected component. Again, $\mathtt{ab|c}$ is not a bridge in $[R_1, A \cup B]$ and we obtain a contradiction to $R_1 \in \min(\mathfrak{sc}(R))$ and Lemma 4.6.

Therefore, if $\{A'', B''\} = \{A, B\}$, then $x, y \notin \alpha \cup \beta \cup \gamma = A'' \cup B''$; a contradiction since we assumed that $\ell^\star_{\mathtt{xy|z}}(R_{\text{new}}) = \{A'', B''\}$ and hence, $x, y \in A'' \cup B''$. – End Proof Claim 4 –

Claim 5: $R_{\text{new}} \in \min(\mathfrak{sc}(R))$.

Proof of Claim 5. In order to show that $R_{\text{new}} \in \min(\mathfrak{sc}(R))$ we use Lemma 4.6 and show that each triple $\mathtt{xy|z} \in R_{\text{new}}$ must be a bridge in $[R_{\text{new}}, A'' \cup B'']$ where $\ell^\star_{\mathtt{xy|z}}(R_{\text{new}}) =$

$\{A'', B''\}$.

Assume for contradiction, that there is a triple $\mathtt{xy|z} \in R_{\text{new}}$ that is not a bridge in $[R_{\text{new}}, A'' \cup B'']$. Claim 4. implies that $\mathtt{xy|z} \neq \mathtt{a'b'|c'}$ and thus, in particular, $\mathtt{xy|z} \in R_1 \setminus \{\mathtt{ab|c}\}$. Moreover, $a', b', c' \in A'' \cup B''$, $\{A'', B''\} \neq \{A, B\}$ and every path P_{xy} in $[R_{\text{new}} \setminus \{\mathtt{xy|z}\}, A'' \cup B'']$ contains the edge (a', b'). Recap that $a, a' \in \alpha$, $b, b' \in \beta$, $c, c' \in \gamma$, $A = \alpha \cup \beta$ and $B = \gamma$.

Recap that $\ell^\star_{\mathtt{ab|c}}(R_1) = \{A, B\} \in \mathfrak{L}^\star_{R_1}(R_1)$. Claim 3 implies that $R_{\text{new}} \in \mathfrak{sc}(R)$. Thus, we can apply Theorem 3.10 and conclude that $\{A'', B''\} = \ell^\star_{\mathtt{xy|z}}(R_{\text{new}}) = \ell^\star_{\mathtt{xy|z}}(R) = \ell^\star_{\mathtt{xy|z}}(R_1)$. Hence, $\{A'', B''\} \in \mathfrak{L}^\star_{R_1}(R_1)$. Moreover, since $a', b', c' \in A'' \cup B''$ as well as $a', b' \in A$ and $c' \in B$ it holds that $A \cap (A'' \cup B'') \neq \emptyset$ and $B \cap (A'' \cup B'') \neq \emptyset$. Thus, we can apply Lemma 3.7 and conclude that $A \cup B \subseteq A''$ or $A \cup B \subseteq B''$. W.l.o.g. assume $A \cup B \subseteq A''$.

Denote one of the paths in $[R_{\text{new}} \setminus \{\mathtt{xy|z}\}, A'' \cup B'']$ that connect x and y by P_{xy}. Claim 4 implies that P_{xy} contains the edge (a', b'). Since $a', b' \in A''$ it must hold that $x, y \in A''$ as otherwise P_{xy} would connect A'' and B'' in $[R_{\text{new}}, A'' \cup B'']$. Therefore, $z \in B''$ and hence $z \notin A \cup B$. Since P_{xy} contains the edge (a', b'), it can be decomposed into the paths $P_{xa'}$ and $P_{b'y}$ (resp. $P_{xb'}$ and $P_{ya'}$) and the edge (a', b'). W.l.o.g. assume that P_{xy} is composed of $P_{xa'}$, (a', b') and $P_{b'y}$. Note, since neither $P_{xa'}$ nor $P_{b'y}$ contains the edge (a', b'), we can conclude that both paths are contained in $[R_{\text{new}} \setminus \{\mathtt{xy|z}, \mathtt{a'b'|c'}\}, A'' \cup B''] = [R_1 \setminus \{\mathtt{xy|z}, \mathtt{ab|c}\}, A'' \cup B''] \subseteq [R_1 \setminus \{\mathtt{xy|z}\}, A'' \cup B'']$. Furthermore, since α and β induce connected subgraphs in $[R_1 \setminus \mathtt{ab|c}, A \cup B]$ and $a, a' \in \alpha$, $b, b' \in \beta$, there are paths $P_{aa'}$ and $P_{bb'}$ in $[R_1 \setminus \{\mathtt{ab|c}\}, A \cup B]$. Since $z \notin A \cup B$ and $A \cup B \subseteq A''$, we have $[R_1 \setminus \{\mathtt{ab|c}\}, A \cup B] = [R_1 \setminus \{\mathtt{ab|c}, \mathtt{xy|z}\}, A \cup B] \subseteq [R_1 \setminus \{\mathtt{xy|z}\}, A'' \cup B'']$. Hence, the paths $P_{aa'}$ and $P_{bb'}$ are also contained in $[R_1 \setminus \{\mathtt{xy|z}\}, A'' \cup B'']$. In summary, $[R_1 \setminus \{\mathtt{xy|z}\}, A'' \cup B'']$ contains the paths $P_{aa'}$, $P_{bb'}$, $P_{xa'}$ and $P_{b'y}$ but also the edge (a, b), since $\mathtt{ab|c} \in R_1 \setminus \{\mathtt{xy|z}\}$ and $a, b, c \in A \cup B \subseteq A''$. Hence, we can combine the four paths and the edge (a, b) to a walk in $[R_1 \setminus \{\mathtt{xy|z}\}, A'' \cup B'']$ that connects x and y. However, this implies that $\mathtt{xy|z}$ is not a bridge in $[R_1, A'' \cup B'']$; a contradiction to $\ell^\star_{\mathtt{xy|z}}(R_1) = \{A'', B''\}$ and Lemma 4.6.

In summary, for all cases for which there is a triple $\mathtt{xy|z} \in R_{\text{new}}$ that is not a bridge in $[R_{\text{new}}, A'' \cup B'']$ we obtain a contradiction. Hence, each triple $\mathtt{xy|z} \in R_{\text{new}}$ must be a bridge in $[R_{\text{new}}, A'' \cup B'']$ and we can apply Lemma 4.6 to conclude that $R_{\text{new}} \in \min(\mathfrak{sc}(R))$.

– End Proof Claim 5 –

We have shown that for any $R_1, R_2 \in \mathcal{B} = \min(\mathfrak{sc}(R))$ and $\mathtt{ab|c} \in R_1 \setminus R_2$ there is a triple $\mathtt{a'b'|c'} \in R_2 \setminus R_1$ such that $R_{\text{new}} = (R_1 \setminus \{\mathtt{ab|c}\}) \cup \{\mathtt{a'b'|c'}\} \in \mathcal{B}$. Hence, we can apply Lemma 4.2 to conclude that $(R, \mathbb{F}_R)$ is a matroid. ■

In order to avoid confusion, we emphasize that the closure operator $\mathrm{cl}(R)$ for rooted triple sets R defined here is not a matroid closure operator cl_M [40, 50]. Note, since $M = (R, \mathbb{F}_R)$ is a matroid, the following property must be satisfied for cl_M, cf. [50, Lemma 1.4.3]:

$$X \subseteq R, r \in R \text{ and } r' \in \mathrm{cl}_M(X \cup \{r\}) \setminus \mathrm{cl}_M(X) \implies r \in \mathrm{cl}_M(X \cup \{r'\}).$$

To see that $\mathrm{cl}(R)$ does not fulfill this property in general, consider the example in Figure 1. To recap, $R_1 = \{\mathtt{ab|c}, \mathtt{ac|d}\}$, $R_3 = \{\mathtt{ac|d}, \mathtt{bc|d}\}$ and $\mathrm{cl}(R_1) = \mathrm{cl}(R) = \{\mathtt{ab|c}, \mathtt{ac|d}, \mathtt{bc|d}, \mathtt{ab|d}\}$, but $\mathrm{cl}(R_3) = \mathrm{cl}(R) \setminus \mathtt{ab|c}$. Now, put $X = \{\mathtt{ac|d}\}$ and $r = \mathtt{ab|c}$. Thus, $r' = \mathtt{bc|d} \in \mathrm{cl}(X \cup \{r\}) \setminus \mathrm{cl}(X) = \mathrm{cl}(R_1) \setminus \{\mathtt{ab|c}\}$. However, $r = \mathtt{ab|c} \notin \mathrm{cl}(X \cup \{r'\}) = \mathrm{cl}(R_3)$. The latter result has already been observed by David Bryant [5], however, the matroid structure of $(R, \mathbb{F}_R)$ was not discovered.

Note, each minimum representative set $R' \in \mathrm{MIN}(\mathfrak{sc}(R))$ is also minimal. Thus, $\mathrm{MIN}(\mathfrak{sc}(R)) \subseteq \min(\mathfrak{sc}(R))$. However, since $(R, \mathbb{F}_R)$ is a matroid with collection of bases $\min(\mathfrak{sc}(R))$, all elements in $\min(\mathfrak{sc}(R))$ have the same cardinality [50]. Therefore, all basis elements of the matroid $(R, \mathbb{F}_R)$ are of minimum size. We summarize this observation in the following

Theorem 4.8. *If R is a consistent triple set, then* $\min(\mathfrak{sc}(R)) = \mathrm{MIN}(\mathfrak{sc}(R))$.

In order to find a minimum representative set R' of R one can apply a simple greedy algorithm. Algorithm 1 computes a basis element of the matroid $(R, \mathbb{F}_R)$ and can easily be adapted to find maximum weighted bases, an issue that might be important for applications in phylogenetics, where the weight of a rooted triple corresponds to a statistical confidence value or any other measure associated with the underlying triples.

Lemma 4.9. *Algorithm 1 computes a subset $R' \subseteq R$ with $\mathrm{cl}(R') = \mathrm{cl}(R)$ of minimum size in $\mathcal{O}(|R|^2 |L_R|)$.*

Algorithm 1 `GREEDY for Minimal/Minimum Representative Triple Sets`

Input: Consistent triple set R;
Output: Minimal Representative Triple set $R_{\min}$;

1: $R_{\text{tmp}} \leftarrow \emptyset$;
2: **for** all $\mathtt{ab|c} \in R$ **do**
3: $\quad R' \leftarrow R \setminus R_{\text{tmp}}$;
4: $\quad$ **if** $R' \setminus \{\mathtt{ab|c}\} \cup \{\mathtt{bc|a}\}$ and $R' \setminus \{\mathtt{ab|c}\} \cup \{\mathtt{ac|b}\}$ are not consistent **then** ▷ Thus, $\mathtt{ab|c} \in \text{cl}(R \setminus (R_{\text{tmp}} \cup \{\mathtt{ab|c}\}))$
5: $\quad\quad R_{\text{tmp}} \leftarrow R_{\text{tmp}} \cup \{\mathtt{ab|c}\}$;
6: **return** $R_{\min} \leftarrow R \setminus R_{\text{tmp}}$;

Proof. By Lemma 2.5, it suffices to decide whether a triple $\mathtt{ab|c}$ is contained in $\text{cl}(R \setminus (R_{\text{tmp}} \cup \{\mathtt{ab|c}\}))$ by the two consistency checks in the IF-condition.

Let $R_{\text{tmp}} = \{r_1, \ldots, r_k\}$ where the indices of the triples are chosen w.r.t. the order in which they are added to R_{tmp}. By construction, $r_i \in R_{\text{tmp}}$ if $r_i \in \text{cl}(R \setminus \{r_1, \ldots, r_i\})$. Lemma 2.7 implies that for the first triple $r_1 \in \text{cl}(R \setminus \{r_1\})$ it holds that $\text{cl}(R \setminus \{r_1\}) = \text{cl}(R)$. Next, r_2 is added to R_{tmp} that is, $r_2 \in \text{cl}(R \setminus \{r_1, r_2\})$ and again by Lemma 2.7, $\text{cl}(R \setminus \{r_1, r_2\}) = \text{cl}(R \setminus \{r_1\}) = \text{cl}(R)$. Inductively, when r_k is chosen we have $r_k \in \text{cl}(R \setminus R_{\text{tmp}}\}) = \text{cl}(R \setminus \{r_1, \ldots, r_{k-1}\}) = \cdots = \text{cl}(R \setminus \{r_1\}) = \text{cl}(R)$. Since by construction, $R_{\min} = R \setminus R_{\text{tmp}}$, it holds that $\text{cl}(R_{\min}) = \text{cl}(R \setminus R_{\text{tmp}}) = \text{cl}(R)$.

We continue to show that $R_{\min} \in \min(\mathfrak{sc}(R))$. Assume for contradiction that there is a subset $R'' \subsetneq R_{\min}$ with $\text{cl}(R'') = \text{cl}(R)$. Note, $R'' = R_{\min} \setminus R'$ for some non-empty subset $R' \subseteq R_{\min}$. Thus, $\text{cl}(R_{\min} \setminus R') = \text{cl}(R) = \text{cl}(R_{\min})$. Lemma 2.7 implies that there is a triple $r \in R'$ such that $\text{cl}(R_{\min} \setminus \{r\}) = \text{cl}(R_{\min})$. Note, since $r \in R' \subseteq R_{\min} = R \setminus R_{\text{tmp}}$ it holds that $r \notin R_{\text{tmp}}$.

Consider the step when r is chosen in Alg. 1. If $R_{\text{tmp}} = \emptyset$ before this step, we would have $r \notin \text{cl}(R \setminus \{r\})$, since r is not added to R_{tmp}. However, since $r \in R_{\min}$ and $R_{\min} \setminus \{r\} \subseteq R \setminus \{r\}$ and it must hold that $r \in \text{cl}(R_{\min}) = \text{cl}(R_{\min} \setminus \{r\}) \subseteq \text{cl}(R \setminus \{r\})$; a contradiction. If R_{tmp} is not empty and thus, $R_{\text{tmp}} = \{r_1, \ldots, r_i\}$ before the step when r is chosen in Alg. 1, we would have $r \notin \text{cl}(R \setminus \{r_1, \ldots, r_i, r\})$, since r is not added to R_{tmp}. However, since $r \in R_{\min}$ and $R_{\min} \setminus \{r\} = R \setminus \{r_1, \ldots, r_k, r\} \subseteq R \setminus \{r_1, \ldots, r_i, r\}$ it must hold that $r \in \text{cl}(R_{\min}) = \text{cl}(R_{\min} \setminus \{r\}) \subseteq \text{cl}(R \setminus \{r_1, \ldots, r_i, r\})$; a contradiction. Therefore, $R_{\min}$ is minimal and we can apply Theorem 4.8 to conclude that $R_{\min}$ is of minimum size.

Concerning the time complexity, observe that the for-loop runs $|R|$ times. In each step of the for-loop, we have to check for consistency which can be done with `BUILD` in $\mathcal{O}(|R||L_R|)$ time. Thus, we end in an overall time complexity $\mathcal{O}(|R|^2|L_R|)$. ■

As a consequence, we obtain the following result:

Theorem 4.10. *Let R_1, R_2 be consistent triple sets such that $\text{cl}(R_1) = \text{cl}(R_2)$. For each $R_1' \in \min(\mathfrak{sc}(R_1))$ and $R_2' \in \min(\mathfrak{sc}(R_2))$ it holds that $|R_1'| = |R_2'|$.*

Proof. Let R_1 and R_2 be consistent triple sets such that $\text{cl}(R_1) = \text{cl}(R_2)$. Set $R = \text{cl}(R_1)$ and apply the greedy method with input R. Since $\text{cl}(R_1) = \text{cl}(R)$ and since the choice of the triples assigned to R_{tmp} is arbitrary as long as $\text{cl}(R \setminus R_{\text{tmp}}) = \text{cl}(R)$, it is possible to obtain greedily a set R_{tmp} for which $R' = R_1 = R \setminus R_{\text{tmp}} \subseteq R$ (in Step 3 of Alg. 1). Now Alg. 1 continues with R_1 in order to find a subset $R' \subseteq R_1$ such that $R' \in \min(\mathfrak{sc}(R)) = \min(\mathfrak{sc}(\text{cl}(R_1)))$. Note, $R' \in \min(\mathfrak{sc}(R_1))$ as otherwise there would be a subset $R'' \subsetneq R' \subseteq R_1 \subseteq \text{cl}(R_1)$ such that $\text{cl}(R'') = \text{cl}(R_1)$; a contradiction to $R' \in \min(\mathfrak{sc}(\text{cl}(R_1)))$ and the correctness of Alg. 1. Hence, for any $R_1' \in \min(\mathfrak{sc}(\text{cl}(R_1)))$ with $R_1' \subseteq R_1$ we also have $R_1' \in \min(\mathfrak{sc}(R_1))$. The same applies to R_2, that is, $R_2' \in \min(\mathfrak{sc}(R_2))$ for any $R_2' \in \min(\mathfrak{sc}(\text{cl}(R_2)))$ with $R_2' \subseteq R_2$. Since $\text{cl}(R_1) = \text{cl}(R_2)$, it holds that $\min(\mathfrak{sc}(\text{cl}(R_1))) = \min(\mathfrak{sc}(\text{cl}(R_2)))$. The latter together with Theorem 4.8 implies that $|R_1'| = |R_2'|$. ■

5 Computing the Closure

The currently fastest algorithm to determine the closure has a time complexity of $\mathcal{O}(|R||L_R|^4)$ and was proposed by Bryant and Steel [6]. In this section, we provide a novel and efficient algorithm to compute the closure. This method is based on the techniques we used to prove the

matroid structure. In particular, the proposed algorithm will rely on computing the set $\mathfrak{L}^\star_R(R)$ and usage of the following theorem.

Theorem 5.1. *Let R be a consistent triple set and define $\mathcal{R}_{A,B} = \{\mathtt{ab|c}\colon a,b \in A, c \in B$ or $a,b \in B, c \in A\}$ for any $A, B \subseteq L_R$. Then,*

$$\mathrm{cl}(R) = \bigcup_{\{A,B\}\in\mathfrak{L}^\star_R(R)} \mathcal{R}_{A,B}.$$

Moreover, for any distinct $\{A,B\},\{A',B'\} \in \mathfrak{L}^\star_R(R)$ it holds that $\mathcal{R}_{A,B} \cap \mathcal{R}_{A',B'} = \emptyset$. In particular,

$$\sum_{\{A,B\}\in\mathfrak{L}^\star_R(R)} |R_{A,B}| \leq |L_R|^3.$$

Proof. Theorem 2.6 immediately implies that $\bigcup_{\{A,B\}\in\mathfrak{L}^\star_R(R)} \mathcal{R}_{A,B} \subseteq \mathrm{cl}(R)$. Thus, it remains to show that $\mathrm{cl}(R) \subseteq \bigcup_{\{A,B\}\in\mathfrak{L}^\star_R(R)} \mathcal{R}_{A,B}$. Let $\mathtt{ab|c} \in \mathrm{cl}(R)$. Lemma 3.3 implies that $\mathfrak{L}_{\mathtt{ab|c}}(R) \neq \emptyset$. Thus, there is also a maximal element $\ell^\star_{\mathtt{ab|c}}(R) = \{A,B\} \in \mathfrak{L}^\star_{\mathrm{cl}(R)}(R)$. Theorem 3.9 implies that $\mathfrak{L}^\star_{\mathrm{cl}(R)}(R) = \mathfrak{L}^\star_R(R)$ and hence, it particularly holds that $\{A,B\} \in \mathfrak{L}^\star_R(R)$ for which $\mathtt{ab|c} \in \mathcal{R}_{A,B}$. Therefore, $\mathtt{ab|c} \in \bigcup_{\{A,B\}\in\mathfrak{L}^\star_R(R)} \mathcal{R}_{A,B}$ and thus, $\mathrm{cl}(R) = \bigcup_{\{A,B\}\in\mathfrak{L}^\star_R(R)} \mathcal{R}_{A,B}$.

We continue by showing that for any distinct $\{A,B\},\{A',B'\} \in \mathfrak{L}^\star_R(R)$ we have $\mathcal{R}_{A,B} \cap \mathcal{R}_{A',B'} = \emptyset$. Assume for contradiction that $\mathtt{ab|c} \in \mathcal{R}_{A,B} \cap \mathcal{R}_{A',B'}$ for some distinct $\{A,B\},\{A',B'\} \in \mathfrak{L}^\star_R(R)$. Thus, $A \cap (A' \cup B') \neq \emptyset$ and $B \cap (A' \cup B') \neq \emptyset$, as well as $A' \cap (A \cup B) \neq \emptyset$ and $B' \cap (A \cup B) \neq \emptyset$. Lemma 3.7 implies that either $A \cup B \subseteq A'$ or $A \cup B \subseteq B'$ and either $A' \cup B' \subseteq A$ or $A' \cup B' \subseteq B$. W.l.o.g. assume that $A' \cup B' \subseteq A$ and therefore, $A', B' \subseteq A$. If $A \cup B \subseteq A'$ (resp. $A \cup B \subseteq B'$), then $B' \subseteq A'$ (resp. $A' \subseteq B'$); a contradiction to the disjointedness of A', B'. Hence, $\mathcal{R}_{A,B} \cap \mathcal{R}_{A',B'} = \emptyset$.

Finally, since $\mathcal{R}_{A,B} \cap \mathcal{R}_{A',B'} = \emptyset$ are disjoint for distinct $\{A,B\},\{A',B'\} \in \mathfrak{L}^\star_R(R)$, the closure $\mathrm{cl}(R)$ is the disjoint union $\uplus_{\{A,B\}\in\mathfrak{L}^\star_R(R)}\mathcal{R}_{A,B}$ and therefore, $|\mathrm{cl}(R)| = \sum_{\{A,B\}\in\mathfrak{L}^\star_R(R)} |R_{A,B}|$. Since $\mathrm{cl}(R)$ can have at most $|L_R|^3$ triples, that is, one triple for each of three-element subsets of L_R, the assertion follows. ■

Lemma 5.2. *Let R be a consistent triple set and $\mathtt{ab|c} \in R$. Moreover, assume that there is a subset $\mathcal{L} \subseteq L_R$ with $a,b,c \in \mathcal{L}$ such that a,b are contained together in some connected component $\mathcal{C}_{a,b}$ of $[R,\mathcal{L}]$. Let $\mathcal{C}_c$ denote the connected component in $[R,\mathcal{L}]$ that contains c. Note, we don't claim that $\mathcal{C}_{a,b} \neq \mathcal{C}_c$.*

Then, $\mathcal{L}' \subseteq \mathcal{C}_{a,b} \cup \mathcal{C}_c$ for all $\mathcal{L}' \subseteq \mathcal{L}$ for which $[R,\mathcal{L}']$ has exactly two connected components, one containing a,b and the other c.

Proof. Assume for contradiction that there is a subset $\mathcal{L}' = A \cup B \subseteq \mathcal{L}$ such that $[R,\mathcal{L}']$ has exactly two connected components A and B with $a,b \in A$ and $c \in B$, but $\mathcal{L}' \not\subseteq \mathcal{C}_{a,b} \cup \mathcal{C}_c$. Thus, there is a vertex $d \in \mathcal{L}' \setminus (\mathcal{C}_{a,b} \cup \mathcal{C}_c)$. Therefore, d is either contained in A or in B, that is, there is either a path P_{da} or P_{dc} in $[R,\mathcal{L}']$. Since $[R,\mathcal{L}']$ is a subgraph of $[R,\mathcal{L}]$ these paths are also contained in $[R,\mathcal{L}]$. But then, $d \in \mathcal{C}_{a,b}$ or $d \in \mathcal{C}_c$; a contradiction. ■

The latter lemma immediately offers a way to compute $\mathfrak{L}^\star_R(R)$ that is summarized in Algorithm 2. For each triple $\mathtt{ab|c} \in R$ start with $[R,L_R]$. If $[R,L_R]$ has already two connected components A and B, one containing a,b and the other c, then $L_R = A \cup B$ clearly maximizes $|A \cup B|$. Thus, $\{A,B\} = \ell^\star_{\mathtt{ab|c}}(R) \in \mathfrak{L}^\star_R(R)$. If $[R,L_R]$ does not have these two connected components A and B, it is, however, still disconnected (cf. Theorem 2.1). Hence, $[R,L_R]$ has two or more connected components. Nevertheless, a,b must be in one connected component $\mathcal{C}_{a,b}$ due to the edge (a,b) implied by $\mathtt{ab|c} \in R$. Moreover, there is a connected component $\mathcal{C}_c$ that contains c. Note, $\mathcal{C}_{a,b} = \mathcal{C}_c$ might be possible. Now set $\mathcal{C} = \mathcal{C}_{a,b} \cup \mathcal{C}_c \subsetneq L_R$. By Lemma 5.2 it holds $\mathcal{L}' \subseteq \mathcal{C}$ for all $\mathcal{L}' \subseteq L_R$ for which $[R,\mathcal{L}']$ satisfies the conditions of Theorem 2.6 when applied to $\mathtt{ab|c} \in R$. Hence, we stepwisely look at these components $\mathcal{C}$ until we have found one $\mathcal{C}$ such that for the particular subset $\mathcal{C}^* = \mathcal{C}_{a,b} \cup \mathcal{C}_c \subsetneq \mathcal{C}$, the Ahograph $[R,\mathcal{C}^*]$ has two connected components A and B, one containing a,b and the other c. Hence, $\mathcal{C}^* = A \cup B$. Since this is the first occurrence of such a set $\mathcal{C}^* \subseteq L_R$ and any further $\mathcal{L}' \subseteq \mathcal{L}$ for which $[R,\mathcal{L}']$ has exactly two connected components, one containing a,b and the other c, must be contained in $\mathcal{C}^*$, $\mathcal{C}^* = A \cup B$ maximizes $|A \cup B|$. Thus, $\{A,B\} = \ell^\star_{\mathtt{ab|c}}(R) \in \mathfrak{L}^\star_R(R)$. By Theorem 2.6 and since $\mathtt{ab|c} \in R \subseteq \mathrm{cl}(R)$, there is indeed a subset $[R,\mathcal{L}']$ that satisfies the conditions of Theorem 2.6 when applied to $\mathtt{ab|c} \in R$. The latter arguments show that Algorithm 2 is correct.

Algorithm 2 `Compute` $\mathfrak{L}^\star_R(R)$

Input: A consistent triple set R;
Output: $\mathfrak{L}^\star_R(R)$;

```
1: L*_R(R) ← ∅;
2: for all ab|c ∈ R do
3:     C ← L_R;
4:     while [R, C] does not have exactly two connected components A, B, one containing a, b and the other c do
5:         C_{a,b} ← connected component in [R, C] that contains a, b;
6:         C_c ← connected component in [R, C] that contains c;
7:         C ← C_{a,b} ∪ C_c;
8:     L*_R(R) ← L*_R(R) ∪ {A, B};
9: return L*_R(R) ← L*_R(R);
```

Algorithm 3 `Compute Closure` $\mathrm{cl}(R)$

Input: A consistent triple set R;
Output: $\mathrm{cl}(R)$;

```
1: Compute L*_R(R) with Algorithm 2;
2: cl(R) ← ∅;
3: for all {A, B} ∈ L*_R(R) do
4:     Compute R_{A,B} (cf. Theorem 5.1);
5:     cl(R) ← cl(R) ∪ R_{A,B};
6: return cl(R);
```

Lemma 5.3. *Let R be a consistent triple set. Algorithm 2 computes $\mathfrak{L}^\star_R(R)$ in $\mathcal{O}(|R|^2|L_R|)$ time.*

Proof. The correctness of the Algorithms follows from Lemma 5.2 and the discussion above.

The FOR-loop runs $\mathcal{O}(|R|)$ times. The WHILE-condition is repeated at most $|L_R|$ times, since $\mathcal{C}$ will in each step have at least one vertex less as otherwise $[R, \mathcal{C}]$ will be connected; a contradiction to Theorem 2.1. For each call of the WHILE-condition we have to construct $[R, \mathcal{C}]$ and keep track of the connected components $\mathcal{C}_{a,b}$ and $\mathcal{C}_c$. The latter task can be done while constructing $[R, \mathcal{C}]$. Thus, both tasks take $\mathcal{O}(|L_R| + |R|)$ time. Since $|L_R| \leq 3|R|$, we have $\mathcal{O}(|L_R| + |R|) = \mathcal{O}(|R|)$. Thus, we end in an overall time complexity $\mathcal{O}(|R|^2|L_R|)$). ■

Lemma 5.4. *Algorithm 3 computes the closure* $\mathrm{cl}(R)$ *of a consistent triple set R in $\mathcal{O}(|R|^2|L_R|)$ time.*

Proof. Given a consistent triple set R, the set $\mathfrak{L}^\star_R(R)$ is computed. For each $\{A, B\} \in \mathfrak{L}^\star_R(R)$ the respective set $\mathcal{R}_{A,B}$ is constructed and attached to $\mathrm{cl}(R)$. By Theorem 5.1, $\mathrm{cl}(R)$ is correctly computed.

Concerning the time complexity, first observe that Algorithm 2 runs in $\mathcal{O}(|R|^2|L_R|)$ time. Furthermore, Theorem 5.1 implies that $\mathcal{R}_{A,B} \cap \mathcal{R}_{A',B'} = \emptyset$ for distinct $\{A, B\}, \{A', B'\} \in \mathfrak{L}^\star_R(R)$. That is, each triple of $\mathrm{cl}(R)$ is computed exactly once in the entire run of the FOR-loop. Since $\mathrm{cl}(R)$ can have at most $|L_R|^3$ triples, the FOR-loop has a time complexity of $\mathcal{O}(|L_R|^3)$. Thus, we end in an overall time complexity $\mathcal{O}(|R|^2|L_R| + |L_R|^3)$. Since $|L_R| \leq 3|R|$, we have therefore $\mathcal{O}(|R|^2|L_R| + |L_R|^3) = \mathcal{O}(|R|^2|L_R|)$. ■

The overall time complexity to compute the closure for a given triple set R is $\mathcal{O}(|R|^2|L_R|)$. In a worst case, $|R|$ is close to $\binom{L_R}{3} = \mathcal{O}(|L_R|^3)$, in which we end in $\mathcal{O}(|L_R|^7)$ time. In this case, the time complexity of our approach is the same as the complexity $\mathcal{O}(|R||L_R|^4) = \mathcal{O}(|L_R|^7)$ of the method proposed by Bryant and Steel [6]. In a best case, however, we have $\mathcal{O}(|R|) = \mathcal{O}(|L_R|)$ and then we obtain $\mathcal{O}(|R|^2|L_R|) = \mathcal{O}(|L_R|^3)$, while the method of Bryant and Steel has a time complexity of $\mathcal{O}(|R||L_R|^4) = \mathcal{O}(|L_R|^5)$. Thus, although the time complexities are asymptotically the same whenever $|R|$ is close to $\binom{L_R}{3}$, our methods outperforms the approach of Bryant and Steel [6] for moderately sized R. In particular, since $|L_R| \leq 3|R|$, our method has always

time complexity $\mathcal{O}(|R|^2|L_R|) \subseteq \mathcal{O}(|R|^3)$, while the method of Bryant and Steel has complexity $\mathcal{O}(|R||L_R|^4) \subseteq \mathcal{O}(|R|^5)$.

For the sake of time complexity one can apply Algorithm 2 and 3 directly on an arbitrary given set $R' \in \mathrm{MIN}(\mathfrak{sc}(R))$, since $\mathfrak{L}^\star_R(R) = \mathfrak{L}^\star_{R'}(R')$ (cf. Thm. 3.10). Note, in many cases $R' \in \mathrm{MIN}(\mathfrak{sc}(R))$ will have cardinality strictly less than $|R|$. By way of example, consider the set of all rooted triples $\mathcal{R}(T)$ that are displayed by a binary rooted tree T. In this case, $\mathcal{R}(T)$ is closed and defines T and for any $R' \in \min(\mathfrak{sc}(\mathcal{R}(T)))$ we have $|R'| = |L_R| - 2$ [25, 58]. Hence, for any subset $R \subseteq \mathcal{R}(T)$ that contains R' the cardinality can vary between $|L_R| - 2$ and $\binom{L_R}{3} \in \mathcal{O}(|L_R|^3)$. Note, $\mathrm{cl}(\mathcal{R}(T)) = \mathrm{cl}(R') \subseteq \mathrm{cl}(R) \subseteq \mathrm{cl}(\mathcal{R}(T)) = \mathcal{R}(T)$ and thus, $\mathrm{cl}(R') = \mathrm{cl}(R) = \mathcal{R}(T)$ and $R' \in \min(\mathfrak{sc}(R))$. Therefore, for any such set $R \subseteq \mathcal{R}(T)$ there is a minimal representative set R' that can have cardinality significantly less compared to $|R|$, while R' still contains all information of the tree structure T that is also provided by R and $\mathcal{R}(T)$. On the one hand, this strongly reduces the space complexity to store the information that is needed to recover T. On the other hand, the closure can be computed in the latter case in $\mathcal{O}(|R'|^2|L_R|) = \mathcal{O}(|L_R|^3)$ time, whenever R' is given, which improves the time complexity $\mathcal{O}(|R||L_R|^4)$ to compute $\mathrm{cl}(R)$ by a factor of $|R||L_R|$. A similar argument applies to the set of all triples $\mathcal{R}(T)$ that are displayed by a non-binary rooted tree T. In this case, $\mathcal{R}(T)$ identifies T. Still, $\mathcal{R}(T)$ can be close to $\binom{L_R}{3} \in \mathcal{O}(|L_R|^3)$, while for any $R' \in \min(\mathfrak{sc}(\mathcal{R}(T)))$ the cardinality is bounded by $B(T) \in \mathcal{O}(|L_R|^2)$, cf. Cor. 6.5.

6 Further Results

6.1 Sufficient Conditions for Minimum Representative Triple Sets

We provide in this subsection easy verifiable conditions that are quite useful to identify triples in R that must be contained in every representative set of R and to check whether R is already a minimal representative of R. In particular, these conditions helped us to construct many (counter)examples when we wrote this paper. For instance, the sets R'_1 and R'_2 in Figure 3 are easily verified to be minimal representatives by using the following results.

Lemma 6.1. *Let R be a consistent triple set and $\mathtt{ab|c} \in R$. Furthermore, let $C_{a,b}$ (resp. C_c) be the connected component in $[R, L_R]$ that contains a and b (resp. c). Note, $C_{a,b} = C_c$ is possible. Then,*

$$\mathtt{ab|c} \in \bigcap_{R' \in \min(\mathfrak{sc}(R))} R'$$

whenever the following two conditions are satisfied:

(S1) $[R, L_R]$ does not contain cycles.

(S2) If $\mathtt{ab|d} \in R$ with $c \neq d$, then $d \notin C_{a,b} \cup C_c$.

Moreover, it holds that

$$\bigcap_{R' \in \min(\mathfrak{sc}(R))} R' = \bigcap_{R' \in \mathfrak{sc}(R)} R'.$$

Proof. Assume that Conditions S1 and S2 are satisfied, but that there exists a triple set $R' \in \min(\mathfrak{sc}(R))$ such that $\mathtt{ab|c} \notin R'$. Let $\{A, B\} = \ell^\star_{\mathtt{ab|c}}(R')$ and assume w.l.o.g. that $a, b \in A$ and $c \in B$. Lemma 5.2 implies that $A \cup B \subseteq C_{a,b} \cup C_c$. Lemma 2.2 implies that $[R', A \cup B]$ is a subgraph of $[R, L_R]$.

Since $a, b \in A$ there is a path P_{ab} from a to b in $[R', A \cup B]$. Note, this path must be the edge (a, b), otherwise $[R, L_R]$ would contain a cycle. Thus, there must be another triple $\mathtt{ab|d} \in R'$ with $c \neq d$ and $d \in A \cup B \subseteq C_{a,b} \cup C_c$ that supports the edge (a, b) in $[R', A \cup B]$; a contradiction to Condition S2.

To verify the last equation, we set $M = \bigcap_{R' \in \min(\mathfrak{sc}(R))} R'$ and $N = \bigcap_{R' \in \mathfrak{sc}(R)} R'$. Observe first that $\min(\mathfrak{sc}(R)) \subseteq \mathfrak{sc}(R)$ implies that $N \subseteq M$. Now assume that there is a triple $r \in R$ such $r \notin N$. Thus, there is a triple set $R' \in \mathfrak{sc}(R)$ with $r \notin R'$. Therefore, $r \notin R''$ for all $R'' \in \min(\mathfrak{sc}(R'))$. Since R'' is already minimal and $\mathrm{cl}(R'') = \mathrm{cl}(R') = \mathrm{cl}(R)$, we have $r \notin M$. Hence, $M \subseteq N$ and thus, $M = N$. ∎

Corollary 6.2. *Let R be a consistent triple set. Then, $\min(\mathfrak{sc}(R)) = \{R\}$ whenever Condition S1 and S2 in Lemma 6.1 are satisfied for all triples in R.*

Note, the example in Figure 2 shows that the Conditions S1 and S2 are not necessary for minimal representatives.

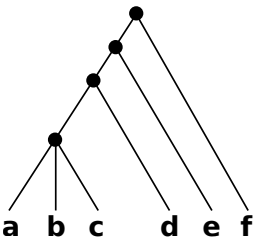

Figure 3: Shown is a tree T on the leaf set $\{a, b, c, d, e, f\}$. Let $R_2 = \mathcal{R}(T)$ and $R_1 = \{\mathtt{ab}|\mathtt{d}, \mathtt{ab}|\mathtt{e}, \mathtt{ab}|\mathtt{f}, \mathtt{bc}|\mathtt{e}, \mathtt{bc}|\mathtt{f}\}$. Although $R_1 \subseteq R_2$, we can observe that for the minimal representative sets $R_1' = R_1 \in \min(\mathfrak{sc}(R_1))$ and $R_2' = \{\mathtt{ab}|\mathtt{d}, \mathtt{bc}|\mathtt{d}, \mathtt{cd}|\mathtt{e}, \mathtt{de}|\mathtt{f}\} \in \min(\mathfrak{sc}(R_2))$ it holds that $|R_1'| > |R_2'|$.

6.2 Triple Sets that Define and Identify a Tree

Here, we are concerned with results established in [58] and [25]. First recall, a triple set R identifies a rooted tree T with leaf set L_R, if T displays R and any other tree that displays R refines T.

In [25] a tight lower bound for the cardinality of triple sets that identify a rooted tree was given. To this end, let $c(v)$ denote the number of children of a vertex v in $T = (V, E)$ and set

$$B(T) = \sum_{(u,v)\in E^0} (c(u) - 1)(c(v) - 1).$$

Theorem 6.3 ([25])**.** *Let R be a consistent triple set. The following properties are satisfied:*

1. $\mathrm{cl}(R) = \mathcal{R}(T)$ *if and only if R identifies T.*
2. *If R identifies T, then $|R| \geq B(T)$.*
3. *For every rooted tree T, there is a triple set R such that R identifies T and $|R| = B(T)$.*

These results allow us to give an exact value and an upper bound for the cardinality of minimal representative triple sets that identify a tree.

Theorem 6.4. *Let T be a tree with maximum degree Δ. Any minimal (and thus, minimum) consistent triple set R that identifies T has cardinality $B(T) \in \mathcal{O}(\Delta \cdot |L_R|) \subseteq \mathcal{O}(|L_R|^2)$.*

Proof. Let R be a minimal consistent triple set R that identifies T. Theorem 6.3(1) implies that $R \in \min(\mathfrak{sc}(\mathcal{R}(T)))$. Theorem 4.8 implies that R has minimum cardinality. Combining Theorem 6.3(2,3) and Theorem 4.8 implies that each $R \in \min(\mathfrak{sc}(\mathcal{R}(T)))$ has cardinality $|R| = B(T)$.

We continue to show that $|B(T)| \in \mathcal{O}(|L_R|^2)$. We will use that $|V^0| \leq |L_R| - 1$, cf. [31, Lemma 1]. Note $\Delta \leq |L_R|$. Moreover, the notation for edges (u, v) is chosen such that u is closer to the root than v.

$$\begin{aligned} B(T) &= \sum_{(u,v)\in E^0} (c(u) - 1)(c(v) - 1) \leq \sum_{(u,v)\in E^0} \deg(u)(c(v) - 1) \leq \Delta \cdot \sum_{v\in V^0} (c(v) - 1) \\ &= \Delta \cdot \Big(- |V^0| + \sum_{v\in V^0} c(v)\Big) = \Delta \cdot (-|V^0| + |E|) \\ &= \Delta \cdot (-|V^0| + |V^0| + |L_R| - 1) = \Delta \cdot (|L_R| - 1) \in \mathcal{O}(|L_R|^2). \end{aligned}$$

■

Corollary 6.5. *Let R be a triple set that identifies the tree T. Then, for any $R' \in \min(\mathfrak{sc}(R))$ we have $|R'| = B(T) \in \mathcal{O}(|L_R|^2)$.*

Proof. Note, $\mathcal{R}(T)$ is closed and therefore, $\mathrm{cl}(\mathcal{R}(T)) = \mathcal{R}(T)$. Theorem 6.3 implies that $\mathrm{cl}(R) = \mathcal{R}(T)$. Now, if $R' \in \min(\mathfrak{sc}(R))$ and $R'' \in \min(\mathfrak{sc}(\mathcal{R}(T)))$, then $\mathrm{cl}(R') = \mathrm{cl}(R'')$. Hence, we can apply Theorem 6.4 and 4.10 to conclude that $|R'| = |R''| = B(T)$. ■

Note, for a rooted binary tree T on L_R and thus, $c(u) = 2$ for each interior vertex, we obtain $B(T) = |L_R| - 2$. Moreover, $B(T) = |L_R| - 2$ shows that $B(T) \in \mathcal{O}(|L_R|)$ is possible. On the other hand, if T is tree for which the root ρ has $c(\rho) = n$ children and each child of ρ is adjacent to exactly two leaves (and hence, $|E^0| = n$ and $|L_R| = 2n$), then we have $B(T) = n(n - 1)$ and therefore, indeed $B(T) \in \Theta(|L_R|^2)$ and thus, $B(T) \notin \mathcal{O}(|L_R|)$ is possible.

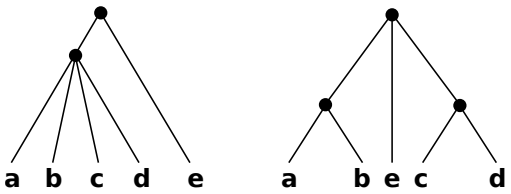

Figure 4: Shown are two trees T_1 (left) and T_2 (right) that display $R = \{\mathtt{ab}|\mathtt{e}, \mathtt{cd}|\mathtt{e}\}$. None of the trees is a refinement of the other one. Hence, R neither identifies T_1 nor T_2. Nevertheless, $\mathbb{C}(R) = \mathcal{C}(T_2)$ and therefore, contains all information to uniquely recover T_2. This example also shows that the converse of Statement (2) in Lemma 6.8 is not satisfied.

We conjecture that for an arbitrary triple set R and $R' \in \min(\mathfrak{sc}(R))$ it always holds that R' is bounded above by $\mathcal{O}(|L_R|^2)$. However, a main difficulty in proving this is the following fact: $R_1 \subseteq R_2$ and $R'_i \in \min(\mathfrak{sc}(R_i))$ with $i = 1, 2$ does not imply that $|R'_1| \leq |R'_2|$; see Figure 3.

Now consider triple sets R that define a rooted tree T with leaf set L_R, that is, T is the unique tree (up to isomorphism) that displays R and thus, T must be binary and $\mathrm{span}(R) = \{T\}$. In [58] it was shown that Theorem 4.8 is always satisfied for triple sets that define a tree.

Theorem 6.6 ([58, Cor. on Page 111])**.** *If R is a triple set that defines the rooted tree $T = (V, E)$ with leaf set L, then $|R'| = |L| - 2$ for any $R' \in \min(\mathfrak{sc}(R))$.*

Note, every triple set R that defines a tree T also identifies T and, by the discussion above and Corollary 6.5, we can conclude that every minimal triple set R that defines T must have cardinality $|R| = B(T) = |L_R| - 2$. Thus, Theorem 6.6 is an immediate consequence of Theorem 4.8 and Corollary 6.5.

A further interesting result is given by Semple [56] and Grünewald et at. [25]:

Lemma 6.7. *For any subset R of $\mathcal{R}(T)$, $\mathrm{cl}(R) = \mathcal{R}(T)$ if and only if R identifies T.*

Furthermore, let A_R denote the unique tree obtained from `BUILD` *with input R. For two sets R_1 and R_2 with $\mathrm{cl}(R_1) = \mathrm{cl}(R_2)$ we have $A_{R_1} = A_{R_2}$. Moreover, if R identifies T, then $A_R = T$.*

The latter result left us with the question if the set $\mathfrak{L}^\star_R(R)$ can be used to obtain similar results. In particular, the question arises under which conditions $\mathfrak{L}^\star_R(R)$ provides the essential information of a hierarchy $\mathcal{C}(T)$ of T. Let

$$\mathbb{C}(R) := \bigcup_{\{A,B\} \in \mathfrak{L}^\star_R(R)} \{A, B\} \cup L_R \cup \Big\{\{x\} \colon x \in L_R\Big\}.$$

There are many examples that show $\mathbb{C}(R) = \mathcal{C}(T)$ for some tree T, in which case $\mathbb{C}(R)$ provides already all information to re-build T. As a simple example consider the set $R_1 = \{\mathtt{ab}|\mathtt{c}\}$ where $\mathcal{C}(T) = \mathbb{C}(R_1) = \{\{a, b\}, \{c\}\} \cup L_{R_1} \cup \{\{a\}, \{b\}, \{c\}\}$ provides the hierarchy of the unique tree $A_{R_1} = \mathtt{ab}|\mathtt{c}$ that displays R_1. A further example is given in Figure 4. Contrary, for $R_2 = \{\mathtt{ab}|\mathtt{d}, \mathtt{bc}|\mathtt{e}\}$ the tree obtained with `BUILD` is $A_{R_2} = ((a, b, c), d, e)$ (given in Newick format). However, the set $\mathbb{C}(R_2)$ does not contain the element $\{a, b, c\}$. Moreover, $\mathbb{C}(R_2)$ contains the elements $\{a, b\}$ and $\{b, c\}$. Hence, $\mathbb{C}(R_2)$ is not a hierarchy. The difference between R_1 and R_2 is simple: R_1 identifies A_{R_1} and R_2 does not identify A_{R_2}. The latter observation leads us to the following result.

Lemma 6.8. *Let R be a consistent triple set and $T \in \mathrm{span}(R)$. Two vertices $u, v \in V(T)$ form a* pair of siblings *$\{u, v\}$, if u and v have a common adjacent vertex w such that $\mathcal{C}(u), \mathcal{C}(v) \subsetneq \mathcal{C}(w)$ and $|\mathcal{C}(u) \cup \mathcal{C}(v)| > 2$, i.e., w is closer to the root than u and v and at least one of u and v is an inner vertex. We denote with $\mathcal{S}(T)$ the set of all such pairs of siblings in T. The following statements are satisfied:*

1. *R identifies T if and only if $\mathfrak{L}^\star_R(R) = \bigcup_{\{u,v\} \in \mathcal{S}(T)} \{\{\mathcal{C}(u), \mathcal{C}(v)\}\}$*
2. *If R identifies T, then $\mathbb{C}(R) = \mathcal{C}(T)$.*

Proof. Let R be a consistent triple set and $T \in \mathrm{span}(R)$. By Lemma 6.7, R identifies T if and only if $\mathrm{cl}(R) = \mathcal{R}(T)$.

We prove first Item (1). Assume that R identifies T. Let $\ell^\star_{\mathtt{ab}|\mathtt{c}}(R) = \{A, B\} \in \mathfrak{L}^\star_R(R)$. W.l.o.g. assume that $a, b \in A$ and $c \in B$. By construction of $\mathfrak{L}^\star_R(R)$ and Theorem 2.6, we have

$\mathtt{ab}|\mathtt{c} \in \mathrm{cl}(R)$. Note, $\mathtt{ab}|\mathtt{c} \in \mathrm{cl}(R) = \mathcal{R}(T)$ if and only if there is a pair of siblings $\{u, v\} \in \mathcal{S}(T)$ such that $a, b \in \mathcal{C}(u)$ and $c \in \mathcal{C}(v)$.

In what follows, we show that $A = \mathcal{C}(u)$ and $B = \mathcal{C}(v)$. Assume for contradiction that $A \not\subseteq \mathcal{C}(u)$. Hence, there is an element $x \in A \setminus \mathcal{C}(u)$. By definition of $\ell^\star_{\mathtt{ab}|\mathtt{c}}(R) = \{A, B\}$ and Theorem 2.6, $\mathtt{ax}|\mathtt{c} \in \mathrm{cl}(R) = \mathcal{R}(T)$. But this immediately implies that $x \in \mathcal{C}(u)$; a contradiction. Hence, $A \subseteq \mathcal{C}(u)$ and, analogously, $B \subseteq \mathcal{C}(v)$. Assume for contradiction that $A \subsetneq \mathcal{C}(u)$. Again, there is an element $x \in \mathcal{C}(u) \setminus A$, which implies that $\mathtt{ax}|\mathtt{c} \in \mathcal{R}(T) = \mathrm{cl}(R)$. Thus, Lemma 3.5 implies that there exists a unique element $\ell^\star_{\mathtt{ax}|\mathtt{c}}(R) = \{A', B'\} \in \mathfrak{L}^\star_R(R)$ such that w.l.o.g. $a, x \in A'$ and $c \in B'$. Thus, $A \neq A'$ as otherwise, $x \in A$. Note, $c \in B \cap B'$. However, Corollary 3.8 implies that $B \cap B'$ must be empty, since $A \cap A' \neq \emptyset$; a contradiction. Therefore, $A = \mathcal{C}(u)$ and, analogously, $B = \mathcal{C}(v)$. In summary, $\mathfrak{L}^\star_R(R) \subseteq \bigcup_{\{u,v\}\in\mathcal{S}(T)} \{\{\mathcal{C}(u), \mathcal{C}(v)\}\}$

Now, let $\{u, v\} \in \mathcal{S}(T)$. Since, $|\mathcal{C}(u) \cup \mathcal{C}(v)| > 2$ we can assume that at least one of $\mathcal{C}(u)$ or $\mathcal{C}(v)$ contains at least two elements, say $\mathcal{C}(u)$. Thus, there are $a, b \in \mathcal{C}(u)$ and $c \in \mathcal{C}(v)$, which implies that $\mathtt{ab}|\mathtt{c} \in \mathcal{R}(T) = \mathrm{cl}(R)$. Lemma 3.5 implies that there exists a unique element $\ell^\star_{\mathtt{ab}|\mathtt{c}}(R) = \{A, B\} \in \mathfrak{L}^\star_R(R)$. Now we can re-use exactly the same arguments as before to show that $\mathcal{C}(u) = A$ and $\mathcal{C}(v) = B$. Thus, $\bigcup_{\{u,v\}\in\mathcal{S}(T)} \{\{\mathcal{C}(u), \mathcal{C}(v)\}\} \subseteq \mathfrak{L}^\star_R(R)$ and therefore, $\bigcup_{\{u,v\}\in\mathcal{S}(T)} \{\{\mathcal{C}(u), \mathcal{C}(v)\}\} = \mathfrak{L}^\star_R(R)$

Conversely, assume that $\mathfrak{L}^\star_R(R) = \bigcup_{\{u,v\}\in\mathcal{S}(T)} \{\{\mathcal{C}(u), \mathcal{C}(v)\}\}$. Let $\mathtt{ab}|\mathtt{c} \in \mathcal{R}(T)$. Again, there must be a pair of siblings $\{u, v\} \in \mathcal{S}(T)$ such that $a, b \in \mathcal{C}(u)$ and $c \in \mathcal{C}(v)$. Hence, $\{\mathcal{C}(u), \mathcal{C}(v)\} \in \mathfrak{L}^\star_R(R)$. Since $\ell^\star_{\mathtt{ab}|\mathtt{c}}(R) \in \mathfrak{L}_{\mathtt{ab}|\mathtt{c}}(R)$, we can apply Lemma 3.3 to conclude that $\mathtt{ab}|\mathtt{c} \in \mathrm{cl}(R)$ and hence, $\mathcal{R}(T) \subseteq \mathrm{cl}(R)$. Moreover, $T \in \mathrm{span}(R)$ implies that $\mathrm{cl}(R) \subseteq \mathcal{R}(T)$ and therefore, $\mathcal{R}(T) = \mathrm{cl}(R)$. Lemma 6.7 implies that R identifies T.

We continue to prove Item (2). Assume that R identifies T. Hence, $\mathfrak{L}^\star_R(R) = \bigcup_{\{u,v\}\in\mathcal{S}(T)} \{\{\mathcal{C}(u), \mathcal{C}(v)\}\}$. Now it is easy to see that $\mathcal{C}^* := \bigcup_{v\in V^0\setminus\{\rho_T\}} \{\mathcal{C}(v)\} = \bigcup_{\{u,v\}\in\mathcal{S}(T)} \{\mathcal{C}(u), \mathcal{C}(v)\}$. Since $L_T = L_R$ we obtain $\mathcal{C}(T) = \mathcal{C}^* \cup L_T \cup \{\{x\} \colon x \in L_T\} = \mathbb{C}(R)$. ■

Note, the converse of Statement (2) in Lemma 6.8 is not satisfied in general, see Figure 4.

6.3 Quartets

Here, we consider unrooted trees in which every inner vertex has degree at least 3. Splits and quartets (unrooted binary trees on four leaves) serve as building blocks for unrooted trees. To be more precise, each edge $e \in E(T)$ of an unrooted tree T gives rise to a split $A|B$, that is, if one removes e from T one obtains two distinct trees T_1 and T_2 with leaf sets $A = L(T_1)$ and $B = L(T_2)$. A tree can be reconstructed in linear time from its set of splits [7, 26, 45]

If there is a split $A|B$ in T such that $a, a' \in A$ and $b, b' \in B$, we say that the quartet $aa'|bb'$ is displayed in T. Equivalently, the quartet $aa'|bb'$ is displayed in T, if $a, a', b, b' \in L(T)$ and the path from a to a' does not intersect the path from b to b' in T. If Q contains all quartets that are displayed in an unrooted tree T, then T is uniquely determined by Q and can be reconstructed in polynomial time [19]. An arbitrary set of quartets Q is called consistent, if there is a tree that displays each quartet in Q, see [57, 60] for further details. Determining whether an arbitrary set of quartets is consistent is an NP-complete problem [58].

Analogously as for rooted triples, we can define the set $\mathrm{span}(Q)$, the closure $\mathrm{cl}(Q)$ and the two sets $\min(\mathfrak{sc}(Q))$ and $\mathrm{MIN}(\mathfrak{sc}(Q))$. Now, consider the ordered pair $(Q, \mathbb{F}_Q)$ where Q is a consistent set of quartets and $\mathbb{F}_Q = \{Q'' \subseteq Q' \colon Q' \in \min(\mathfrak{sc}(Q))\}$. Of course, one might ask whether $(Q, \mathbb{F}_Q)$ is a matroid as well and thus, whether minimal representative sets $Q' \in \min(\mathfrak{sc}(Q))$ have all the same cardinality.

A counterexample, which we recall here for the sake of completeness, is given in [58]: Let $\mathcal{Q}(T)$ be the set of all quartets that are displayed in a binary unrooted tree T with leaf set L. Thus, $\mathrm{span}(\mathcal{Q}(T)) = \{T\}$. Proposition 2(3) in [58] implies that there is a minimal subset $Q' \subseteq \mathcal{Q}(T)$ of size $|Q'| = |L| - 3$ such that $\mathrm{span}(Q') = \{T\}$ and hence, $\mathrm{cl}(\mathcal{Q}(T)) = \mathrm{cl}(Q')$. Thus, for the tree T in Figure 5 there is a minimal representative quartet set of size 4. However, the set $Q' = \{57|24, 15|67, 12|35, 47|13, 34|56\}$ is also a minimal representative quartet set of $Q(T)$, but has size 5. Thus, the basis elements of $(Q, \mathbb{F}_Q)$ don't have the same size, in general. Therefore, $(Q, \mathbb{F}_Q)$ is not a matroid.

7 Conclusion and Outlook

In this contribution, we were concerned with minimum representative triple sets, that is, subsets $R' \subseteq R$ that have minimum cardinality and for which $\mathrm{cl}(R') = \mathrm{cl}(R)$. We have shown that it is

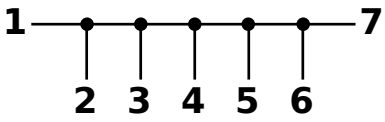

Figure 5: Shown is a binary unrooted tree T with leaf set $L = \{1, 2, \ldots, 7\}$.

possible to compute minimum representative triple sets in polynomial time via a simple greedy approach. To prove the correctness of this method, we showed that minimal representative sets (and its subsets) form a matroid $(R, \mathbb{F}_R)$. Minimal representative sets contain minimum representative sets and since they form the basis of the matroid $(R, \mathbb{F}_R)$, they all must have the same cardinality. The techniques we used to show the matroid structure have been utilized to provide a novel and efficient method to compute the closure $\mathrm{cl}(R)$ of a consistent triple set R. For this algorithm, minimum representative triple sets $R' \in \min(\mathfrak{sc}(R))$ can be used as input, which significantly improves the runtime of the closure computation. Hence, a particular problem that might be addressed in future work is the design of a more efficient algorithm to compute $R' \in \min(\mathfrak{sc}(R))$. Furthermore, the size of $R' \in \min(\mathfrak{sc}(R))$ is not known *a priori*. Boundaries for such sets R' have not been established so-far, except for some rare examples as "defining" or "identifying" triple sets [25, 58]. Thus, in order to understand minimal representative triple sets in more detail, a more thorough analysis of the structure of the matroid $(R, \mathbb{F}_R)$, its collection of bases $\min(\mathfrak{sc}(R))$ or its dual $(R, \mathbb{F}_R)^*$ is needed.

We also assume that the runtime of Algorithm 2 can be improved, which would immediately lead to a faster method to compute $\mathrm{cl}(R)$.

An interesting starting point for future research might be the investigation of the sets $\mathfrak{L}^\star_R(R)$ in more detail and finding a characterization for sets R where $\mathbb{C}(R)$ provides a hierarchy $\mathcal{C}(T)$ of some tree T.

Moreover, generalizations of the established results would be of interest, for instance, is there still a matroid structure if one does not insist that for the subset R' of R we have $\mathrm{cl}(R') = \mathrm{cl}(R)$? What can be said about the structure of representative sets for non-consistent triple sets, see e.g. [25]? Although minimal representative sets of quartets do not provide a matroid structure, it might be useful to figure out which of the other established result are satisfied for quartets as well.

Acknowledgment

We are grateful to Volkmar Liebscher, Mike Steel, Annemarie Luise Kühn and the anonymous referees for their constructive comments and suggestions which has led to a significant improvement of this paper.

References

[1] A. V. Aho, Y. Sagiv, T. G. Szymanski, and J. D. Ullman. Inferring a tree from lowest common ancestors with an application to the optimization of relational expressions. *SIAM J. Comput.*, 10:405–421, 1981.

[2] Federico Ardila. Subdominant matroid ultrametrics. *Annals of Combinatorics*, 8(4):379–389, Jan 2005.

[3] Federico Ardila and Caroline J. Klivans. The bergman complex of a matroid and phylogenetic trees. *Journal of Combinatorial Theory, Series B*, 96(1):38 – 49, 2006.

[4] S. Böcker, D. Bryant, A.W.M. Dress, and M.A. Steel. Algorithmic aspects of tree amalgamation. *Journal of Algorithms*, 37(2):522 – 537, 2000.

[5] D. Bryant. *Building trees, hunting for trees, and comparing trees: theory and methods in phylogenetic analysis.* PhD thesis, University of Canterbury, 1997.

[6] D. Bryant and M. Steel. Extension operations on sets of leaf-labelled trees. *Adv. Appl. Math.*, 16(4):425–453, 1995.

[7] Peter Buneman. A characterisation of rigid circuit graphs. *Discrete Mathematics*, 9(3):205 – 212, 1974.

[8] Jaroslaw Byrka, Sylvain Guillemot, and Jesper Jansson. New results on optimizing rooted triplets consistency. *Discrete Applied Mathematics*, 158(11):1136 – 1147, 2010.

[9] Miklós Csűrös. Fast recovery of evolutionary trees with thousands of nodes. *Journal of Computational Biology*, 9(2):277–297, 2002.

[10] Miklós Csűrös and Ming-Yang Kao. Recovering evolutionary trees through harmonic greedy triplets. In *Proceedings of the Tenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '99, pages 261–270, Philadelphia, PA, USA, 1999. Society for Industrial and Applied Mathematics.

[11] Miklós Csűrös and Ming-Yang Kao. Provably fast and accurate recovery of evolutionary trees through harmonic greedy triplets. *SIAM Journal on Computing*, 31(1):306–322, 2001.

[12] Michael DeGiorgio and James H. Degnan. Fast and consistent estimation of species trees using supermatrix rooted triples. *Molecular Biology and Evolution*, 27(3):552–569, 2010.

[13] M. C. H. Dekker. Reconstruction methods for derivation trees. Master's thesis, Vrije Universiteit, Amsterdam, Netherlands, 1986.

[14] Yu Deng and David Fernández-Baca. Fast Compatibility Testing for Rooted Phylogenetic Trees. In Roberto Grossi and Moshe Lewenstein, editors, *27th Annual Symposium on Combinatorial Pattern Matching (CPM 2016)*, volume 54 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 12:1–12:12, Dagstuhl, Germany, 2016. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

[15] Tobias Dezulian and Mike Steel. Phylogenetic closure operations and homoplasy-free evolution. In David Banks, Frederick R. McMorris, Phipps Arabie, and Wolfgang Gaul, editors, *Classification, Clustering, and Data Mining Applications: Proceedings of the Meeting of the International Federation of Classification Societies (IFCS), Illinois Institute of Technology, Chicago, 15–18 July 2004*, pages 395–416. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.

[16] Riccardo Dondi, Nadia El-Mabrouk, and Manuel Lafond. Correction of weighted orthology and paralogy relations-complexity and algorithmic results. In *International Workshop on Algorithms in Bioinformatics*, pages 121–136. Springer, 2016.

[17] Andreas Dress, Katharina Huber, and Mike Steel. A matroid associated with a phylogenetic tree. *Discrete Mathematics & Theoretical Computer Science*, Vol. 16 no. 2, 2014.

[18] Andreas Dress, Katharina T. Huber, Jacobus Koolen, Vincent Moulton, and Andreas Spillner. *Basic Phylogenetic Combinatorics*. Oxford University Press, 2012.

[19] P. L. Erdős, M. A. Steel, L. Székely, and T. J. Warnow. A few logs suffice to build (almost) all trees (I). *Random Structures & Algorithms*, 14(2):153–184, 1999.

[20] P. L. Erdős, M. A. Steel, L. Székely, and T. J. Warnow. A few logs suffice to build (almost) all trees: Part ii. *Theoretical Computer Science*, 221(1):77 – 118, 1999.

[21] Gregory B. Ewing, Ingo Ebersberger, Heiko A. Schmidt, and Arndt von Haeseler. Rooted triple consensus and anomalous gene trees. *BMC Evolutionary Biology*, 8(1):118, Apr 2008.

[22] Alexandre P. Francisco, Miguel Bugalho, Mário Ramirez, and João A. Carriço. Global optimal eburst analysis of multilocus typing data using a graphic matroid approach. *BMC Bioinformatics*, 10(1):152, May 2009.

[23] M. Geiß, J. Anders, P.F. Stadler, N. Wieseke, and M. Hellmuth. Reconstructing gene trees from fitchs xenology relation. *J. Math. Biology*, 2018. (to appear, arXiv:1711.02152).

[24] Ilan Gronau and Shlomo Moran. On the hardness of inferring phylogenies from triplet-dissimilarities. *Theoretical Computer Science*, 389(1):44 – 55, 2007.

[25] S. Grünewald, M. Steel, and M.S. Swenson. Closure operations in phylogenetics. *Mathematical Biosciences*, 208(2):521 – 537, 2007.

[26] Dan Gusfield. Efficient algorithms for inferring evolutionary trees. *Networks*, 21(1):19–28, 1991.

[27] Dan Gusfield. Haplotyping as perfect phylogeny: Conceptual framework and efficient solutions. In *Proceedings of the Sixth Annual International Conference on Computational Biology*, RECOMB '02, pages 166–175, New York, NY, USA, 2002. ACM.

[28] M. Hellmuth, M. Hernandez-Rosales, K. T. Huber, V. Moulton, P. F. Stadler, and N. Wieseke. Orthology relations, symbolic ultrametrics, and cographs. *J. Math. Biology*, 66(1-2):399–420, 2013.

[29] M. Hellmuth and N. Wieseke. From sequence data including orthologs, paralogs, and xenologs to gene and species trees. In P. Pontarotti, editor, *Evolutionary Biology: Convergent Evolution, Evolution of Complex Traits, Concepts and Methods*, pages 373–392, Cham, 2016. Springer.

[30] Marc Hellmuth. Biologically feasible gene trees, reconciliation maps and informative triples. *Algorithms for Molecular Biology*, 12(1):23, 2017.

[31] Marc Hellmuth, Nicolas Wieseke, Marcus Lechner, Hans-Peter Lenhof, Martin Middendorf, and Peter F. Stadler. Phylogenomics with paralogs. *Proceedings of the National Academy of Sciences*, 112(7):2058–2063, 2015. DOI: 10.1073/pnas.1412770112.

[32] M. Hernandez-Rosales, M. Hellmuth, N. Wieseke, K. T. Huber, and P. F. Moulton, V.and Stadler. From event-labeled gene trees to species trees. *BMC Bioinformatics*, 13(Suppl 19):S6, 2012.

[33] J. Holm, K. de Lichtenberg, and M. Thorup. Poly-logarithmic deterministic fully-dynamic algorithms for connectivity, minimum spanning tree, 2-edge, and biconnectivity. *J. ACM*, 48(4):723–760, 2001.

[34] K.T. Huber, V. Moulton, C. Semple, and M. Steel. Recovering a phylogenetic tree using pairwise closure operations. *Applied mathematics letters*, 18(3):361–366, 2005.

[35] D. H. Huson, T. Dezulian, T. Klopper, and M. A. Steel. Phylogenetic super-networks from partial trees. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 1(4):151–158, 2004.

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

[37] J Jansson, J. H.-K. Ng, K. Sadakane, and W.-K. Sung. Rooted maximum agreement supertrees. *Algorithmica*, 43:293–307, 2005.

[38] Jesper Jansson and Wing-Kin Sung. Inferring a level-1 phylogenetic network from a dense set of rooted triplets. *Theoretical Computer Science*, 363(1):60 – 68, 2006. Computing and Combinatorics.

[39] Sampath K. Kannan, Eugene L. Lawler, and Tandy J. Warnow. Determining the evolutionary tree using experiments. *Journal of Algorithms*, 21(1):26 – 50, 1996.

[40] Bernhard Korte and Jens Vygen. *Combinatorial optimization*, volume 2. Springer, Berlin Heidelberg, 2012.

[41] Anne Kupczok, Heiko A. Schmidt, and Arndt von Haeseler. Accuracy of phylogeny reconstruction methods combining overlapping gene data sets. *Algorithms for Molecular Biology*, 5(1):37, Dec 2010.

[42] Manuel Lafond, Riccardo Dondi, and Nadia El-Mabrouk. The link between orthology relations and gene trees: a correction perspective. *Algorithms for Molecular Biology*, 11(1):1, 2016.

[43] Manuel Lafond and Nadia El-Mabrouk. Orthology and paralogy constraints: satisfiability and consistency. *BMC Genomics*, 15(6):S12, 2014.

[44] Manuel Lafond and Nadia El-Mabrouk. Orthology relation and gene tree correction: complexity results. In *International Workshop on Algorithms in Bioinformatics*, pages 66–79. Springer, 2015.

[45] C. A. Meacham and G. F. Estabrook. Compatibility methods in systematics. *Annual Review of Ecology and Systematics*, 16(1):431–446, 1985.

[46] Christopher A. Meacham. Theoretical and computational considerations of the compatibility of qualitative taxonomic characters. In Joseph Felsenstein, editor, *Numerical Taxonomy*, pages 304–314. Springer Berlin Heidelberg, Berlin, Heidelberg, 1983.

[47] Elchanan Mossel and Mike Steel. A phase transition for a random cluster model on phylogenetic trees. *Mathematical Biosciences*, 187(2):189 – 203, 2004.

[48] Vincent Moulton, Charles Semple, and Mike Steel. Optimizing phylogenetic diversity under constraints. *Journal of Theoretical Biology*, 246(1):186 – 194, 2007.

[49] Vincent Moulton and Andreas Spillner. Phylogenetic diversity and the maximum coverage problem. *Applied Mathematics Letters*, 22(10):1496 – 1499, 2009.

[50] J. Oxley. *Matroid Theory*. Oxford University Press, Oxford, UK, 2011.

[51] Fabio Pardi and Nick Goldman. Species choice for comparative genomics: Being greedy works. *PLOS Genetics*, 1(6):1–1, 12 2005.

[52] Hervé Philippe and Mathieu Blanchette. Overview of the first phylogenomics conference. *BMC Evolutionary Biology*, 7(1):S1, 2007.

[53] Vincent Ranwez, Vincent Berry, Alexis Criscuolo, Pierre-Henri Fabre, Sylvain Guillemot, Celine Scornavacca, and Emmanuel J. P. Douzery. PhySIC: A veto supertree method with desirable properties. *Systematic Biology*, 56(5):798, 2007.

[54] Monika Rauch Henzinger, Valerie King, and Tandy Warnow. Constructing a tree from homeomorphic subtrees, with applications to computational evolutionary biology. *Algorithmica*, 24:1–13, 1999.

[55] C. Scornavacca, V. Berry, and V. Ranwez. Building species trees from larger parts of phylogenomic databases. *Information and Computation*, 209(3):590 – 605, 2011.

[56] Charles Semple. Reconstructing minimal rooted trees. *Discrete Applied Mathematics*, 127(3):489 – 503, 2003.

[57] Charles Semple and Mike Steel. *Phylogenetics*, volume 24 of *Oxford Lecture Series in Mathematics and its Applications*. Oxford University Press, Oxford, UK, 2003.

[58] Michael Steel. The complexity of reconstructing trees from qualitative characters and subtrees. *Journal of Classification*, 9(1):91–116, 1992.

[59] Mike Steel. Phylogenetic diversity and the greedy algorithm. *Systematic Biology*, 54(4):527–529, 2005.

[60] Mike Steel. *Phylogeny: Discrete and Random Processes in Evolution*. CBMS-NSF Regional conference series in Applied Mathematics. SIAM, Philadelphia, USA, 2016.

[61] L. van Iersel, J. Keijsper, S. Kelk, L. Stougie, F. Hagen, and T. Boekhout. Constructing level-2 phylogenetic networks from triplets. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 6(4):667–681, 2009.

[62] Mark Wilkinson, James A. Cotton, and Joseph L. Thorley. The information content of trees and their matrix representations. *Systematic Biology*, 53(6):989, 2004.