

# BINARY NORMAL NETWORKS WITHOUT NEAR RETICULATIONS CAN BE RECONSTRUCTED FROM THEIR ROOTED TRIPLES

ANDREW FRANCIS AND CHARLES SEMPLE

ABSTRACT. Normal networks are an important class of phylogenetic networks that have compelling mathematical properties which align with intuition about inference from genetic data. While tools enabling widespread use of phylogenetic networks in the biological literature are still under mathematical, statistical, and computational development, many such results are being assembled, and in particular for normal phylogenetic networks. For instance, it has been shown that binary normal networks can be reconstructed from the sets of three- and four-leaf rooted phylogenetic trees that they display. It is also known that one can reconstruct particular *subclasses* of normal networks from just the displayed rooted triples. This applies, for instance, to rooted binary phylogenetic trees and to binary level-1 normal networks. In this paper we address the question of how much of the class of binary normal networks can be reconstructed from just the rooted triples that they display. We find that all except those with substructures that we call “near-sibling reticulations” and “near-stack reticulations” can be reconstructed just from their rooted triples. This goes some way to answering the natural question of how much information can be extracted from a set of displayed rooted triples, which are arguably the simplest substructure that one may hope for in a phylogenetic object.

## 1. INTRODUCTION

A significant challenge for the widespread adoption of phylogenetic networks for biological inference is the difficulty in reconstructing them from available data. Such data typically includes alignments of genetic sequence, either at the whole genome level or, more commonly, at the gene level. Such alignments can readily produce small substructures such as a rooted triple. Much of the difficulty lies in combining these substructures in a meaningful way. We can always reconstruct a phylogenetic network that infers, for example, each of the rooted triples in a given collection by simply choosing a network that infers *all* possible rooted triples. But this is neither meaningful nor informative. Thus we need to constrain the phylogenetic network that we reconstruct. One way to do this is to restrict the reconstructed network to a certain class. This is not unusual, as most current reconstruction methods in computational biology restrict to the class of phylogenetic trees. Furthermore, for the purposes of consistency, the substructures that we use for reconstruction need to have the property of being able to determine the phylogenetic networks of interest.

In the last twenty years, a wild assortment of classes of phylogenetic networks have been introduced and studied. Amongst these classes, normal and tree-child networks

---

*Date:* December 16, 2025.

*Key words and phrases.* Phylogenetic networks, normal networks, rooted triples.

The first author is supported by the Australian Research Council, DP260102678. The second author was supported by the New Zealand Marsden Fund.

have arguably been the most prominent. Introduced by Willson [9, 10] and Cardona et al. [4], respectively, these two classes of phylogenetic networks are sufficiently complex to capture reticulate evolution but also sufficiently constrained to be meaningful. They also have the compelling biological property of “visibility”, which allows the present to ‘see’ all past speciation and reticulation events from extant species. However, of these two classes, normal networks are more adaptable to reconstruction [5]. For example, unlike tree-child networks, normal networks are reconstructable from the small subtrees that they infer [7]. The purpose of this paper is to investigate this adaptability in the context of rooted triples, the first step in establishing a general method for taking a given collection of rooted triples and outputting a normal network that is representative of the initial data.

For a binary phylogenetic network  $N$ , let  $R(N)$  denote the set of rooted triples displayed by  $N$ . It is well known that if  $T$  is a rooted binary phylogenetic tree, then  $R(T)$  determines  $T$  (see, for example, [8]). In particular, we have the following theorem.

**Theorem 1.1.** *Let  $T_1$  and  $T_2$  be two rooted binary phylogenetic  $X$ -trees. Then  $R(T_1) = R(T_2)$  if and only if  $T_1 \cong T_2$ .*

While not difficult to prove, Theorem 1.1 underlies many so-called supertree methods. These methods take as input a collection (of not necessarily compatible) rooted phylogenetic trees on overlapping leaf sets and output a single rooted phylogenetic tree that “best” represents the input collection. For an excellent overview of supertree methods, see [1]. Furthermore, Theorem 1.1 justifies the use of rooted triples in the input collection, as the “true” tree (at least theoretically) is recoverable from the rooted triples it displays. More generally than trees, if  $N$  is a binary level-1 normal network, then  $R(N)$  determines  $N$  [6]. However, not surprisingly, if  $N$  is an arbitrary binary normal network, then  $N$  cannot typically be determined by  $R(N)$  alone (see, for example, the counterexample given by Lemma 3.1).

While rooted triples alone are insufficient, it has recently been shown that displayed trees with at most four leaves are nevertheless sufficient for determining an arbitrary binary normal network. Let  $Q(N)$  denote the set of 4-leaf caterpillars, that is, rooted binary phylogenetic trees on 4 leaves with exactly one cherry, displayed by a binary normal network  $N$ . The next theorem is established in [7, Theorem 2].

**Theorem 1.2.** *Let  $N_1$  and  $N_2$  be two binary normal networks on  $X$ . Then  $R(N_1) = R(N_2)$  and  $Q(N_1) = Q(N_2)$  if and only if  $N_1 \cong N_2$ .*

Using an explicit example, it is shown in [7] that the inclusion of 4-leaf caterpillars is necessary in the statement of Theorem 1.2. This begs the question *to what extent* are such caterpillars necessary? In this paper, we consider this question. More particularly, we investigate *which binary normal networks are determined by the set of rooted triples they display?*

Let  $u$  and  $v$  be reticulations of a binary phylogenetic network  $N$ . If there is a tree vertex in  $N$  such that one of its children is a tree vertex, and this tree vertex is a parent of  $u$ , and the other child is  $v$ , then we say that  $u$  and  $v$  are *near-sibling reticulations*. Furthermore, if the unique child of  $u$  is a tree vertex and this tree vertex is a parent of  $v$ , then we say that  $u$  and  $v$  are *near-stack reticulations*. These concepts are illustrated in

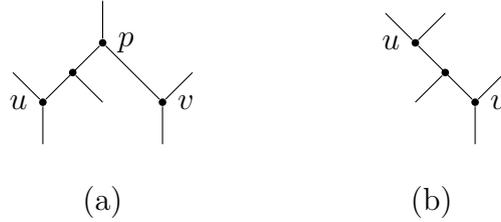


FIGURE 1. The reticulate vertices  $u$  and  $v$  in each figure are near-reticulations. In (a) they are near-sibling reticulations, where the vertex  $p$  is a parent of  $v$  and a grandparent of  $u$ , and in (b) they are near-stack reticulations.

Figure 1. If  $N$  has no near-sibling and no near-stack reticulations, we say  $N$  has no *near reticulations*. Note that binary level-1 normal networks have no near reticulations. The main result of this paper is the following theorem.

**Theorem 1.3.** *Let  $N_1$  and  $N_2$  be two binary normal networks on  $X$  with no near reticulations. Then  $R(N_1) = R(N_2)$  if and only if  $N_1 \cong N_2$ .*

As we show in Section 3, the condition in the statement of Theorem 1.3 that  $N_1$  and  $N_2$  have no near-sibling reticulations is a necessary condition for the theorem to hold. In particular, if  $N$  is a normal network on  $X$  with a pair of near-sibling reticulations, then, unless the reticulations have a certain structural property, there is a normal network  $N'$  on  $X$  not isomorphic to  $N$  such that  $R(N) = R(N')$ . However, it remains open as to whether excluding near-stack reticulations is also necessary in the statement. We discuss this further in the last section. For now, we make the following conjecture.

**Conjecture 1.4.** *Let  $N_1$  and  $N_2$  be two binary normal networks on  $X$  with no near-sibling reticulations. Then  $R(N_1) = R(N_2)$  if and only if  $N_1 \cong N_2$ .*

The paper is organised as follows. The next section contains some necessary preliminaries including formal definitions. Section 3 reviews and establishes some properties of normal networks. In particular, we show that having no near-sibling reticulations in the statement of Theorem 1.3 is a necessary condition. The proof of Theorem 1.3, which follows the model of the proof of Theorem 1.2, relies on being able to recognise cherries and reticulated cherries of a binary normal network  $N$  with no near reticulations using only the set  $R(N)$  of rooted triples displayed by  $N$ . This recognition is established in Section 4, while the proof of Theorem 1.3 is given in Section 5. We end the paper with a discussion in Section 6.

## 2. PRELIMINARIES

Throughout the paper  $X$  denotes a non-empty finite set.

**Phylogenetic networks and trees.** A *binary phylogenetic network*  $N$  on  $X$  is an acyclic directed graph with (i) a single vertex of in-degree 0, which has out-degree 2 and is called the *root*, (ii)  $|X|$  vertices of in-degree 1 and out-degree 0 labelled bijectively by the elements in  $X$ , and (iii) all other vertices having either in-degree 1 and out-degree 2,

called *tree vertices*, or in-degree 2 and out-degree 1, called *reticulations*. The set  $X$  is the *leaf set* of  $N$ , and the elements in  $X$  are called *leaves*. Furthermore, the arcs directed into a reticulation are called *reticulation arcs*. For technical reasons, if  $|X| = 1$ , then the acyclic directed graph consisting of a single vertex labelled by the element in  $X$  is a binary phylogenetic network on  $X$ . For a vertex  $u$  of  $N$ , the *cluster set* of  $u$ , denoted by  $C_u$ , is the subset of  $X$  consisting of the leaves of  $N$  that are descendants of  $u$  (those leaves for which there is a directed path from  $u$ ). A *rooted binary phylogenetic  $X$ -tree* is a phylogenetic network on  $X$  with no reticulations. A *rooted triple* is a binary phylogenetic tree  $T$  on three leaves. If the leaf set of a rooted triple  $T$  is  $\{x, y, z\}$  and  $z$  is adjacent to the root, then we denote  $T$  by  $xy|z$  or, equivalently,  $yx|z$ . Since all phylogenetic networks and phylogenetic trees considered in this paper are rooted and binary, we refer to a rooted binary phylogenetic network and a rooted binary phylogenetic tree as a phylogenetic network and phylogenetic tree, respectively.

Let  $N_1$  and  $N_2$  be two phylogenetic networks on  $X$  with vertex sets  $V_1$  and  $V_2$ , respectively. Then  $N_1$  is *isomorphic* to  $N_2$  if there is bijective map  $\varphi : V_1 \rightarrow V_2$  such that  $\varphi(x) = x$  for all  $x \in X$ , and  $(u, v)$  is an arc of  $N_1$  if and only if  $(\varphi(u), \varphi(v))$  is an arc of  $N_2$ .

**Tree-child and normal networks.** A phylogenetic network  $N$  is *tree-child* if every non-leaf vertex has at least one child that is a tree vertex or a leaf. Equivalently, each vertex  $v$  of  $N$  has a *tree path* to a leaf, that is, a path from  $v$  to a leaf whose non-terminal vertices are tree vertices. A reticulation arc  $(u, v)$  is a *shortcut* if there is a (directed) path from  $u$  to  $v$  that does not traverse  $(u, v)$ . A *normal network* is a tree-child network without shortcuts. Tree-child and normal networks were introduced in [4] and [9, 10], respectively.

Tree-child networks and normal networks have the important property that every vertex is “visible”. A vertex  $u$  of a phylogenetic network on  $X$  is *visible* if there is a leaf  $\ell$  such that every path from the root of  $N$  to  $\ell$  traverses  $u$ , in which case  $\ell$  *verifies the visibility* of  $u$ . The *visibility set* of  $u$ , denote by  $V_u$ , is the subset of  $X$  consisting of those leaves that verify the visibility of  $u$ . Observe that  $V_u \subseteq C_u$  and that  $C_u - V_u$  is not necessarily empty. It turns out that visibility characterises the class of tree-child networks. The next lemma is well known and gives this characterisation as well as the other characterisation that is central to this paper. Its proof is an almost immediate consequence of the definition. A pair of reticulations in a phylogenetic network are *sibling reticulations* if they share a parent and *stack reticulations* if one of the reticulations is a parent of the other reticulation.

**Lemma 2.1.** *Let  $N$  be a phylogenetic network. Then the following statements are equivalent:*

- (i)  $N$  is tree-child.
- (ii) Every vertex of  $N$  is visible.
- (iii)  $N$  has no sibling and no stack reticulations.

Let  $u$  and  $v$  be reticulations of a phylogenetic network  $N$ . Repeating the definitions given in the introduction, if there is a tree vertex of  $N$  whose two children are  $v$  and a tree vertex parent of  $u$ , then  $u$  and  $v$  are *near-sibling reticulations*. Furthermore, if the child of  $u$  is a tree vertex and a parent of  $v$ , then  $u$  and  $v$  are *near-stack reticulations*. A

phylogenetic network with no near-sibling and no near-stack reticulations is said to have no *near reticulations*.

Let  $N$  be a phylogenetic network on  $X$ , and let  $\{a, b\} \subseteq X$ . We say that  $\{a, b\}$  is a *cherry* of  $N$  if  $a$  and  $b$  have the same parent. Furthermore,  $\{a, b\}$  is a *reticulated cherry* if the parent  $p_b$  of  $b$  is a reticulation and the parent  $p_a$  of  $a$  is a parent of  $p_b$ , in which case  $b$  is the *reticulation leaf* of the reticulated cherry. Observe that  $p_a$  is necessarily a tree vertex. The next lemma shows that every tree-child network contains either a cherry or a reticulated cherry [3, Lemma 4.1].

**Lemma 2.2.** *Let  $N$  be a tree-child network on  $X$ . If  $|X| \geq 2$ , then  $N$  contains either a cherry or a reticulated cherry.*

**Displaying and embedding phylogenetic trees.** Let  $N$  be a phylogenetic network on  $X$ , and let  $T$  be a phylogenetic  $X'$ -tree, where  $X' \subseteq X$ . We say  $N$  *displays*  $T$  if there is a subdigraph  $N'$  of  $N$  such that, up to suppressing vertices of in-degree one and out-degree one,  $T$  is isomorphic to  $N'$ , in which case the arc set of  $N'$  is an *embedding* of  $T$  in  $N$ .

In this paper, it is the embedding of rooted triples that is of interest. Let  $N$  be a phylogenetic network, and suppose that  $E_{xyz}$  is an embedding of the rooted triple  $xy|z$  in  $N$ . If  $e$  is an arc of  $N$ , we say that  $E_{xyz}$  *uses*  $e$  if  $e \in E_{xyz}$ . Similarly, if  $u$  is a vertex of  $N$ , we say that  $E_{xyz}$  *uses*  $u$  if  $u$  is an end vertex of an arc in  $E_{xyz}$ . Extending this terminology, let  $f$  be an arc of  $xy|z$ . If the arcs of  $E_{xyz}$  corresponding to  $f$  use an arc  $e$  of  $N$ , we say  $f$  *uses*  $e$  in  $E_{xyz}$ . Similarly, if  $v$  is a vertex of  $N$  and the arcs of  $E_{xyz}$  corresponding to  $f$  use  $v$ , we say  $f$  *uses*  $v$  in  $E_{xyz}$ .

### 3. PROPERTIES OF NORMAL NETWORKS

We begin this section by showing that the condition of having no near-sibling reticulations is a necessary condition in the statement of Theorem 1.3. Let  $N$  be a normal network on  $X$ , and suppose that  $u$  and  $v$  are near-sibling reticulations of  $N$ , where a parent  $t$  of  $v$  is a grandparent of  $u$ , as shown in the left of Figure 2. Let  $s$  denote the parent of  $u$  that is a child of  $t$ , and let  $w$  denote the child of  $s$  that is not  $u$ . Since  $N$  is normal,  $w$  is either a tree vertex or a leaf. Let  $N'$  be the phylogenetic network on  $X$  obtained from  $N$  by deleting  $(s, w)$ , suppressing  $s$ , subdividing the edge  $(t, v)$  with a new vertex  $s'$ , and adjoining  $w$  to this new vertex with a new edge. This operation on  $N$  is illustrated in Figure 2 and we say that  $N'$  has been obtained from  $N$  by a *nearest neighbour interchange relative to  $\{u, v\}$* . Certainly,  $N$  is not isomorphic to  $N'$  and, provided  $(t, u)$  is not a shortcut in  $N'$ , it is easily checked that, as  $N$  is normal,  $N'$  is also normal. Note that if  $(t, u)$  is a shortcut in  $N'$ , then  $v$  is an ancestor of  $u$ , that is,  $u$  and  $v$  are *comparable* in  $N$  (and  $N'$ ). The next lemma shows that we cannot distinguish  $N$  and  $N'$  based only on their sets of rooted triples. The case  $(t, u)$  is a shortcut is discussed in Section 6.

**Lemma 3.1.** *Let  $N$  be a normal network with near-sibling reticulations  $u$  and  $v$ , and let  $N'$  be the phylogenetic network obtained from  $N$  by a nearest-neighbour interchange relative to  $\{u, v\}$ . Then  $R(N) = R(N')$ . In particular, if  $u$  and  $v$  are non-comparable, then  $N'$  is normal and  $R(N) = R(N')$ .*

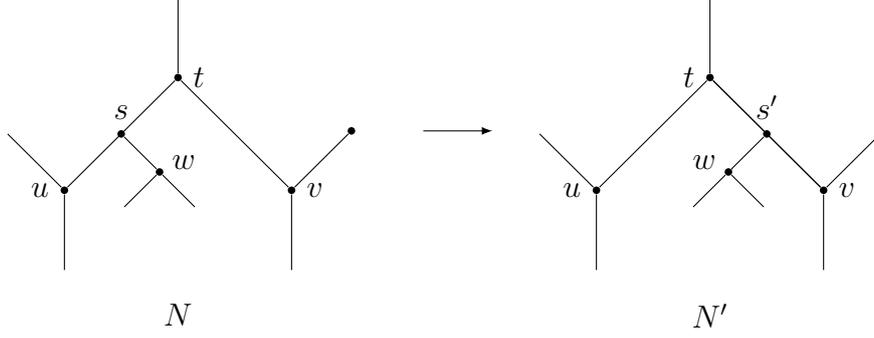


FIGURE 2. Normal networks  $N$  and  $N'$ , where  $N'$  has been obtained from  $N$  by a nearest neighbour interchange relative to  $\{u, v\}$ .

*Proof.* Without loss of generality, we may assume that a parent of  $v$  is a grandparent of  $u$ . In  $N$ , let  $s$  and  $t$  denote parents of  $u$  and  $v$ , respectively, such that  $t$  is a parent of  $s$ . Let  $w$  denote the tree vertex or leaf child of  $s$ , and let  $e = (s, w)$ .

In  $N'$ , let  $t$  and  $s'$  denote the parents of  $u$  and  $v$ , respectively, such that  $t$  is a parent of  $s'$ . Let  $xy|z \in R(N)$  and let  $E_{xyz}$  be an embedding of  $xy|z$  in  $N$ . Using  $E_{xyz}$ , we will show that there is an embedding  $E'_{xyz}$  of  $xy|z$  in  $N'$ . If  $E_{xyz}$  avoids the arc  $e$ , then it is clear that  $xy|z$  is a rooted triple of  $N'$ , so suppose that  $E_{xyz}$  uses  $e$ .

Label the arcs of the rooted triple  $xy|z$  by  $f_x, f_y, f_{xy}$ , and  $f_z$ , where  $f_x, f_y$ , and  $f_z$  are the pendant arcs incident with  $x, y$ , and  $z$ , respectively (see Figure 3). First assume that  $f_z$  uses  $e$ . Say  $f_z$  uses  $t$ . If  $f_{xy}$  avoids  $t$ , then set

$$E'_{xyz} = (E_{xyz} - \{(t, s), (s, w)\}) \cup \{(t, s'), (s', w)\},$$

while if  $f_{xy}$  uses  $t$ , then set

$$E'_{xyz} = (E_{xyz} - \{(t, s), (s, w), (t, v)\}) \cup \{(s', w), (s', v)\}.$$

Otherwise, if  $f_z$  avoids  $t$ , then set

$$E'_{xyz} = (E_{xyz} - \{(s, w), (s, u)\}) \cup \{(t, u), (t, s'), (s', w)\}.$$

Since  $E_{xyz}$  is an embedding of  $xy|z$  in  $N$ , it is easily checked that  $E'_{xyz}$  is an embedding of  $xy|z$  in  $N'$ . Taking a similar approach, if  $f_{xy}$  uses  $e$ , it is straightforward to modify  $E_{xyz}$  to get an embedding of  $xy|z$  in  $N'$ .

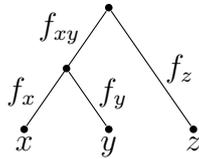


FIGURE 3. Arc labels of the rooted triple  $xy|z$  used in the proof of Lemma 3.1.

Now assume that  $f_x$  uses  $e$ . If  $f_x$  uses  $t$ , then a similar modification to that in the previous paragraph gives an embedding of  $xy|z$  in  $N'$ . Say  $f_x$  avoids  $t$ . Then  $f_{xy}$  uses  $(t, s)$ . If  $f_z$  avoids  $(t, v)$ , then again a modification of  $E_{xyz}$  similar to that in the previous paragraph gives an embedding of  $xy|z$  in  $N'$ . Say  $f_z$  uses  $(t, v)$ . Let  $P_t$  be a path from

the root of  $N'$  to  $t$  and let  $P_v$  be a path from the root of  $N'$  to  $v$  using the reticulation arc directed into  $v$  that is not  $(s', v)$ .

If  $r$  denotes the last vertex common to  $P_t$  and  $P_v$ , let  $E_t$  and  $E_v$  denote the arcs of  $P_t$  and  $P_v$  from  $r$  to  $t$  and from  $r$  to  $v$ , respectively. Set

$$E'_{xyz} = (E_{xyz} - \{(s, u), (s, w), (t, s), (t, v)\}) \cup \{t, u), (s', w), (t, s')\} \cup E_t \cup E_v.$$

Then  $E'_{xyz}$  is an embedding of  $xy|z$  in  $N'$  unless  $v$  is a descendant of either  $u$  or  $w$  in  $N$ . But then  $(t, v)$  is a shortcut in  $N$ , contradicting that  $N$  is normal. By symmetry, if  $f_y$  uses  $e$ , then we can modify  $E_{xyz}$  to get an embedding of  $xy|z$  in  $N'$ . Thus  $xy|z \in R(N')$  and so  $R(N) \subseteq R(N')$ . The same argument but with the roles of  $N$  and  $N'$  interchanged gives  $R(N') \subseteq R(N)$ . This completes the proof of the lemma.  $\square$

The next lemma relates a reticulated cherry of a normal network with the rooted triples it displays.

**Lemma 3.2.** *Let  $N$  be a normal network on  $X$ . Let  $\{a, b\}$  be a reticulated cherry of  $N$  with reticulation leaf  $b$ , and let  $g_b$  be the grandparent of  $b$  that is not the parent of  $a$ . Then each of the following hold:*

- (i) *For all  $x \in X - \{a, b\}$ , we have  $ab|x \in R(N)$ .*
- (ii) *For all  $c \in V_{g_b}$  and  $x \in X - (V_{g_b} \cup \{b\})$ , we have  $bc|x \in R(N)$  and  $ac|b \notin R(N)$ .*
- (iii) *For all distinct  $c, c' \in V_{g_b}$ , we have  $bc|c' \notin R(N)$ .*
- (iv) *For all  $c \in V_{g_b}$  and  $x \in X - (V_{g_b} \cup \{a, b\})$ , we have  $cx|a \in R(N)$  if and only if  $bx|a \in R(N)$ .*
- (v) *If  $bx|y \in R(N)$  and  $ax|y \notin R(N)$ , where  $x, y \in X - (V_{g_b} \cup \{a, b\})$ , then  $cx|y \in R(N)$  for all  $c \in V_{g_b}$ .*
- (vi) *If  $N$  is normal with no near reticulations,  $c \in V_{g_b}$ , and  $ac|x \in R(N)$  for all  $x \in X - (V_{g_b} \cup \{a, b\})$ , then  $ax|c, ax|b \notin R(N)$ .*

*Proof.* Items (i)–(iii) are established in [7]. For parts (iv)–(vi), Figure 4 can be helpful in visualising the arguments below.

Let  $p_a$  and  $p_b$  denote the parents of  $a$  and  $b$  in  $N$ . To prove (iv), let  $c \in V_{g_b}$  and  $x \in X - (V_{g_b} \cup \{a, b\})$ . If  $cx|a \in R(N)$ , then, as  $c$  is in the visibility set of  $g_b$ , an embedding  $E_{cxa}$  of  $cx|a$  in  $N$  uses  $g_b$  and avoids  $p_b$ . Let  $f_c$  denote the arc of  $cx|a$  incident with  $c$ . If  $f_c$  avoids  $g_b$  in  $E_{cxa}$ , in which case  $x \in C_{g_b}$ , set  $E_c$  to be the subset of arcs of  $E_{cxa}$  used by  $f_c$ . On the other hand, if  $f_c$  uses  $g_b$  in  $E_{cxa}$ , set  $E_c$  to be the subset of arcs of  $E_{cxa}$  that induce a path in  $N$  from  $g_b$  to  $c$ . In both instances, it is easily seen that

$$(E_{cxa} - E_c) \cup \{(g_b, p_b), (p_b, b)\}$$

is an embedding of  $bx|a$  in  $N$ . The converse is simpler and omitted, completing the proof of (iv).

For the proof of (v), let  $x, y \in X - (V_{g_b} \cup \{a, b\})$ , and suppose that  $bx|y \in R(N)$  but  $ax|y \notin R(N)$ . Let  $E_{bxy}$  be an embedding of  $bx|y$  in  $N$ . Since  $ax|y \notin R(N)$ , it follows that  $E_{bxy}$  uses  $g_b$  and the arc directed into  $g_b$ . Therefore, if  $c \in V_{g_b}$  and  $P_c$  is a path from  $g_b$  to  $c$ , then

$$(E_{bxy} - \{(g_b, p_b), (p_b, b)\}) \cup E(P_c)$$

is an embedding of  $cx|y$  in  $N$  unless  $P_c$  intersects the path in  $E_{bxy}$  corresponding to the pendant edge of  $bx|y$  incident with  $y$ . But  $y \notin V_{g_b}$ , so this is not possible; otherwise, there is a path from the root of  $N$  to  $c$  avoiding  $g_b$ . This completes the proof of (v).

To prove (vi), suppose that  $N$  has no near reticulations,  $c \in V_{g_b}$ , and  $ac|x \in R(N)$  for all  $x \in X - (V_{g_b} \cup \{a, b\})$ . Let  $g_a$  denote the parent of  $p_a$ , as shown in Figure 4. If  $g_a$  is not an ancestor of  $g_b$  (Figure 4(b)), then  $g_a$  is a tree vertex; otherwise,  $g_a$  and  $p_b$  are near-stack reticulations. If  $g_a$  is the parent of a reticulation, then this reticulation and  $p_b$  are near-sibling reticulations, a contradiction. It now follows that there is a tree path from  $g_a$  to a leaf  $\ell \notin V_{g_b} \cup \{a, b\}$ . But then, if  $c \in V_{g_b}$ , the rooted triple  $ac|\ell \notin R(N)$ , a contradiction.

Now assume that  $g_a$  is an ancestor of  $g_b$  (Figure 4(c)). A similar argument to that in the last paragraph shows that the parent of  $g_b$  is an ancestor of  $a$ . Since  $N$  is acyclic, this implies that  $g_a$  is the parent of  $g_b$ . It now follows that  $ax|c, ax|b \notin R(N)$  for all  $x \in X - (V_{g_b} \cup \{a, b\})$ . This completes the proof of (vi) and the lemma.  $\square$

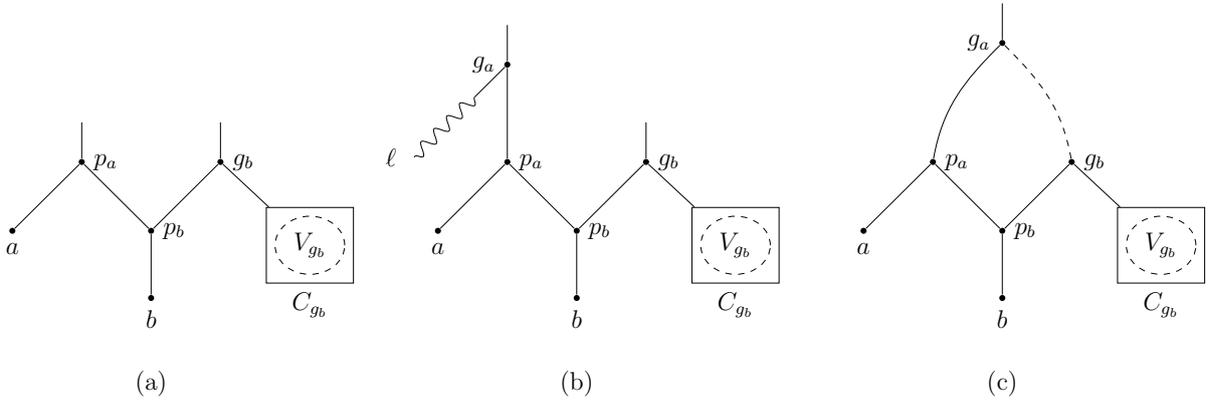


FIGURE 4. (a) shows the generic reticulated cherry scenario for Lemma 3.2, while (b) and (c) illustrate scenarios for the proof of part (vi) of Lemma 3.2. The wavy line in (b) indicates a tree path from  $g_a$  to leaf  $\ell$ , while the dashed line in (c) indicates a (directed) path from  $g_a$  to  $g_b$ .

**Remark.** The scenario in Lemma 3.2(vi) also implies  $cx|a, cx|b \notin R(N)$  as an outcome, but this is not needed for our later arguments.

#### 4. DETECTING CHERRIES AND RETICULATED CHERRIES FROM ROOTED TRIPLES

A key part of proving Theorem 1.3 is to show that if  $N$  is a normal network with no near reticulations, then cherries and reticulated cherries are recognisable using only the rooted triples of  $N$ . The next lemma [7, Lemma 2] shows that *cherries* are recognisable from the set of rooted triples of an arbitrary normal network.

**Lemma 4.1.** *Let  $N$  be a normal network on  $X$  with  $\{a, b\} \subseteq X$ . Then  $\{a, b\}$  is a cherry of  $N$  if and only if, whenever  $xy|z \in R(N)$  and  $a, b \in \{x, y, z\}$ , we have  $\{a, b\} = \{x, y\}$ .*

For the proof of Theorem 1.3, in addition to recognising reticulated cherries  $\{a, b\}$  in a normal network with no near reticulations, where  $b$  is the reticulation leaf, we need to

recognise the visibility set of the grandparent  $g_b$  of  $b$  that is not the parent of  $a$ . The following definition captures sufficient properties of  $V_{g_b}$  for this task.

**Definition 4.2.** Let  $N$  be a normal network with no near reticulations. A *candidate set* for  $b$  is a non-empty subset  $W_b$  of  $X - \{a, b\}$  with the following properties:

- (W1) For all  $c \in W_b$  and  $x \in X - (W_b \cup \{b\})$ , we have  $bc|x \in R(N)$  and  $ac|b \notin R(N)$ .
- (W2) For distinct  $c, c' \in W_b$ , we have  $bc|c' \notin R(N)$ .
- (W3) For all  $c \in W_b$  and  $x \in X - (W_b \cup \{a, b\})$ , we have  $cx|a \in R(N)$  if and only if  $bx|a \in R(N)$ .
- (W4) If  $bx|y \in R(N)$  and  $ax|y \notin R(N)$ , where  $x, y \in X - (W_b \cup \{a, b\})$ , then  $cx|y \in R$  for all  $c \in W_b$ .
- (W5) If  $c \in W_b$  and  $ac|x \in R(N)$  for all  $x \in X - (W_b \cup \{a, b\})$ , then  $ax|c, ax|b \notin R(N)$ .

Note that these five properties for  $W_b$  are also the properties (ii)–(vi) satisfied by the visibility set of  $g_b$  in the context of Lemma 3.2.

**Lemma 4.3.** Let  $N$  be a normal network on  $X$  with no near reticulations. Let  $\{a, b\} \subseteq X$  and let  $W_b$  be a candidate set for  $b$ . Then  $\{a, b\}$  is a reticulated cherry of  $N$  with reticulation leaf  $b$ , and  $V_{g_b} = W_b$ , where  $g_b$  is the grandparent of  $b$  that is not the parent of  $a$ , if and only if  $ab|x \in R(N)$  for all  $x \in X - \{a, b\}$ .

*Proof.* For the forward direction, if  $\{a, b\}$  is a reticulated cherry, then, by Lemma 3.2(i),  $ab|x \in R(N)$  for all  $x \in X - \{a, b\}$ . For the backward direction, suppose that  $ab|x \in R(N)$  for all  $x \in X - \{a, b\}$ . We first show that the parent  $p_b$  of  $b$  is a reticulation.

Suppose to the contrary that  $p_b$  is a tree vertex. Let  $v$  denote the child of  $p_b$  that is not  $b$ . Over the next several sub-lemmas, we deduce consequences of the assumption that  $p_b$  is a tree vertex. The first is that this implies that  $v$  is a reticulation.

#### 4.3.1. $v$ is a reticulation.

If  $v$  is a leaf, then  $v = a$ ; otherwise,  $ab|v \notin R(N)$ , a contradiction. But if  $v = a$ , then,  $\{a, b\}$  is a cherry and we cannot have  $bc|a \in R(N)$ , where  $c \in W_b$ , contradicting (W1) for  $W_b$ . So  $v$  is not a leaf.

Suppose instead that  $v$  is a tree vertex. Since  $N$  is normal, there is a tree path from  $v$  to a leaf, say  $\ell$ . If  $a \neq \ell$ , then either  $a\ell|b \in R(N)$  or  $b\ell|a \in R(N)$  but  $ab|\ell \notin R(N)$ , a contradiction. So  $a = \ell$ . But then, similarly,  $bc|a \notin R(N)$ , where  $c \in W_b$ , contradicting (W1) for  $W_b$ . Therefore  $v$  is a reticulation and (4.3.1) holds.

Let  $p_v$  denote the parent of  $v$  that is not  $p_b$ , and let  $\ell$  and  $m$  be leaves at the end of tree paths starting at  $v$  and  $p_v$ , respectively, as shown in Figure 5(a). Observe that  $\ell \neq m$ . Let  $h_b$  denote the parent of  $p_b$  in  $N$ . Since  $N$  has no near-stack reticulations,  $h_b$  is a tree vertex or the root of  $N$ . Furthermore, as  $N$  has no near-sibling reticulations, the child of  $h_b$  that is not  $p_b$  is either a tree vertex or a leaf. Thus there is a tree path from  $h_b$  to a leaf, say  $z$ , avoiding  $p_b$ . In particular,  $z \neq b$ . We next consider four cases depending on whether  $a \in C_v$  and whether  $W_b \cap C_v$  is non-empty.

#### 4.3.2. Either $a \notin C_v$ or $W_b \cap C_v$ is empty.

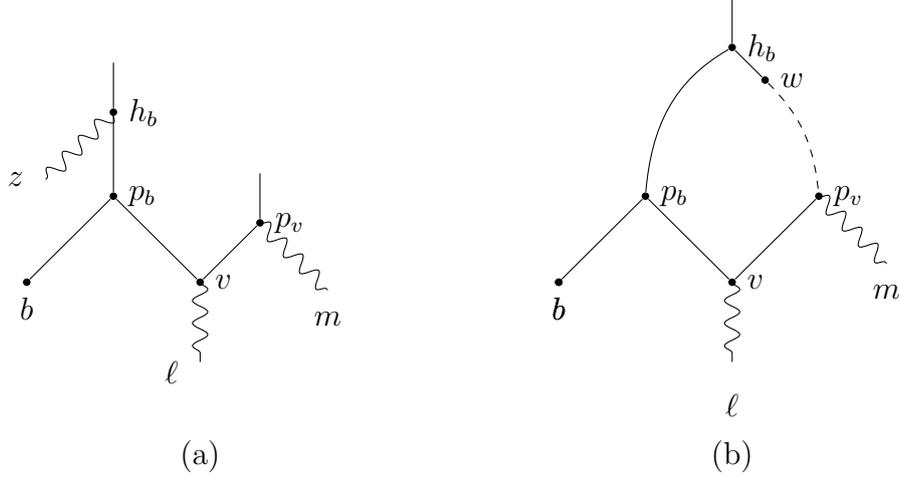


FIGURE 5. Scenarios for (4.3.2), (4.3.3), (4.3.4), and (4.3.5), in which it is assumed that  $p_b$  is a tree vertex. The parent  $h_b$  of  $p_b$  must also be a tree vertex, otherwise  $N$  has a near-stack reticulation. (a) There is a tree path from  $h_b$  to a leaf  $z$ , and from  $p_v$  to a leaf  $m$ . (b) The scenario in which  $h_b$  is an ancestor of  $p_v$ , in which case there is a tree path from  $h_b$  to  $z$  traversing  $w$ .

If  $a \in C_v$  and there exists  $c \in W_b \cap C_v$ , then  $ac|b \in R(N)$ , contradicting (W1) for  $W_b$ . Thus (4.3.2) holds.

**4.3.3.** *Either  $a \in C_v$  or  $W_b \cap C_v$  is non-empty.*

Assume that  $a \notin C$  and  $W_b \cap C_v$  is empty. Say  $h_b$  is not an ancestor of  $p_v$ . Since  $ab|x \in R(N)$  for all  $x \in X - \{a, b\}$ , it follows that  $a \in C_{h_b} - (C_v \cup \{b\})$ . In particular,  $z = a$ ; otherwise,  $ab|z \notin R(N)$ . But then, as  $W_b \cap C_v$  is empty, either  $ac|b \in R(N)$  or  $ab|c \in R(N)$  but  $bc|a \notin R(N)$ , contradicting (W1) for  $W_b$ . Now say  $h_b$  is an ancestor of  $p_v$ , as in Figure 5(b), and note that  $m$  and  $z$  need not be distinct. Again, as  $ab|x \in R(N)$  for all  $x \in X - \{a, b\}$ , we have  $z = a$ . But, as  $W_b \cap C_v$  is empty,  $bc|a \notin R(N)$  for all  $c \in W_b$ , contradicting (W1) for  $W_b$ . Thus (4.3.3) holds.

It now follows that either  $a \notin C_v$  and  $W_b \cap C_v$  is non-empty, or  $a \in C_v$  and  $W_b \cap C_v$  is empty.

**4.3.4.**  *$a \in C_v$  and  $W_b \cap C_v$  is empty.*

Assume that  $a \notin C_v$  and  $W_b \cap C_v$  is non-empty. Note that  $W_b \subseteq C_v$ ; otherwise, there is a  $c' \in W_b$  such that  $bc'|c' \in R(N)$ , contradicting (W2) for  $W_b$ . Say  $h_b$  is not an ancestor of  $p_v$ . If  $a \neq z$ , then, as  $a \notin C_v$ , it follows that  $ab|z \notin R(N)$ . So  $a = z$ . If  $c \in W_b \cap C_v$ , then  $cm|a \in R(N)$ , but  $bm|a \notin R(N)$ , contradicting (W3) for  $W_b$ .

Now say  $h_b$  is an ancestor of  $p_v$ , as in Figure 5(b) and, again, note that  $m$  and  $z$  may not be distinct. Consider  $a, b$ , and  $z$ . If  $a \neq z$ , then  $ab|z \notin R(N)$ , a contradiction, and so  $a = z$ . But then  $ac|b \in R(N)$ , where  $c \in W_b$ , contradicting (W1) for  $W_b$ . Hence (4.3.4) holds.

We now work towards a contradiction that will allow us to conclude that  $p_b$  must be a reticulation. By (4.3.4),  $a \in C_v$  and  $W_b \cap C_v$  is empty. First assume that  $h_b$  is not an ancestor of  $p_v$ . Consider  $a$ ,  $b$ , and  $z$ . If  $z \notin W_b$ , then, for all  $c \in W_b$ , we have  $bc|z \notin R(N)$  as  $W_b \cap C_v$  is empty, contradicting (W1) for  $W_b$ . So  $z \in W_b$ . Also, if there is a  $t \in V_{h_b} - (W_b \cup \{b\})$ , then  $bc|t \notin R(N)$  for all  $c \in W_b$ , a contradiction. So  $V_{h_b} \subseteq W_b \cup \{b\}$ .

We proceed by showing that  $ac|x \in R(N)$  for all  $x \in X - (W_b \cup \{a, b\})$ , where  $c = z$ , and then use (W5) for  $W_b$  to get a contradiction. Let  $y \in X - (W_b \cup \{a, b\})$ . By assumption,  $ab|y \in R(N)$ , and so there is an embedding  $E_{aby}$  of  $ab|y$  in  $N$ . Label the arcs of the rooted triple  $ab|y$  by  $f_a$ ,  $f_b$ ,  $f_{ab}$ , and  $f_y$ , where  $f_a$ ,  $f_b$ , and  $f_y$  are the pendant edges incident with  $a$ ,  $b$  and  $y$ , respectively (along the lines of the arc-labelling in Figure 3). If  $E_{aby}$  uses the arc  $(p_b, v)$ , and so  $f_{ab}$  uses  $(h_b, p_b)$  in  $E_{aby}$ , then it is straightforward to modify  $E_{aby}$  to get an embedding of  $ac|y$  in  $N$  unless  $f_y$  meets  $f_{ab}$  at  $h_b$ . In this case, as  $y \notin V_{h_b}$ , there is a path  $P_y$  from the root of  $N$  to  $y$  avoiding  $h_b$ . Let  $P_c$  denote the tree path from  $h_b$  to  $c$ , and let  $P_a$  denote the path that is the union of  $\{(h_b, p_b)\}$  and the subset of arcs of  $E_{aby}$  corresponding to  $f_a$ . If  $P_y$  and  $P_c$  meet, then  $P_c$  contains a reticulation, a contradiction. Hence if  $P_{h_b}$  is a path from the root of  $N$  to  $h_b$ , it is now easily seen that

$$E(P_a) \cup E(P_c) \cup E(P_{h_b}) \cup E(P_y)$$

contains an embedding of  $ac|y$  in  $N$ , and so  $ac|y \in R(N)$ . If  $E_{aby}$  does not use the arc  $(p_b, v)$ , then  $f_b$  uses  $h_b$  in  $E_{aby}$ . So replace the arcs corresponding to  $f_a$  in  $E_{aby}$  with the arcs in the tree path from  $h_b$  to  $c$  to get an embedding of  $ac|y$  in  $N$ . It now follows that  $ac|x \in R(N)$  for all  $x \in X - (W_b \cup \{a, b\})$ , contradicting (W5) for  $W_b$ .

Now assume that  $h_b$  is an ancestor of  $p_v$ , as in Figure 5(b). If  $z \neq c$ , then, as  $W_b \cap C_v$  is empty,  $bc|z \notin R(N)$ , a contradiction to (W2) for  $W_b$ . So  $z = c$  and, as  $a \in C_v$ , we have  $ac|b \in R(N)$ , contradicting (W1) for  $W_b$ .

We conclude that  $p_b$  is not a tree vertex, and so we have (4.3.5):

**4.3.5.**  $p_b$  is a reticulation.

Let  $p_1$  and  $p_2$  denote the parents of  $p_b$ , and let  $\ell_1$  and  $\ell_2$  denote the leaves at the end of tree paths starting at  $p_1$  and  $p_2$ , respectively, as shown in Figure 6(a). To complete the proof of the lemma, it remains to show that, for  $\{i, j\} = \{1, 2\}$ , the vertex  $p_i$  is the parent of  $a$  and  $V_{p_j} = W_b$ .

**4.3.6.** If  $W_b \cap V_{p_i}$  is non-empty for some  $i \in \{1, 2\}$ , then  $W_b \subseteq V_{p_i}$ .

Say  $W_b \cap V_{p_i}$  is non-empty for some  $i \in \{1, 2\}$ . If there is an element  $c' \in W_p - V_{p_i}$ , then  $bc|c' \in R(N)$ , where  $c \in W_b \cap V_{p_i}$ , contradicting (W2). So (4.3.6) holds.

**4.3.7.** If  $a \in C_{p_i}$  for some  $i \in \{1, 2\}$ , then  $C_{p_i} = \{a, b\}$  and  $W_b = V_{p_j}$ , where  $\{i, j\} = \{1, 2\}$ .

Say  $a \in C_{p_i}$ . If  $\ell_j \notin W_b$ , then, as  $bc|\ell_j \in R(N)$ , an embedding  $E_{bc|\ell_j}$  of  $bc|\ell_j$  in  $N$  must use  $(p_i, p_b)$ , where  $c \in W_b$ . But then we can modify  $E_{bc|\ell_j}$  to get an embedding of  $ac|\ell_j$  that uses  $p_i$  which in turn can be modified to get an embedding of  $ac|b$  in  $N$ , a contradiction to (W1) for  $W_b$ . Thus  $\ell_j \in W_b$ . In particular,  $V_{p_j} \subseteq W_b$ , and so, by (4.3.6),  $V_{p_j} = W_b$ .

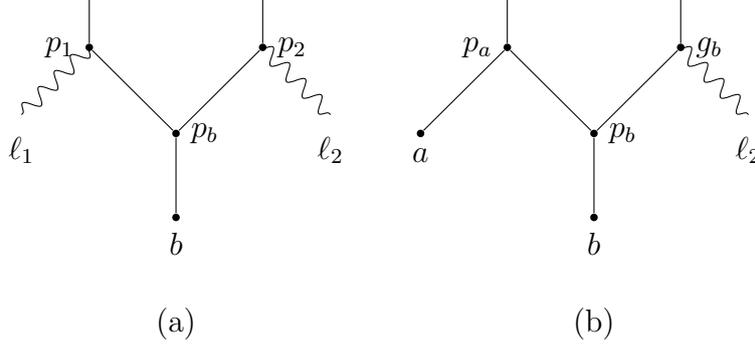


FIGURE 6. Illustration of scenarios for the argument after the parent  $p_b$  of  $b$  has been identified as a reticulation. (a) Once  $b$  is identified as a reticulation, we need to be able to use the rooted triples to show that for some  $i \in \{1, 2\}$ , the vertex  $p_i$  is the parent of  $a$  and (b) to show that, for the other parent of  $p_b$ , its visibility set is  $W_b$ .

Now consider  $C_{p_i}$ . If  $a \notin V_{p_i}$ , then  $a \neq \ell_i$  and an embedding  $E_{ab\ell_i}$  of  $ab|\ell_i$  in  $N$  uses  $(p_j, p_b)$ . But then  $ac|\ell_i \in R(N)$  and, in turn,  $ac|b \in R(N)$ , where  $c \in W_b$ , a contradiction. Thus  $a \in V_{p_i}$ . If  $|V_{p_i}| \geq 2$ , then, for some element  $z \in V_{p_i} - \{a\}$ , we have  $ab|z \notin R(N)$ , a contradiction. So  $|V_{p_i}| = 1$  and  $V_{p_i} = \{a\}$ . If  $|C_{p_i}| \geq 3$ , then the child of  $p_i$  that is not  $p_b$ , say  $q_i$ , is a tree vertex as  $N$  has no sibling reticulations. If there is a tree path from  $q_i$  to a leaf other than  $a$ , then  $|V_{p_i}| \geq 2$ , a contradiction. So  $q_i$  is the parent of a reticulation, and this reticulation and  $p_b$  are near-sibling reticulations, a contradiction. Hence  $C_{p_i} = \{a, b\}$ , and so (4.3.7) holds.

**4.3.8.** *If  $W_b \cap C_{p_i}$  is non-empty for some  $i \in \{1, 2\}$ , then  $W_b = V_{p_i}$  and  $C_{p_j} = \{a, b\}$ , where  $\{i, j\} = \{1, 2\}$ .*

Say  $W_b \cap C_{p_i}$  is non-empty. If  $a \neq \ell_j$ , then an embedding  $E_{ab\ell_j}$  of  $ab|\ell_j$  in  $N$  uses  $(p_i, p_b)$ . But then we can modify  $E_{ab\ell_j}$  to get an embedding of  $ac|\ell_j$  that uses  $p_i$ , which in turn gives an embedding of  $ac|b$  in  $N$ , a contradiction to (W1) for  $W_b$ . So  $a = \ell_j$ . If  $|V_{p_j}| \geq 2$ , then there is an element  $z \in V_{p_j} - \{a\}$  and  $ab|z \notin R(N)$ , a contradiction. So  $|V_{p_j}| = 1$ . If  $C_{p_j} - \{a, b\}$  is non-empty, then the child of  $p_j$  that is not  $p_b$  is a tree vertex and, as  $N$  has no near-sibling reticulations, it follows that there is a tree path from  $p_j$  to a leaf that is not  $a$ , contradicting  $|V_{p_j}| = 1$ . Hence  $C_{p_j} = \{a, b\}$ . By (4.3.7),  $W_b = V_{p_i}$ , and (4.3.8) holds.

We complete the proof of the lemma by showing that, for some  $i \in \{1, 2\}$ , either  $a \in C_{p_i}$  or  $W_b \cap C_{p_i}$  is non-empty and then apply either (4.3.7) or (4.3.8). First assume that  $p_1$  and  $p_2$  have the same parent. If  $a \notin C_{p_1} \cup C_{p_2}$ , then  $ab|\ell_1 \notin R(N)$ , a contradiction. Thus either  $a \in C_{p_1}$  or  $a \in C_{p_2}$ , and the lemma follows by (4.3.7).

We may now assume that  $p_1$  and  $p_2$  do not have the same parent. Let  $h_1$  and  $h_2$  denote the parents of  $p_1$  and  $p_2$ . Since  $N$  has no near-stack reticulations,  $h_1$  and  $h_2$  are both tree vertices and so, as  $N$  is acyclic, either  $h_1$  is not an ancestor of  $p_2$  or  $h_2$  is not an ancestor of  $p_1$ . Without loss of generality, we may assume that  $h_1$  is not an ancestor of  $p_2$ . Since  $N$  has no near-sibling reticulations, the child of  $h_1$  that is not  $p_1$  is a tree vertex or a leaf, and so there is a tree path starting at  $h_1$ , avoiding  $p_1$ , and ending at a leaf  $m_1 \neq \ell_1$ .

If  $m_1 \in W_b \cup \{a\}$ , then either  $ab|_{\ell_1} \notin R(N)$  or  $bc|_{\ell_1} \notin R(N)$  for some  $c \in W_b$ , a contradiction. So  $m_1 \notin W_b \cup \{a\}$ . If  $\ell_1 \notin W_b \cup \{a\}$ , then, as  $b\ell_1|m_1 \in R(N)$  and  $W_b$  is a candidate set for  $b$ , it follows by (W4) for  $W_b$  that either  $a\ell_1|m_1 \in R(N)$  or  $c\ell_1|m_1 \in R(N)$  for all  $c \in W_b$ . This implies that  $(W_b \cup \{a\}) \cap C_{p_1}$  is non-empty and the lemma follows by (4.3.7) and (4.3.8). This completes the proof of Lemma 4.3.  $\square$

## 5. RECONSTRUCTING NORMAL NETWORKS FROM ROOTED TRIPLES

In this section, we prove Theorem 1.3. We begin with a lemma that allows us to use induction for its proof.

**Lemma 5.1.** *Let  $N$  be a normal network on  $X$  with no near reticulations, and let  $\{a, b\}$  be a reticulated cherry of  $N$  with reticulation leaf  $b$ . Let  $N'$  be the phylogenetic network obtained from  $N$  by deleting  $b$  and its parent, and suppressing the resulting two vertices of degree 2. Then  $N'$  is a normal network on  $X - \{b\}$  with no near reticulations.*

*Proof.* It follows by [2, Lemma 3.2] that  $N'$  is normal. To see that  $N'$  has no near reticulations, suppose to the contrary that  $u$  and  $v$  are either near-sibling or near-stack reticulations of  $N'$ . Let  $p_a$  and  $p_b$  denote the parents of  $a$  and  $b$ , respectively, in  $N$  and let  $g_b$  denote the parent of  $p_b$  that is not  $p_a$ . Say  $u$  and  $v$  are near-sibling reticulations of  $N'$ , where the parent  $p_v$  of  $v$  is the parent of a parent  $p_u$  of  $u$ . Then, as  $N$  has no near reticulations, to obtain  $N$  from  $N'$  it follows that  $g_b$  subdivides one  $(p_v, p_u)$ ,  $(p_u, u)$ , and  $(p_v, v)$  in  $N$ . If  $g_b$  subdivides either  $(p_u, u)$  or  $(p_v, v)$ , then either  $u$  and  $p_b$  or  $v$  and  $p_b$  are sibling reticulations in  $N$ , a contradiction. If  $g_b$  subdivides  $(p_v, p_u)$ , then  $v$  and  $p_b$  are near-sibling reticulations in  $N$ , another contradiction. So  $u$  and  $v$  are not near-sibling reticulations. Now say  $u$  and  $v$  are near-stack reticulations of  $N'$ , where the child of  $u$  is a parent  $p_v$  of  $v$ . Since  $N$  has no near reticulations, to obtain  $N$  from  $N'$ , the vertex  $g_b$  subdivides either  $(u, p_v)$  or  $(p_v, v)$  in  $N$ . But then either  $u$  and  $p_b$  are near-stack reticulations or  $v$  and  $p_b$  are sibling reticulations, respectively, in  $N$ . This last contradiction implies that  $N'$  has no near reticulations, thereby completing the proof of the lemma.  $\square$

We now prove Theorem 1.3.

*Proof of Theorem 1.3.* Let  $N_1$  and  $N_2$  be two normal networks on  $X$  with no near reticulations. If  $N_1 \cong N_2$ , then  $R(N_1) = R(N_2)$ . The proof of the converse is by induction on  $|X|$ .

Suppose that  $R(N_1) = R(N_2)$ . If  $|X| \in \{1, 2\}$ , then neither  $N_1$  nor  $N_2$  has any reticulations and so  $N_1 \cong N_2$ , and the theorem holds. Now suppose that  $|X| \geq 3$  and that the theorem holds for all normal networks with no near reticulations on  $|X| - 1$  leaves. By Lemma 2.2,  $N_1$  has either a cherry or a reticulated cherry. If  $N_1$  has a cherry  $\{a, b\}$ , then, by Lemma 4.1,  $\{a, b\}$  is recognised by  $R(N_1)$ . Since  $R(N_1) = R(N_2)$ , Lemma 4.1 also implies that  $\{a, b\}$  is a cherry of  $N_2$ . Let  $N'_1$  and  $N'_2$  denote the normal networks obtained from  $N_1$  and  $N_2$ , respectively, by deleting  $b$  and suppressing the resulting degree-two vertex. Since neither  $N_1$  nor  $N_2$  has a near reticulation, we have from Lemma 5.1 that neither  $N'_1$  nor  $N'_2$  has a near reticulation. Furthermore, as  $R(N_1) = R(N_2)$ , it follows

that  $R(N'_1) = R(N'_2)$ , in particular, for each  $i \in \{1, 2\}$ ,

$$R(N'_i) = R(N_i) - \{bx|y : x, y \in X - \{b\}\}.$$

Therefore, by induction,  $N'_1 \cong N'_2$ . For each  $i \in \{1, 2\}$ , as  $\{a, b\}$  is a cherry of  $N_i$ , the only way  $b$  can be adjoined to  $N'_i$  to get  $N_i$  is to subdivide the arc directed into  $a$  and adjoin  $b$  to this new vertex with a new arc. Thus  $N_1 \cong N_2$ .

Now assume that  $N_1$  has a reticulated cherry  $\{a, b\}$  with reticulation leaf  $b$ . By Lemma 4.3, this reticulated cherry, as well as the visibility set  $V_{g_b}$  of the grandparent  $g_b$  of  $b$  that is not the parent of  $a$ , is recognised by  $R(N_1)$ . Since  $R(N_1) = R(N_2)$ , it follows that  $\{a, b\}$  is also a reticulated cherry of  $N_2$  with reticulation leaf  $b$  and the visibility set of the grandparent of  $b$  that is not the parent of  $a$  is  $V_{g_b}$ . Let  $N'_1$  and  $N'_2$  denote the phylogenetic networks obtained from  $N_1$  and  $N_2$ , respectively, by deleting  $b$  and its parent, and suppressing the two resulting degree-two vertices. By Lemma 5.1,  $N'_1$  and  $N'_2$  are normal networks with no near reticulations. For each  $i \in \{1, 2\}$ , as the rooted triples displayed by  $N'_i$  are exactly the rooted triples displayed by  $N_i$  not having  $b$  as a leaf,  $R(N'_1) = R(N'_2)$ . Therefore, by induction,  $N'_1 \cong N'_2$ .

Let  $p_a$  and  $p_b$  denote the parents of  $a$  and  $b$ , respectively, in  $N_1$ , and recall that  $g_b$  is the parent of  $p_b$  that is not  $p_a$  in  $N_1$ . Let  $e_1 = (p_a, p_b)$  and  $f_1 = (g_b, p_b)$  denote the reticulation arcs in  $N_1$  directed into  $p_b$ . Since  $N'_1 \cong N'_2$ , to show that  $N_1 \cong N_2$ , it suffices to show that there is exactly one arc in  $N'_1$  to adjoin  $e_1$  and exactly one arc in  $N'_1$  to adjoin  $f_1$ . Since, up to suppressing degree-two vertices,  $\{a, b\}$  is a cherry of the normal network obtained from  $N_1$  by deleting  $f_1$ , the only arc to adjoin  $e_1$  to in  $N'_1$  is the arc directed into  $a$ .

We now consider the possible arcs of  $N'_1$  to adjoin  $f_1$  to. Since  $g_b$  is a parent of the reticulation  $p_b$  in  $N_1$ , it follows that the child of  $g_b$  that is not  $p_b$  is a tree vertex or a leaf, and so it has visibility set  $V_{g_b}$ . Thus there is at least one vertex in  $N'_1$  with visibility set  $V_{g_b}$ . Let  $U$  denote the subset of vertices of  $N'_1$  with visibility sets  $V_{g_b}$ . Let  $u \in U$ . Since  $N'_1$  is acyclic, we may assume that  $u$  is chosen so that no ancestor of  $u$  is in  $U$ .

**5.1.1.**  *$|U| \leq 2$  and, if  $U = \{u, u'\}$  and  $u$  is a tree vertex, then  $u'$  is a tree vertex child of  $u$  and the other child of  $u$  is a reticulation.*

If  $|U| = 1$ , then (5.1.1) is proved. Assume that  $|U| \geq 2$ , and let  $u' \in U - \{u\}$ . If  $u'$  is not a descendant of  $u$ , then, by the choice of  $u$ , for each leaf  $\ell \in V_u$ , there is a path in  $N'_1$  from the root to  $\ell$  traversing  $u$  but not  $u'$ , a contradiction. Thus  $u'$  is a descendant of  $u$ . Let  $P = u, v_1, v_2, \dots, v_k, u'$  be a (directed) path from  $u$  to  $u'$  in  $N'_1$ . If there is a tree path from  $u$  to a leaf  $m$  that avoids traversing  $v_1$ , then, by the choice of  $u$ , we have  $m \in V_u$  but  $m \notin V_{u'}$ , a contradiction. Therefore if  $u$  is a tree vertex, it is the parent of a reticulation. In turn, this implies that, regardless of whether  $u$  is a tree vertex or a reticulation,  $v_1$  is not a reticulation as  $N'_1$  is normal. So  $v_1$  is a tree vertex. Let  $w_1$  be the child of  $v_1$  that is not  $v_2$ . If  $w_1$  is a tree vertex, then there is a tree path from  $v_1$  to a leaf  $m_1$  that avoids traversing  $v_2$ , and so  $m_1 \in V_u$  but  $m_1 \notin V_{u'}$ , a contradiction. Thus  $w_1$  is a reticulation. If  $u$  is a reticulation,  $u$  and  $w_1$  are near-stack reticulations, while if  $u$  is a tree vertex,  $w_1$  and the child of  $u$  that is not  $v_1$  are near-sibling reticulations, a contradiction. Hence  $v_1 = u'$ .

It now follows that if  $u$  is a reticulation, then  $|U| = 2$ . Say  $u$  is a tree vertex. Let  $v$  be the child of  $u$  that is not  $u'$  and let  $m$  be a leaf at the end of a tree path starting at  $v$ . If  $v$  is a tree vertex, then  $m \in V_u$  but  $m \notin V_{u'}$ , a contradiction. So  $v$  is a reticulation. As  $N'_1$  has no shortcuts,  $m \notin V_u$  but  $m$  is an element of the visibility set of  $v$ . Thus  $v \notin U$ , so  $|U| = 2$  and (5.1.1) holds.

If  $u$  is a reticulation, then, by (5.1.1),  $U = \{u, u'\}$  and  $f_1$  adjoins to either one of the arcs directed into  $u$ , in which case  $u$  and  $p_b$  are sibling reticulations of  $N_1$ , or the arc  $(u, u')$ , in which case  $u$  and  $p_b$  are near-stack reticulations of  $N_1$ , a contradiction. Thus  $u$  is not a reticulation, and so  $u$  is a tree vertex. If  $|U| = 2$ , then, by (5.1.1),  $f_1$  adjoins to either the arc directed into  $u$  or the arc  $(u, u')$ . In both instances,  $p_b$  and the reticulation child of  $u$  are near-sibling reticulations of  $N_1$ , another contradiction. Hence  $|U| = 1$ , and so there is exactly one arc to adjoin  $f_1$ , namely, the arc directed into  $u$ . We conclude that  $N_1 \cong N_2$ , thereby completing the proof of Theorem 1.3.  $\square$

## 6. DISCUSSION

The capability to reliably reconstruct an explicit phylogenetic network from a set of genetic sequences is a key missing element preventing their widespread use in biological applications. Acquiring this capability involves two logical steps.

The first step is obtaining results that show phylogenetic networks *can* be reconstructed given particular information (and preferably giving a procedure for doing so). The second step is being able to tell when one has the right information in the first place. For instance, the present paper has shown that a certain class of phylogenetic networks can be reconstructed from the sets of rooted triples that they display. But it hasn't give a straightforward method to determine whether a given set of rooted triples might be displayed by a network of that class.

Another way to view these two steps is in terms of mathematical maps. In showing that a class  $\mathcal{N}$  of phylogenetic networks can be reconstructed from the rooted triples it displays, we are effectively showing that the map from  $\mathcal{N}$  to the power set of rooted triples is injective. In doing that, we also show how to reconstruct the pre-image of the set of rooted triples. The second step would be to describe the image of this map, so that we have a bijection between the class of normal networks without near reticulations, and a well-described subset of the power set of the rooted triples.

The second step is crucial for applications: a scientist who gathers a set of genetic sequences does not *a priori* know whether their data evolved on a normal network without near reticulations. Ideally, they would be able to compute the set of rooted triples represented by their data, and be able to tell from those rooted triples whether they have the capability to reconstruct. With the results in this paper, the test for whether a given set of rooted triples is the (complete) set of rooted triples displayed by a normal network,  $N$  say, with no near reticulations, would involve finding a cherry or a set  $W_b$  among the rooted triples, performing a cherry or reticulated cherry reduction, and repeating until it were no longer possible. A more practical test, and one not requiring the complete set of rooted triples displayed by  $N$ , would be better.



## DATA AVAILABILITY STATEMENT

This manuscript has no associated data.

## REFERENCES

- [1] Olaf RP Bininda-Emonds. The evolution of supertrees. *Trends in Ecology & Evolution*, 19(6):315–322, 2004.
- [2] Magnus Bordewich, Katharina T Huber, Vincent Moulton, and Charles Semple. Recovering normal networks from shortest inter-taxa distance information. *Journal of Mathematical Biology*, 77(3):571–594, February 2018.
- [3] Magnus Bordewich and Charles Semple. Determining phylogenetic networks from inter-taxa distances. *Journal of Mathematical Biology*, 73(2):283–303, 2016.
- [4] Gabriel Cardona, Francesc Rosselló, and Gabriel Valiente. Comparison of tree-child phylogenetic networks. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 6(4):552–569, 2009.
- [5] Andrew Francis. “Normal” phylogenetic networks may be emerging as the leading class. *Journal of Theoretical Biology*, page 112236, 2025.
- [6] Philippe Gambette and Katharina T. Huber. On encodings of phylogenetic networks of bounded level. *Journal of Mathematical Biology*, 65(1):157–180, Jul 2012.
- [7] Simone Linz and Charles Semple. Caterpillars on three and four leaves are sufficient to reconstruct binary normal networks. *Journal of Mathematical Biology*, 81:961–980, 2020.
- [8] Charles Semple and Mike Steel. *Phylogenetics*. Oxford University Press, 2003.
- [9] S. J. Willson. Properties of normal phylogenetic networks. *Bulletin of Mathematical Biology*, 72(2):340–358, 2010.
- [10] Stephen J Willson. Reconstruction of certain phylogenetic networks from the genomes at their leaves. *Journal of Theoretical Biology*, 252(2):338–349, 2008.

SCHOOL OF MATHEMATICS AND STATISTICS, UNIVERSITY OF NEW SOUTH WALES, AUSTRALIA

*Email address:* a.francis@unsw.edu.au

SCHOOL OF MATHEMATICS AND STATISTICS, UNIVERSITY OF CANTERBURY, NEW ZEALAND

*Email address:* charles.semple@canterbury.ac.nz