LEAF-RECONSTRUCTIBILITY OF PHYLOGENETIC NETWORKS*

LEO VAN IERSEL † AND VINCENT MOULTON ‡

Abstract. An important problem in evolutionary biology is to reconstruct the evolutionary history of a set X of species. This history is often represented as a phylogenetic network, that is, a connected graph with leaves labelled by elements in X (for example, an evolutionary tree), which is usually also binary, i.e. all vertices have degree 1 or 3. A common approach used in phylogenetics to build a phylogenetic network on X involves constructing it from networks on subsets of X. Here we consider the question of which (unrooted) phylogenetic networks are *leaf-reconstructible*, i.e. which networks can be uniquely reconstructed from the set of networks obtained from it by deleting a single leaf (its X-*deck*). This problem is closely related to the (in)famous reconstruction conjecture in graph theory but, as we shall show, presents distinct challenges. We show that some large classes of phylogenetic networks are reconstructible from their X-deck. This includes phylogenetic trees, binary networks containing at least one non-trivial cut-edge, and binary level-4 networks (the level of a network measures how far it is from being a tree). We also show that for fixed k, almost all binary level-k phylogenetic networks are leaf-reconstructible. As an application of our results, we show that a level-3 network N can be reconstructed from its quarnets, that is, 4-leaved networks that are induced by N in a certain recursive fashion. Our results lead to several interesting open problems which we discuss, including the conjecture that all phylogenetic networks with at least five leaves are leaf-reconstructible.

Key words. phylogenetic trees, phylogenetic networks, graph reconstruction, reconstruction conjecture

AMS subject classifications. 05C60, 92D15

1. Introduction.

An important problem in evolutionary biology is to reconstruct the evolutionary history of a set of species. This commonly involves constructing some form of phylogenetic network, that is, a graph (often a tree) labeled by a set X of species, for which some data (e.g. molecular sequences) has been collected. Over the past four decades several ways have been introduced to construct phylogenetic trees (see e.g. [4]) and, more recently, methods have been developed to construct more general phylogenetic networks (see e.g. [7, 8]).

One particular approach for constructing phylogenetic networks involves building them up from smaller networks. This approach is particularly useful when it is only feasible to compute networks from the biological data on small datasets (e.g. when using likelihood approaches). The problem of building trees from smaller trees has been studied for some time (where it is commonly known as the supertree problem; cf. e.g. [16, Chapter 6]) but the related problem for networks has been only considered more recently (see e.g. [9, 10] focussing on directed phylogenetic networks and [18] focussing on pedigrees). Even so, this problem can be extremely challenging.

*Submitted to the editors 16 January 2017.

Funding: Part of this work was conducted while Vincent Moulton was visiting the TU Delft on a visitors grant funded by the Netherlands Organization for Scientific Research (NWO). Leo van Iersel was partially supported by NWO, including Vidi grant 639.072.602, and partially by the 4TU Applied Mathematics Institute.

†Delft Institute of Applied Mathematics, Delft University of Technology, The Netherlands (l.j.j.v.iersel@gmail.com)

‡School of Computing Sciences, University of East Anglia, Norwich, United Kingdom (v.moulton@uea.ac.uk)

In this paper, we shall present a unified approach to constructing phylogenetic networks from smaller networks. We shall consider *unrooted* phylogenetic networks (cf. [6]). Essentially, these are connected graphs with leaf-set labelled by a set X; they are called *binary* if the degree of every vertex is 1 or 3. For such networks, we focus on the problem of reconstructing a phylogenetic network from its X*-deck*, roughly speaking, this is the collection of networks that is obtained by deleting one leaf and supressing the resulting degree-2 vertex. We call a network that can be reconstructed from its X-deck *leaf-reconstructible*. See Sections 2 and 3 for formal definitions.

Intriguingly, the problem of reconstructing a graph from its vertex deleted subgraphs has been studied for over 75 years (it was introduced in 1941 by Kelly and Ulam [3]), where it is known as the *reconstruction conjecture*. In particular, this conjecture states that every finite simple undirected graph on three of more vertices can be constructed from its collection of vertex deleted subgraphs. This conjecture remains open, but has been shown to hold for several large and important classes of graphs [3]. Even so, as we shall see, although determining leaf-reconstructibilty of a phylogenetic network is closely related to the reconstruction conjecture, there are several key differences which mean that they need to be treated as quite distinct problems.

We now summarize the contents of the rest of the paper. In the next section, we present some preliminaries concerning phylogenetic networks. In Section 3, we then formally define leaf-reconstructibility and explain why this concept is distinct from the notion of end-vertex reconstructibilty a well-studied concept in graph reconstruction theory (see [3, p.237]). (While the notions *end-vertex* and *leaf* have the same meaning, the difference comes from the fact that end-vertex reconstructibility is applied to graphs without leaf-labels, while leaf-reconstructibility is applied to networks where the leaves are labelled.) In addition, we show that certain key features of a binary phylogenetic network (such as its level and reticulation number) can be reconstructed from its X-deck.

In Section 4, we then show that a large class of phylogenetic networks, which we call *decomposable networks* are leaf-reconstructible. These are networks containing at least one cut-edge not incident to a leaf. To show this we first show that any phylogenetic tree with at least 5 leaves is leaf-reconstructible. We also note that phylogenetic trees with 4 leaves are not leaf-reconstructible. Our result concerning decomposable networks is analogous to a result by Yongzhi [21] who showed that the graph reconstruction conjecture can be restricted to considering 2-connected graphs.

The fact that decomposable networks are reconstructible implies that we can restrict our attention to leaf-reconstructibility of *simple* networks, that is, non-decomposable networks. An important feature of a phylogenetic network N is its *level*, which measures how far away the network is from being a phylogenetic tree (in particular, trees are level-0 networks). By considering certain subconfigurations in simple networks, in Section 5, we prove that, for fixed k, almost all binary level-k networks are leaf-reconstructible.

In Section 6, we then turn to the problem of computing the smallest number of elements in the X-deck of a leaf-reconstructible network that are required to reconstruct it, which we call its leaf-reconstruction number. This is analogous to the so-called reconstruction number of a graph (cf. [1] for a survey on these numbers). In particular, we show that the leaf-reconstruction number of any phylogenetic tree on 5 or more leaves is 2, unless it is a star-tree in which case this number is 3. We also show that this implies that the leaf-reconstruction number of any decomposable phylogenetic network with at least 5 leaves is 2.

In Section 7, we turn our attention to low-level networks, showing that all binary level-4 networks with at least five leaves have leaf-reconstruction number at most 2. The proof uses several lemmas that could be useful in studying the leaf-reconstructibility of higher-level networks.

In practice, most methods for constructing phylogenetic networks from smaller networks to date have focussed on using networks with small numbers of leaves (in the rooted case, often 3-leaved networks). In Section 8, by using a recursive argument and our previous results, we show that any level-3 network can be reconstructed from its set of *quarnets*. Essentially, these are 4-leaved networks which are obtained from N by selecting 4 leaves in the network, removing all other leaves and suppressing degree-2 vertices, multi-edges and biconnected components with two incident cut-edges. Our result on quartnets is analogous to results presented in [12] for level-2 rooted phylogenetic networks.

Several variants of the reconstruction conjecture have been considered in the literature (see [3]). We can also consider variants for phylogenetic networks. In Section 9, we consider the problem of reconstructing a phylogenetic network from its collection of edge-deleted subgraphs, showing that in this setting we can sharpen the leaf-reconstructibility bounds that we previously obtained. We then conclude in the last section by discussing the problem of reconstructing directed phylogenetic networks, as well as various open problems.

2. Preliminaries.

In this section, we present some preliminaries concerning phylogenetic networks (cf. [6])

Let X be a finite set with $|X| \geq 2$.

Definition 2.1. *A* phylogenetic tree *on X is a tree with no degree-2 vertices in which the leaves (degree-1 vertices) are bijectively labelled by the elements of X.*

A *biconnected component* of a graph is a maximal 2-connected subgraph and it is called a *blob* if it contains at least two edges.

Definition 2.2. *A* phylogenetic network *on X is a connected graph N such that contracting each blob (one by one) into a single vertex gives a phylogenetic tree on X.*

A bipartition $A|B$ of X, with $A, B \neq \emptyset$ is a *split* of a phylogenetic network N if N contains a cut-edge e such that the elements of A and B are the leaf-labels of the two connected components of $N - e$. If this is the case, we also say that the split $A|B$ is *induced* by e. From the definition of a phylogenetic network it follows that each of its cut-edges induces a split and no two cut-edges induce the same split. Moreover, the phylogenetic tree obtained by contracting each blob of N into a single vertex is the unique phylogenetic tree that has precisely the same splits as N. This phylogenetic tree is denoted $T(N)$, see Figure 1 for an example.

A cut-edge is called *trivial* if at least one of its endpoints is a leaf. A phylogenetic network with at least one nontrivial cut-edge is called *decomposable*. We call a phylogenetic network *simple* if it has precisely one blob.

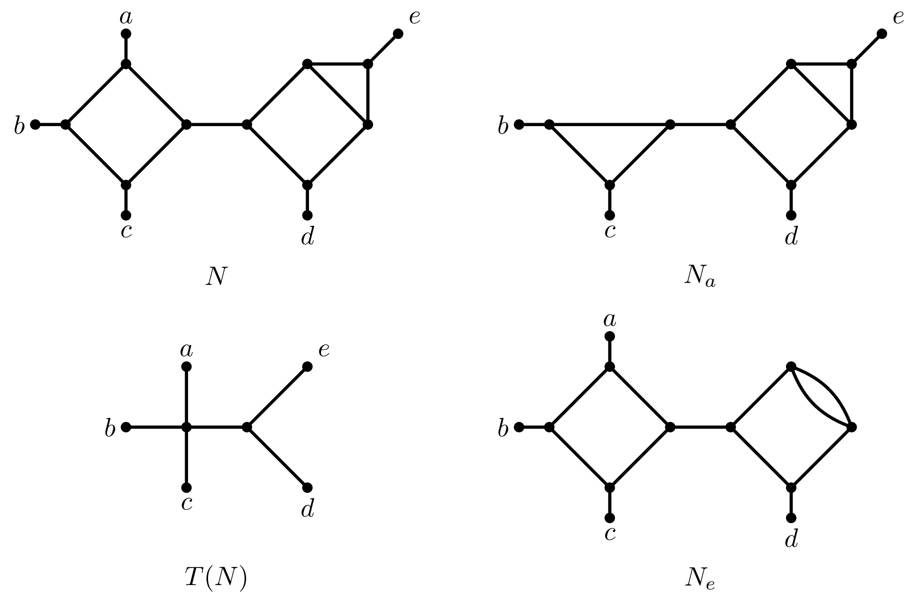

FIG. 1. *A binary phylogenetic network N, the phylogenetic tree $T(N)$, and two elements of the X-deck of N: the phylogenetic network N_a and the pseudo-network N_e.*

DEFINITION 2.3. *A* pseudo-network *on X is a multigraph with no degree-2 vertices in which the leaves (degree-1 vertices) are bijectively labelled by the elements of X.*

Hence, each phylogenetic tree is a phylogenetic network and each phylogenetic network is a pseudo-network. We let $L(N), V(N), E(N)$ denote, respectively, the set of leaves, vertices and edges of a pseudo-network N. In addition, the phylogenetic tree $T(N)$ is defined as the phylogenetic tree obtained by contracting each blob of N into a single vertex and suppressing any resulting degree-2 vertices. Two pseudo-networks N, N' are *equivalent*, denoted $N \sim N'$ if there exists a graph isomorphism between N and N' that is the identity on X.

A pseudo-network is called *binary* if every non-leaf vertex has degree 3. Note that our definition of a binary phylogenetic network is slightly different from the one presented in [6], and has the advantage that for fixed X, there are only finitely many phylogenetic networks with fixed level and leaf-set X (essentially because the number of phylogenetic trees with leaf set X is finite cf. [16]). Note also that a binary phylogenetic network is simple precisely when it is not decomposable and not a star tree. However, this is not the case for nonbinary networks (because then there can be blobs that overlap in a single vertex).

3. X-decks and leaf-reconstructibility.

In this section we introduce the concept of leaf-reconstructibility. We begin by defining the X-deck for a phylogenetic network on X.

Given a phylogenetic network N and a vertex $v \in V(N)$, the pseudo-network N_v is the result of deleting vertex v from N, together with its incident edges, and suppressing resulting degree-2 vertices. See Figure 1 for an example. Given a phylogenetic network N on X and $U \subseteq V(N)$, the *U-deck* of N is the multiset $\{N_u \mid u \in U\}$.

FIG. 2. *A pair of phylogenetic networks that are not leaf-reconstructible (and not even $V(N)$-reconstructible) but that are end-vertex reconstructible (when ignoring the leaf-labels).*

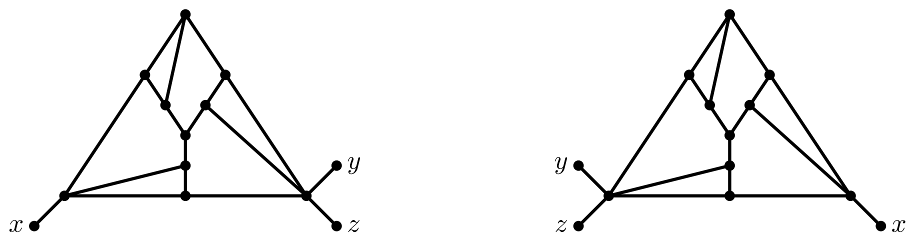

FIG. 3. *A pair of phylogenetic networks that are not end-vertex reconstructible (when ignoring the leaf-lables) but that are leaf-reconstructible.*

A *U-reconstruction* of a network N on X is a network N' on X with $V(N') = V(N)$ and $N'_u \sim N_u$ for all $u \in U$. We call a phylogenetic network N *U-reconstructible* if every U-reconstruction of N is equivalent to N. The *U-reconstruction number* of a network N on X is the smallest k for which there is a subset $U' \subseteq U$ with $|U'| = k$ such that N is U'-reconstructible.

We are usually interested in the case that $U \subseteq X$. For the case that $U = X$, we will also refer to X-reconstruction, X-reconstructible and X-reconstruction number as *leaf-reconstruction*, *leaf-reconstructible* and *leaf-reconstruction number*, respectively. It could also be interesting to take $U = V(N)$, but we shall not consider this possibility in this paper.

If N is a binary network on X and $x \in X$ then N can be obtained from N_x by *attaching* x to some edge e, i.e., to subdivide e by a new vertex v and adding a vertex labelled x and an edge between v and x. For example, the network N in Figure 1 is $\{e\}$-reconstructible since it can be uniquely reconstructed from N_e by attaching leaf e to one of the multi-edges. Hence, this network has leaf-reconstruction number 1. The networks in Figure 2 are not leaf-reconstructible since both networks have the same X-deck.

REMARK 1. *At first sight it might appear that leaf-reconstructibility of a phylogenetic network could be equivalent to end-vertex reconstructibility (where one tries to reconstruct a graph from the deck obtained by deleting only its end-vertices, i.e. leaves, cf. [3, p.237]). However, these are distinct concepts. For example, the phylogenetic networks in Figure 3 are leaf-reconstructible. However, considered as graphs (with no labels), they are not end-vertex reconstructible, as they both have the same end-vertex deck (the multiset of graphs obtained by deleting a single leaf) [15, p.313]. Conversely, the networks in Figure 2 are end-vertex reconstructible but not leaf-reconstructible. Leaf-reconstructibility is also different from reconstructibility, because the latter aims at reconstructing a graph from subgraphs obtained by deleting any vertex (not necessarily a leaf) and without suppressing any resulting degree-2 vertices.*

We call a class $\mathcal{N}$ of phylogenetic networks *leaf-reconstructible* if each $N \in \mathcal{N}$ is leaf-reconstructible. Class $\mathcal{N}$ is *weakly leaf-reconstructible* if, for each network $N \in \mathcal{N}$, all leaf-reconstructions of N that are in $\mathcal{N}$ are equivalent to N. Class $\mathcal{N}$ is *leaf-recognizable* if, for each network $N \in \mathcal{N}$, every leaf-reconstruction of N is also in $\mathcal{N}$.

OBSERVATION 1. *A class $\mathcal{N}$ of phylogenetic networks is leaf-reconstructible if and only if it is leaf-recognizable and weakly leaf-reconstructible.*

We conclude this section by showing that certain features of a binary phylogenetic network on X can be reconstructed from its X-deck. The *reticulation number* of a pseudo-network N is defined as $|E(N)| - |V(N)| + 1$. The *level* of N is the maximum reticulation number of a biconnected component of N. A phylogenetic network is called a *level-k* network, with $k \in \mathbb{N}$, if its level is at most k. A phylogenetic network is called a *simple level-k network* if it is simple and has level exactly k.

A function f defined on a class $\mathcal{N}$ of phylogenetic networks is *leaf-reconstructible* if for each $N \in \mathcal{N}$ and for any leaf-reconstrution M of N we have $f(N) = f(M)$.

PROPOSITION 3.1. *The functions assigning to each binary phylogenetic network its number of edges, number of vertices, reticulation number or level are all leaf-reconstructible.*

Proof. Let N be any phylogenetic network and $x \in L(N)$.

If $|V(N)| = 2$, then $|V(N_x)| = |V(N)| - 1$ and $|E(N_x)| = |E(N)| - 1$. Moreover, the level and reticulation number of N_x are 0, the same as the reticulation number and level of N.

If $|V(N)| \geq 3$, then $|V(N_x)| = |V(N)| - 2$ and $|E(N_x)| = |E(N)| - 2$. Moreover, the level and reticulation number of N_x are the same as the reticulation number and, respectively, level of N.

In both cases, the proposition follows directly. ∎

The following is a direct consequence.

COROLLARY 3.2. *For each $k \in \mathbb{N}$, the class of binary level-k phylogenetic networks is leaf-recognizable.*

4. Decomposable networks. In this section we will consider decomposable networks, that is, networks with at least one nontrivial cut-edge (that is, a cut-edge which does not contain a leaf). We start with a few simple observations. Note that, for $|X| \leq 3$, there exists a unique phylogenetic tree on X which is therefore X-reconstructible. For $|X| = 4$, no binary phylogenetic tree on X is X-reconstructible, but all phylogenetic trees T on X are $V(T)$-reconstructible.

THEOREM 4.1. *Any phylogenetic tree with at least five leaves is leaf-reconstructible.*

Proof. The class of phylogenetic trees is leaf-recognizable by Corollary 3.2. To show weak-reconstructibility, suppose that there exist phylogenetic trees $T \not\sim T'$ on X such that T and T' have the same X-deck. Then there is at least one nontrivial split $A|B$ that is a split of, without loss of generality, T but not of T'. Since $|X| \geq 5$, at least one of A and B contains at least three elements. The other side contains at least two elements since the split is nontrivial. Assume $a_1, a_2, a_3 \in A$ and $b_1, b_2 \in B$. Then T_{a_1} has split $A \setminus \{a_1\}|B$ and T_{a_2} has split $A \setminus \{a_2\}|B$. Hence, T'_{a_1} and T'_{a_2} have the same splits, respectively. This implies that T' has a split that can be obtained from $A \setminus \{a_1\}|B$ by inserting a_1. Since it does not have split $A|B$, it must have split $A \setminus \{a_1\}|B \cup \{a_1\}$. Similarly, T' must have the split $A \setminus \{a_2\}|B \cup \{a_2\}$. This leads to a contradiction because these splits are incompatible (see e.g. [16]). ∎

REMARK 2. *It is known that any tree is reconstructible [14]. A proof of this result is given in [3, p.232], which uses a generalization of Kelly's Lemma [14]. Kelly's Lemma is key to proving several results in graph reconstructibility. We were unable to derive an analogous result for leaf-reconstructibility – it would be interesting to know if some such result exists. Note also that trees are known to be end-vertex reconstructible [11].*

To extend Theorem 4.1 to decomposable networks, we will use the following observation.

OBSERVATION 2. *For any phylogenetic network N on X and any leaf $x \in X$ we have*

$$(T(N))_x = T(N_x)$$

COROLLARY 4.2. *The function mapping a phylogenetic network N with at least five leaves to $T(N)$ is leaf-reconstructible.*

Proof. By Observation 2 and Theorem 4.1. ∎

THEOREM 4.3. *Any decomposable phylogenetic network with at least five leaves is leaf-reconstructible.*

Proof. Let $\mathcal{N}$ be the class of phylogenetic networks with at least five leaves and at least one nontrivial cut-edge. This class is leaf-recognizable since a phylogenetic network on X belongs to this class if and only if every element of its X-deck has at least four leaves and at most two elements of its X-deck have no nontrivial cut-edges.

It remains to show weak leaf-reconstructibility. Suppose $|X| \geq 5$ and let N be a phylogenetic network on X with some nontrivial cut-edge e. Let $A|B$ be the split induced by e. By Corollary 4.2, $T(N)$ is X-reconstructible. Hence, any reconstruction N' of N contains a unique edge e' representing split $A|B$. Since e is nontrivial, there exist leaves $a_1, a_2 \in A$ and $b_1, b_2 \in B$. Pseudo-network N_{a_1} contains a unique edge f inducing split $A \setminus \{a_1\}|B$. Since $N_{a_1} \sim N'_{a_1}$, the connected component of $N_{a_1} - f$ containing B is equivalent to the connected component of $N' - e'$ containing B. Call this connected component N_B and let u be the endpoint of f that it contains. Similarly, pseudo-network N_{b_1} contains a unique edge g inducing split $A|B \setminus \{b_1\}$ and the connected component of $N_{b_1} - g$ containing A is equivalent to the connected component of $N' - e'$ containing A. Call this connected component N_A and let v be the endoint of g that it contains. Then, N' can be obtained from N_A and N_B by adding an edge between u and v. Therefore, $N' \sim N$. ∎

5. Simple networks.

When considering leaf-reconstructability of binary networks we can, by Theorem 4.3, restrict to simple networks, which are binary networks containing precisely one blob. Therefore, in this section we focus on leaf-reconstructibility of simple binary networks. The class of such networks is clearly leaf-recognizable since a phylogenetic network on X is contained in this class if and only if each element of its X-deck is binary and has precisely one blob.

We say that (x, y, z) is a *3-chain* of a phylogenetic network N on X if $x, y, z \in X$ and N contains a path (u, v, w) such that x, y and z are respectively a neighbour of u, v and w.

LEMMA 5.1. *Any simple binary level-k phylogenetic network containing a 3-chain is leaf-reconstructible if it has at least 4 leaves and at least 5 leaves if $k = 1$.*

Proof. The class $\mathcal{N}$ of such networks is leaf-recognizable since a simple binary level-k phylogenetic network on X, with $|X| \geq 4$ and $|X| \geq 5$ if $k = 1$, is contained in $\mathcal{N}$ if and only if at most three elements of its X-deck do not contain a 3-chain.

To show weak leaf-reconstructibility, let $N \in \mathcal{N}$ be a phylogenetic network on X and let (x, y, z) be a 3-chain in N. Since $|X| \geq 4$, there exists at least one other leaf $a \in X$. Consider N_y and N_a. First observe that N_a contains a 3-chain (x, y, z). In N_y, there is a unique edge e between the neighbours of x and z. Moreover, in N_y there is no 3-chain (x, a, z) by the assumption that $|X| \geq 5$ if $k = 1$. Let $N' \in \mathcal{N}$ be a $\{y, a\}$-reconstruction of N. Then N' contains a 3-chain (x, y, z) since N_a contains a 3-chain (x, y, z) and N_y does not contain a 3-chain (x, a, z). Hence, N' can be reconstructed from N_y by attaching y to edge e. Therefore, $N' \sim N$. ∎

COROLLARY 5.2. *Any simple binary level-k phylogenetic network with at least $6k - 5$ leaves and $k \geq 2$ is leaf-reconstructible.*

Proof. Leaf-recognizability is clear. Let N be a simple binary level-k phylogenetic network on X with $k \geq 2$ and $|X| \geq 6k - 5$. Deleting all leaves from N and suppressing all degree-2 vertices gives a 3-regular multigraph G. Since N is simple level-k, $|E(N)| - |V(N)| + 1 = k$ and hence $|E(G)| - |V(G)| + 1 = k$. Combining this with the fact that, since G is 3-regular, $3|V(G)| = 2|E(G)|$ gives that $|E(G)| = 3k - 3$. Suppose that N contains no 3-chain. Then it could have at most two leaves per edge of G, implying that $|X| \leq 6k - 6$. Hence, N contains a 3-chain and is therefore X-reconstructible by Lemma 5.1. ∎

COROLLARY 5.3. *Any binary phylogenetic network $N = (V, E)$ on X with $|X| \geq \max\{6(|E| - |V|) + 1, 5\}$ is leaf-reconstructible.*

Proof. If N contains a nontrivial cut-edge, then apply Theorem 4.3. If it is simple level-1, then apply Lemma 5.1. If it is simple level-k with $k \geq 2$ then $|E| - |V| + 1 = k$ and hence $|X| \geq 6k - 5$ and therefore we can apply Corollary 5.2. ∎

We say that *almost all* phylogenetic networks from a certain class $\mathcal{N}$ are leaf-reconstructible, if the probability that a network drawn uniformly at random out of all networks in $\mathcal{N}$ with n leaves is leaf-reconstructible goes to 1 when n goes to infinity.

COROLLARY 5.4. *For any fixed k, almost all binary level-k phylogenetic networks are leaf-reconstructible.*

Proof. All networks with at least five leaves and some nontrivial cut-edge are leaf-reconstructible by Theorem 4.3. For a simple binary level-k phylogenetic network $N = (V, E)$ on X, with $k \geq 1$ we have (similar to in the proof of Corollary 5.2)

$$|V| = 2k - 2 + 2|X|.$$

Hence, when $|V| \to \infty$ then $|X| \to \infty$. When $|X| \geq \max\{6k - 5, 5\}$ then N is X-reconstructible by Lemma 5.1 and Corollary 5.2. The corollary follows. ∎

6. Reconstruction numbers of decomposable networks. In this section, we shall show that the reconstruction number of a decomposable phylogenetic network with at least five leaves is at most two.

Observation 3. *Let $k \geq 0$. To recognize that a phylogenetic network N is level-k it suffices to check that any element of its X-deck is level-k.*

We start by determining the reconstruction number of binary trees.

The *median* of three leaves $x, y, z \in L(T)$ in a phylogenetic tree T is the unique vertex that lies on each of the paths between all pairs of leaves in $\{x, y, z\}$.

Lemma 6.1. *Any binary phylogenetic tree T with at least five leaves has leaf-reconstruction number 2.*

Proof. The class of phylogenetic trees on X is $\{x\}$-recognizable for any $x \in X$ by Observation 3. No phylogenetic tree on X with $|X| \geq 5$ is $\{x\}$-reconstructible for any $x \in X$ since attaching x to different edges in T_x gives different non-equivalent trees. Hence, the leaf-reconstruction number of such trees is at least 2. It remains to show that it is exactly 2.

Consider a binary phylogenetic tree T on X with $|X| \geq 5$. Take any two leaves $x, y \in X$ such that the distance between them is at least 4. Such leaves exist since $|X| \geq 5$. We will show that T can be uniquely reconstructed from T_x and T_y. First observe that any leaf-reconstruction of T is binary since T_x and T_y are binary and x and y do not have a common neighbour.

Let w be the neighbour of x in T and u, v the other two neighbours of w. Then T_x has an edge $\{u, v\}$.

First assume that neither u nor v is a leaf. Then there exist leaves $a, b \neq y$ such that the path between a and b (in T) contains u but not w and there exist leaves $c, d \neq y$ such the path between c and d (in T) contains v but not w. Then u is the median of a, b, c and v is the median of a, c, d in T. Call in T_x and T_y the median of a, b, c also u and the median of a, c, d also v. Then, in T_y, the neighbour of x is adjacent to u and v. Hence, we can reconstruct T from T_x by attaching x to the edge $\{u, v\}$.

Now assume that u is a leaf. Then there again exist leaves $c, d \neq y$ such that v is on the path between c and d (in T). In this case, v is the median of u, c, d in T. Call the median of u, c, d in T_x and T_y also v. Then, since the neighbour of x in T_y is adjacent to u and v, we can again uniquely reconstruct T from T_x by attaching x to the edge $\{u, v\}$. ◻

We now consider nonbinary trees.

Theorem 6.2. *Any phylogenetic tree with at least five leaves has leaf-reconstruction number 2 unless it is a star, in which case it has leaf-reconstruction number 3.*

Proof. As in the proof of Lemma 6.1, it is clear that, for any $x \in X$, the class of phylogenetic trees on X is $\{x\}$-recognizable and no phylogenetic tree on X is $\{x\}$-reconstructible if $|X| \geq 5$. Consider a phylogenetic tree T on X with $|X| \geq 5$.

First consider the case that T is a star. Then, for any $x, y \in X$, there exists a phylogenetic tree $T' \not\sim T$ on X such that $T'_x \sim T_x$ and $T'_y \sim T_y$ (T' has two internal vertices, leaves x and y are adjacent to one of these internal vertices while all other leaves are adjacent to the other internal vertex). Hence, the X-reconstruction number of T is at least 3. To see that it is exactly 3, note that any phylogenetic tree that is not a star has at most two elements in its X-deck that are stars. Hence, since there exists a unique phylogenetic star tree on X, the reconstruction number of T is 3.

Now consider the case that T contains exactly one nontrivial cut-edge $\{u, v\}$. Take one leaf x adjacent to u and one leaf y adjacent to v. First suppose that u has degree 3. Then v has degree at least 4. Hence, T_x is a star tree and T_y has exactly one nontrivial cut-edge $\{u', v'\}$. Suppose x is adjacent to u'. Then u' is adjacent to exactly one other leaf z. Hence, we can uniquely reconstruct T from T_x by attaching x to the edge incident to z. Now suppose that both u and v have degree at least 3. Then T_x and T_y both have exactly one nontrivial cut-edge. Let z be any leaf adjacent to the neighbour of x in T_y. Then we can uniquely reconstruct T from T_x by adding x with an edge to the neighbour of z.

Finally, assume that T has at least two nontrivial cut-edges. Then there exist two leaves $x, y \in X$ such that the distance between them is at least 4. Let w be the neighbour of x in T and $u, v \neq x$ two other neighbours of w.

If w has degree 3, then we can proceed as in the proof of Lemma 6.1.

Now assume w has degree at least 4. Then it has a neighbour $z \notin \{u, v, x\}$. Then there exist leaves $a, b, c \notin \{x, y\}$ reachable by paths from u, v and z respectively that do not contain w. Therefore, the median of a, b and c in T is w. Hence, we can uniquely reconstruct T from T_x by adding x with an edge to the median of a, b and c. ◻

Corollary 6.3. *Any decomposable phylogenetic network with at least five leaves has leaf-reconstruction number at most 2.*

Proof. Let N be a phylogenetic network that has at least five leaves and at least one nontrivial cut-edge and let x and y be maximum distance apart in $T(N)$. Then any $\{x, y\}$-reconstruction has a nontrivial cut-edge. Moreover, since the distance between x and y in $T(N)$ is at least 3, $T(N)$ is $\{x, y\}$-reconstructable by the proof of Theorem 6.2. Moreover, by the proof of Theorem 4.3, it now follows that N is $\{x, y\}$-reconstructable. ◻

7. Low-level networks.

In this section we show that all binary networks with at least five leaves and level at most 4 are leaf-reconstructible and, moreover, have leaf-reconstruction number at most 2. The proofs are based on the following notions.

Definition 7.1. *A binary level-k* generator*, for $k \geq 2$, is a 2-connected 3-regular multigraph $G = (V, E)$ with $|E| - |V| + 1 = k$. The* underlying generator *of a binary simple level-k network N is the generator obtained from N by deleting all leaves and suppressing resulting degree-2 vertices. For an edge e of G, we say that a leaf x* is on edge *e in N if the neighbour of x is on a path that is suppressed into edge e. If x is on edge e then we also say that e* contains *x and we refer to e as the* x-edge.

See Figure 4 for all binary level-k generators, for $2 \leq k \leq 4$.

We say that two cycles are *similar* if they have the same number of vertices and the same number of vertices that are neighbours of leaves, and hence also the same number of generator vertices (i.e. vertices that are not neighbours of leaves).

The following three lemmas show several special cases of simple level-k networks that are leaf-reconstructible. We will use these lemmas to show that all simple level-4 networks are leaf-reconstructible, if they have at least five leaves.

Lemma 7.2. *Let N be a binary simple level-k network on X, with $k \geq 2$ and $|X| \geq 5$. If N contains a cycle C containing the neighbours of leaves a, b, c and d and either*

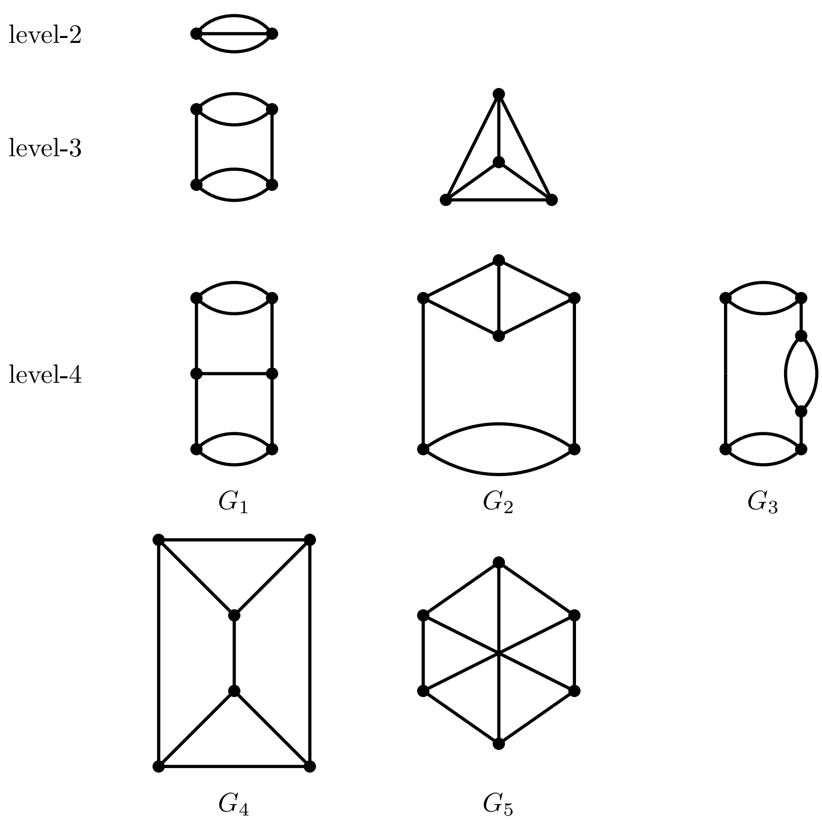

FIG. 4. *All binary level-k generators, for* $2 \leq k \leq 4$.

(i) there is no cycle $C' \neq C$ *in* N *that is similar to* C *and contains the neighbours of* a, b *and* c*; or*

(ii) c *and* d *are on the same edge of the underlying generator and there is no cycle* $C' \neq C$ *in* N *that is similar to* C *and contains the neighbours of* a, b, c *and* d *in a different order,*

then N *is* $\{d, e\}$*-reconstructible, for any* $e \in X \setminus \{a, b, c, d\}$.

Proof. (i) Note that N_e has a cycle C_e containing the neighbours of a, b, c and d and no other cycle that is similar to C_e and contains the neighbours of a, b, c and d. Assume without loss of generality that these neighbours are visited in this order. Suppose that the neighbour of d is the i-th vertex on the path from the neighbour of c to the neighbour of a on C_e. Now consider N_d, which contains a cycle C_d containing the neighbours of a, b and c and no other cycle similar to C_d that contains the neighbours of a, b and c. Let P be the path from the neighbour of c to the neighbour of a on C_d, not via the neighbour of b. If the neighbour of e is among the first i vertices of P then we let f be the i-th edge on P. Otherwise, we let f be the $(i-1)$-th edge on P. Then the unique way to insert d into N_d is by attaching it to edge f.

(ii) Assume without loss of generality that the distance between c and d is 3. Note that N_e has a cycle C_e containing the neighbours of a, b, c and d and no cycle that is similar to C_e and contains the neighbours of a, b, c and d in a different order. Assume again that C_e visits a, b, c and d in this order. Now consider N_d and choose any cycle C_d containing the neighbours of a, b and c. Let f be the first edge on the path from the neighbour of c to the neighbour of a along C_d, not via the neighbour of b. Then the unique way to insert d into N_d is by attaching it to edge f. ◻

LEMMA 7.3. *Let N be a binary simple level-k network on X, with $k \geq 2$ and $|X| \geq 5$. If the underlying generator of N has a pair of multi-edges e_1, e_2 then, unless one of e_1, e_2 contains two leaves and the other one no leaves in N, then N has leaf-reconstruction number at most 2.*

Proof. First suppose that there is exactly one leaf x that is on one of the multi-edges. Then N_x has multi-edges. Since multi-edges are not allowed in phylogenetic networks, the unique way to insert x into N_x is by attaching it to one of the multi-edges.

Now suppose that there is exactly one leaf x on e_1 and exactly one leaf a on e_2. Let y be any other leaf. Then N_y contains a unique 4-cycle containing the neighbours of x and a, and these neighbours are not adjacent. Since N_x contains a unique 3-cycle C containing the neighbour of a, the only way to insert x into N_x is by attaching it to the unique edge on C that is not incident to the neighbour of a.

Now suppose that there are exactly two leaves a, b on e_1 and exactly one leaf x on e_2. Let $y \in X \setminus \{a, b, x\}$. Then, N_y contains a unique 5-cycle containing the neighbours of a, b and x and the neighbour of x is not adjacent to the neighbours of a and b. Since N_x contains a unique 4-cycle C containing the neighbours of a and b, the unique way to insert x into N_x is by attaching it to the unique edge on C that is not incident to the neighbours of a and b.

Now suppose that there are exactly two leaves a, b on e_1 and exactly two leaves c, d on e_2. This case is handled by Lemma 7.2 (i).

The only remaining possibility is that there is a 3-chain, which is handled by the proof of Lemma 5.1. ◻

LEMMA 7.4. *Let N be a binary simple level-k network on X, with $k \geq 2$ and $|X| \geq 5$. If the underlying generator of N has three pairwise incident edges and N has at least three leaves on these edges, then N has leaf-reconstruction number at most 2.*

Proof. First suppose that all three edges are incident to some vertex v and the other three endpoints are all distinct. If each edge contains at least one leaf, let a, b, c be the leaves closest to v on each of the edges. Then N is $\{a, d\}$-reconstructible for any $d \in X \setminus \{a, b, c\}$, since we can reconstruct N from N_a by attaching a to the edge that is incident to the vertex v' that is incident to the b-edge and to the c-edge, making a the leaf closest to v' on that edge. Similarly, if one edge contains at least two leaves a, b and another edge at least one leaf c, then N is again $\{a, d\}$-reconstructible for any $d \in X \setminus \{a, b, c\}$.

A similar argument can be used to handle the case that the three edges form a triangle.

Finally, suppose that at least two of the three edges are multi-edges. Then, by Lemma 7.3, exactly two of the three edges form multi-edges, one of them containing two leaves, the other one no leaves, and the third edge of the three pairwise incident edges contains at least one leaf. Then again it can be seen that N has leaf-reconstruction number at most 2 by using a similar argument as above. ◻

THEOREM 7.5. *Any binary level-4 phylogenetic network with at least five leaves has leaf-reconstruction number at most 2.*

Proof. Let N be such a network. By Corollary 6.3, we may assume that N has no nontrivial cut-edges, i.e. N is simple.

If N is a simple level-1 network, pick any two x, y that are distance at least 4 apart. The fact that N is simple is $\{x, y\}$-recognizable. Moreover, using the fact that N has at least five leaves, it can easily be shown that N can be uniquely reconstructed from N_x and N_y.

Now suppose that N is a simple level-k network, with $k \geq 2$.

If N has a 3-chain (x, y, z) and $a \in X \setminus \{x, y, z\}$, then any $\{y, a\}$-reconstruction of N is simple. Moreover, by the proof of Lemma 5.1 it can be concluded that N is $\{y, a\}$-reconstructible. Hence, we may assume that N contains no 3-chains.

If $k = 2$, then, considering the unique level-2 generator in Figure 4, we are done by Lemma 7.3.

If $k = 3$, then there are two possible underlying generators, see Figure 4. First suppose the underlying generator G is not K_4 and thus has two pairs of multi-edges. Then, by Lemma 7.3, we may assume that each pair of multi-edges has one edge containing exactly two leaves. Hence, we are done by Lemma 7.2 (i). Now suppose that $G = K_4$. Since $|X| \geq 5$, it is straightforward to check that at least one 3-cycle C of G contains at least three leaves in N. By Lemma 7.2, it contains exactly 3 leaves. There are two cases (by Lemma 5.1). Either each edge of C contains exactly one leaf, or one edge contains two leaves and one edge one leaf. In either case, it is easy to check that wherever the other two leaves are, we can apply Lemma 7.2 to see that N has reconstruction number at most 2.

Finally, suppose $k = 4$. Then there are five possibilities for the underlying generator G, see Figure 4. If $G \in \{G_1, G_2, G_3\}$ then, by Lemma 7.3, each pair of multi-edges has one edge containing exactly two leaves and one edge containing no leaves. If $G = G_1$ or G_3, then we are done by Lemma 7.2 (i). If $G = G_2$, then it is straightforward to check that, since $|X| \geq 5$, there must exist some cycle that satisfies the condition of Lemma 7.2 (ii).

Now suppose that $G = G_4$. Observe that G_4 consists of two disjoint 3-cycles and three other edges, which we will call the *middle edges*. For every vertex of G_4, at most two edges incident to this vertex contain leaves by Lemma 7.4. Since $|X| \geq 5$, it is straightforward to check that there is at least one vertex v of G_4 with exactly two leaves a, b on the edges incident to v.

First assume that a is on a middle edge and b is on a triangle edge. Then there is a unique Hamiltonian cycle C of G containing the a-edge and the b-edge. First suppose that there is at least one leaf $c \in X \setminus \{a, b\}$ on an edge of C. Assume that c is the first such leaf on the path along C between the neighbour of b and the neighbour of a not containing v. Let i be the distance from the neighbour of b to the neighbour of c on this path. Let $d \in X \setminus \{a, b, c\}$. Then N is $\{c, d\}$-reconstructible, since the unique way to insert c into N_c is by attaching it to the i-th edge of the path along C from the neighbour of b to the neighbour of a not containing v. Now suppose that none of the leaves in $X \setminus \{a, b\}$ are on edges of C. By Lemma 7.4 there are no leaves on the third edge incident to v. Hence, since $|X| \geq 5$, there at least three leaves on the two edges of G that are not on C and not incident to v. It is now straightforward to check that N has reconstruction number 2 by Lemma 7.2 (i).

Now assume that a and b are both on the same triangle-edge. Then, if the previous case is not applicable for any vertex v' of G_4, the only remaining possibility is that the other triangle also has an edge containg two leaves and we can apply Lemma 7.2.

Now assume that a and b are on different triangle edges (of the same triangle). Then, if the previous cases are not applicable, all other leaves must be on the other triangle and we can use Lemma 7.4.

Finally, assume that a and b are both on the same middle edge. Then, if the previous cases are not applicable, the only remaining possibility is that some other middle edge also contains two leaves and we can apply Lemma 7.2.

Now consider the last level-4 generator $G_5 = K_{3,3}$. As before, it is straightforward to check that there is at least one vertex v of G_5 with exactly two leaves a, b on the edges incident to v.

First suppose that a and b are on different edges incident to v. Observe that there are precisely two Hamiltonian cycles C and D of G_5 containing the a-edge and the b-edge. Since each leaf is on an edge of at least one of C and D, at least one edge of C and D contains a third leaf $c \in X \setminus \{a, b\}$. Suppose that c is on an edge of C. First suppose that all leaves are on edges of C. Then we can use a similar argument as for the Hamiltonian cycle in G_4 to show that N is $\{c, d\}$-reconstructible, for some $d \in X \setminus \{a, b, c\}$. If at least one leaf $e \in X \setminus \{a, b, c\}$ is on an edge that is not also on D, then we choose the Hamiltonian cycle containing the e-edge, and choose $d \neq e$. Otherwise, all leaves are also on edges of D. Observet that there are precisely four edges that are on both C and D, which are two pairs of incident edges. Since $|X| \geq 5$, it then follows by Lemma 7.4 that N has leaf-reconstruction number 2. Now suppose that at least one leaf $e \in X \setminus \{a, b, c\}$ is not on an edge of C. Then N is $\{c, d\}$-reconstructible, with $d \in X \setminus \{a, b, c, e\}$, again using a similar argument as for the Hamiltonian cycle in G_4, choosing the Hamiltonian cycle of G not containing the e-edge.

Finally, suppose that a and b are on the same edge incident to v. Then, if the previous case is not applicable for any vertex v' of G_5, the only remaining possibility is that there is some other edge of G_5 containing two leaves and we can apply Lemma 7.2 (ii). ∎

8. Reconstructing networks from quarnets. We have focussed so far on reconstructing networks from their X-deck. We could try to use a recursive argument in order to reconstruct networks from smaller subnetworks, with less than $|X| - 1$ leaves. However, this approach does not work in general since there are networks for which no elements of its X-deck are phylogenetic networks, see Figure 5. Nevertheless, it is possible to apply a recursive approach if we use the following variant of the X-deck of a network.

DEFINITION 8.1. *Given a phylogenetic network N on X and a leaf $x \in X$, the phylogenetic network $N_x^{\mathcal{P}}$ is the result of deleting leaf x from N, together with its incident edge, and applying the following three operations until none is applicable:*

(i) suppress a degree-2 vertex;

(ii) replace a pair of multi-edges by a single edge;

(iii) collapse a blob with precisely two incident cut-edges into a single vertex.

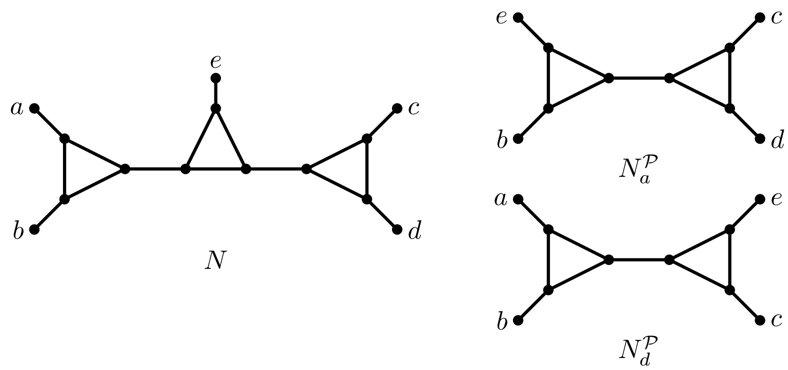

FIG. 5. *An example of a level-1 phylogenetic network N on X such that no elements of its X-deck are phylogenetic networks. Nevertheless, it is possible to reconstruct N from the quarnets $N_a^{\mathcal{P}}$ and $N_d^{\mathcal{P}}$.*

Given a phylogenetic network N on X and $X' \subseteq X$, the phylogenetic X'-deck *of N is the set $\{N_x^{\mathcal{P}} \mid x \in X'\}$.*

See again Figure 5 for an example. Note that this form of leaf-deletion was introduced for directed level-1 phylogenetic networks in [10] – see also [9] for more details for general phylogenetic networks.

All elements of a phylogenetic X-deck are phylogenetic networks by the following observation, which is easily verified.

OBSERVATION 4. *Let N be a phylogenetic network N on X, with $|X| \geq 3$, and $x \in X$. Then $N_x^{\mathcal{P}}$ is a phylogenetic network on $X \setminus \{x\}$.*

This opens the door to reconstructing networks from smaller subnetworks. A *quarnet* is a phylogenetic network with precisely four leaves. The set of quarnets $Q(N)$ of a phylogenetic network N on X is defined recursively by $Q(N) = \{N\}$ if $|X| = 4$ and

$$Q(N) = \bigcup_{x \in X} Q(N_x^{\mathcal{P}}) \quad \text{if } |X| \geq 5.$$

Here, the union operation keeps one phylogenetic network from each group of equivalent phylogenetic networks. We say that two sets $\mathcal{N}, \mathcal{N}'$ of phylogenetic networks are *equivalent*, denoted $\mathcal{N} \sim \mathcal{N}'$, if there exists a bijection $f : \mathcal{N} \to \mathcal{N}'$ with $N \sim f(N)$ for all $N \in \mathcal{N}$.

We say that a network N is *reconstructible from its quarnets* if every phylogenetic network N' with $Q(N) \sim Q(N')$ is equivalent to N. Moreover, a class $\mathcal{N}$ of phylogenetic networks is *quarnet-reconstructible* if each $N \in \mathcal{N}$ is reconstructible from its quarnets.

Similarly, N is *reconstructible from its phylogenetic X-deck* if every phylogenetic network N', whose phylogenetic X-deck is equivalent to the phylogenetic X-deck of N, is equivalent to N. Moreover, a class $\mathcal{N}$ of phylogenetic networks is *phylogenetically reconstructible* if each $N \in \mathcal{N}$ is reconstructible from its phylogenetic X-deck.

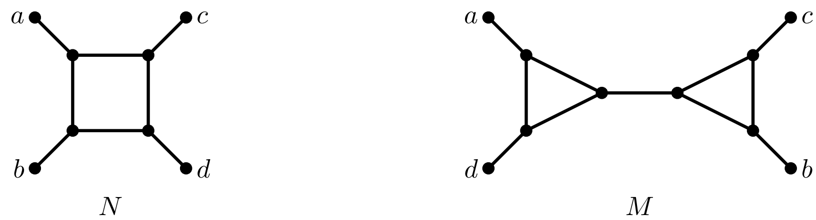

FIG. 6. *Two phylogenetic networks that have the same phylogenetic X-deck but not the same X-deck (even though the X-deck and phylogenetic X-deck of N are equivalent). Network N is neither X-reconstructible nor reconstructible from its phylogenetic X-deck, while M is X-reconstructible but not reconstructible from its phylogenetic X-deck.*

If two phylogenetic networks on X have equivalent X-decks, then they have equivalent phylogenetic X-decks (but not conversely, see Figure 6). Consequently, if a phylogenetic network on X is reconstructible from its phylogenetic X-deck, then it is X-reconstructible. The following proposition, which shows that the converse is also true in some cases, will permit us to apply results from previous sections.

PROPOSITION 8.2. *Let N be a phylogenetic network on X with $|X| \geq 4$. If N is Y-reconstructible for some $Y \subseteq X$ with $|Y| \geq 2$ and $N_y^{\mathcal{P}} \sim N_y$ for all $y \in Y$, then N is reconstructible from its phylogenetic X-deck.*

Proof. Suppose that there exists a network M that is not equivalent to N but has an equivalent phylogenetic X-deck. Since N is Y-reconstructible, there exists a $y \in Y$ such that $N_y \not\sim M_y$. Since $M_y^{\mathcal{P}} \sim N_y^{\mathcal{P}} \sim N_y$, it follows that $M_y^{\mathcal{P}} \not\sim M_y$ and hence that the neighbour of y in M is in a triangle. Moreover, since N_y has the same reticulation number as N, $M_y^{\mathcal{P}}$ also has the same reticulation number as N. Since, in M, the neighbour of y is in a triangle, M has a higher reticulation number than $M_y^{\mathcal{P}}$ and N. Take any $z \in Y \setminus \{y\}$. Then, since $M_z^{\mathcal{P}} \sim N_z^{\mathcal{P}} \sim N_z$, $M_z^{\mathcal{P}}$ has the same reticulation number as N and $M_y^{\mathcal{P}}$ and hence a lower reticulation number than M. It follows that the neighbour of z in M is also in a triangle. We distinguish two cases.

First assume that the neighbours of y and z are both in the same triangle in M. Consider any two leaves $x, p \in X \setminus \{y, z\}$. Then, the neighbours of y and z are together in the same triangle in $M_x^{\mathcal{P}} \sim N_x^{\mathcal{P}}$ and in $M_p^{\mathcal{P}} \sim N_p^{\mathcal{P}}$. On the other hand, neither of the neighbours of y and z is in a triangle in N, since $N_z^{\mathcal{P}} \sim N_z$ and $N_y^{\mathcal{P}} \sim N_y$. This is only possible when N is a simple level-1 network on $X = \{x, y, z, p\}$. This contradicts the assumption that N is Y-reconstructible, with $Y \subseteq X$, and hence X-reconstructible.

Now assume that the neighbours of y and z are in different triangles in M. Then, the neighbour of z is also in a triangle in $M_y^{\mathcal{P}} \sim N_y$. On the other hand, the neighbour of z is not in a triangle in N, since $N_z^{\mathcal{P}} \sim N_z$. Hence, in N, the neighbours of y and z are part of a 4-cycle. Consider again two leaves $x, p \in X \setminus \{y, z\}$. In $N_x^{\mathcal{P}} \sim M_x^{\mathcal{P}}$ and in $N_p^{\mathcal{P}} \sim M_p^{\mathcal{P}}$, the neighbours of y and z are in a triangle or 4-cycle. This is only possible when, in M, the neighbours of (without loss of generality) x and y are in one triangle while the neighbours of p and z are in a different triangle, and the two triangles are adjacent. This implies that there are no other leaves, i.e. $X = \{x, y, z, p\}$, and again N is a simple level-1 network on X. This again leads to a contradiction since N is X-reconstructible. ∎

In particular, we have the following.

FIG. 7. *Phylogenetic networks on $X = \{a, b, c\}$ that are X-reconstructible but not reconstructible from their phylogenetic X-deck.*

COROLLARY 8.3. *Let N be a phylogenetic network on X with $|X| \geq 4$. If the X-deck of N consists of only phylogenetic networks, then N is reconstructible from its phylogenetic X-deck if and only if N is X-reconstructible.*

Note that Corollary 8.3 does not hold when $|X| = 3$, see Figure 7.

THEOREM 8.4. *Let $\mathcal{N}$ be a class of phylogenetic networks such that each element of $\mathcal{N}$ has at least five leaves and, for each element N of $\mathcal{N}$ with at least six leaves, the phylogenetic X-deck of N is equivalent to a subset of $\mathcal{N}$. Then $\mathcal{N}$ is phylogenetically-reconstructible if and only if it is quarnet-reconstructible.*

Proof. If $\mathcal{N}$ is quarnet-reconstructible then it is phylogenetically-reconstructible since if two phylogenetic networks $N, N' \in \mathcal{N}$ have equivalent phylogenetic X-decks then it follows directly that $Q(N) \sim Q(N')$.

Now suppose that $\mathcal{N}$ is phylogenetically-reconstructible. We prove by induction on i that each $N \in \mathcal{N}$ with at most i leaves is quarnet-reconstructible. If $i = 5$ then the phylogenetic X-deck of N is equal to $Q(N)$ and therefore N is quarnet-reconstructible. Now suppose $i \geq 6$. Since N is reconstructible from its X-deck and each element of its X-deck is, by induction, quarnet-reconstructible, N is quarnet-reconstructible. ◻

First observe that each phylogenetic tree on X with $|X| \geq 5$ is reconstructible from its phylogenetic X-deck by Theorem 4.1 and Proposition 8.2. Hence, the class of phylogenetic trees with at least five leaves is phylogenetically reconstructible.

However, a similar argument cannot be used to show that even the class of level-1 networks is phylogenetically reconstructible. Therefore, it is interesting to study which classes of networks are phylogenetically reconstructible.

THEOREM 8.5. *The class of level-3 phylogenetic networks with at least five leaves is phylogenetically reconstructible.*

To prove this theorem, we will first show that an analogue of Theorem 4.3 holds.

THEOREM 8.6. *The class of decomposable phylogenetic networks with at least five leaves is phylogenetically reconstructible.*

Proof. The proof is very similar to that of Theorem 4.3. As in that proof, first note that a phylogenetic network has at least one nontrivial cut-edge if and only if at most two elements of its phylogenetic X-deck do not. Let N be some phylogenetic network on X with at least one nontrivial cut-edge and $|X| \geq 5$. Since $(T(N))^{\mathcal{P}}_x = T(N^{\mathcal{P}}_x)$, for all $x \in X$, we can reconstruct $T(N)$ from the phylogenetic X-deck of N. We can then use exactly the same argument as in the last part of the proof of Theorem 4.3

to show that N is reconstructible from its phylogenetic X-deck (see Figure 5 for an illustration). ∎

We now prove Theorem 8.5.

Proof. By Theorem 8.6, it suffices to consider simple level-k networks with $1 \leq k \leq 3$. For simple level-1 networks, the phylogenetic X-deck is precisely equal to the X-deck and we are done by Proposition 8.2.

Now consider a simple level-2 network N and its underlying generator G. If the phylogenetic X-deck of N is not equal to its X-deck then one of the three edges of G contains exactly one leaf x, another edge of G contains no leaves, and the third edge of G contains all other leaves $X \setminus \{x\}$. Then N is $\{y,z\}$-reconstructible for any $y, z \in X \setminus \{x\}$ with distance between them at least 4. Since $N_y^{\mathcal{P}} = N_y$ and $N_z^{\mathcal{P}} = N_z$ we are done by Proposition 8.2.

Therefore, we may assume that N is a simple level-3 network. Suppose the phylogenetic X-deck of N is not equal to its X-deck. Then the underlying generator G of N is not equal to K_4 (since K_4 does not have any multi-edges). Hence, G is the other level-3 generator, see Figure 4. Moreover, at least one pair of multi-edges contains precisely one leaf, say leaf x. The other pair of multi-edges contains at least one leaf y.

If there is at least one leaf z on an edge that is not in a pair of multi-edges, then it is straightforward to check that, wherever you put leaves $p, q \in X \setminus \{x, y, z\}$, there is a cycle containing the neighbours of leaves a, b, c, d satisfying the conditions of Lemma 7.2(i) and a fifth leaf e such that $N_d^{\mathcal{P}} = N_d$ and $N_e^{\mathcal{P}} = N_e$, and we are done by Proposition 8.2.

The only remaining case is that all leaves in $X \setminus \{x\}$ are on the pair of multi-edges not containing x. Then there is again a cycle containing the neighbours of leaves a, b, c, d satisfying the conditions of Lemma 7.2(i) and a fifth leaf e such that $N_d^{\mathcal{P}} = N_d$. However, if $|X| = 5$ then the only choice for e is $e = x$ and hence $N_e^{\mathcal{P}} \not\sim N_e$. Nevertheless, we can use a similar argument as in the proof of Lemma 7.2(i) since $N_e^{\mathcal{P}}$ does contain a unique cycle containing the neighbours of a, b, c and d. ∎

Corollary 8.7. *Any level-3 phylogenetic network is reconstructible from its quarnets.*

9. Edge-reconstructibility.

In this section we shall consider the problem of reconstructing a phylogenetic network from its edge-deleted networks. We first formalize this concept (cf. [3, Section 2] for a review of edge-reconstruction in graphs).

Given a phylogenetic network N and an edge $e \in E(N)$, the pseudo-network N_e is the result of deleting edge e from N and suppressing resulting degree-2 vertices. The *edge-deck* of N is the multiset $\{N_e \mid e \in E(N)\}$. An *edge-reconstruction* of a network N on X is a network N' on X with $E(N') = E(N)$ and $N'_e \sim N_e$ for all $e \in E(N)$. Note that by $E(N') = E(N)$ we do not mean that the edges of N are the same pairs of vertices as the edges of N', but that there exists a bijection $f : E(N) \to E(N')$ which we assume to be the identity. We call a phylogenetic network N *edge-reconstructible* if every edge-reconstruction of N is equivalent to N.

Lemma 9.1. *Let N be a phylogenetic network on X. If N is leaf-reconstructible then it is edge-reconstructible.*

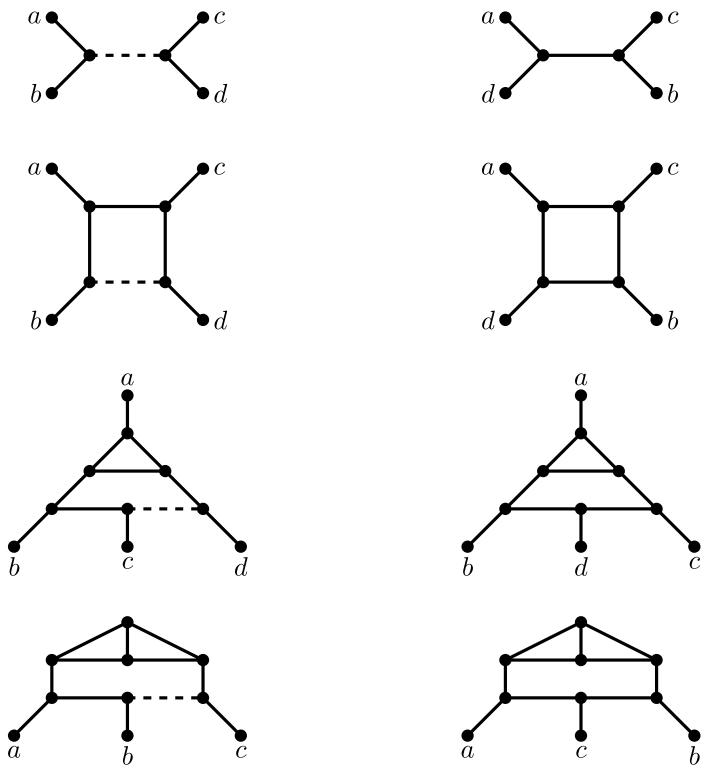

FIG. 8. *Pairs of phylogenetic networks that are not leaf-reconstructible but that are edge-reconstructible. The dashed edges indicate an edge e such that N_e is not contained in the edge-deck of the other network of the pair.*

Proof. This follows directly from the observation that $N_e \sim N'_e$ if and only if $N_x \sim N'_x$ for each edge e that has an endpoint $x \in X$ in both N and N'. ∎

However, there exist edge-reconstructible networks that are not leaf-reconstructible, see the examples in Figure 8.

When considering edge-reconstructability of binary networks we can, by Theorem 4.3 and Lemma 9.1, again restrict to simple networks.

We say that (x, y) is a *2-chain* of a phylogenetic network N on X if $x, y \in X$ and the distance between x and y in N is 3.

PROPOSITION 9.2. *Any simple binary phylogenetic network on X containing a 2-chain is edge-reconstructible.*

Proof. The fact that N is simple can be recognized by considering three elements of its edge-deck $N_{e_1}, N_{e_2}, N_{e_3}$ such that each of e_1, e_2, e_3 is incident to a leaf. Since each of $N_{e_1}, N_{e_2}, N_{e_3}$ consists of a simple network and an isolated vertex, any edge-reconstruction of N is simple.

Suppose that N has a 2-chain (x, y). Let u and v be the neighbours of x and y in N respectively and $e = \{u, v\}$. Let u' and v' be the neighbours of x and y in N_e respectively.

First suppose that (x, y) is not a 2-chain in N_e. There exists at least one edge f that is not incident to u or v. Since (x, y) is a 2-chain in N_f, we can uniquely reconstruct N from N_e by subdividing the edges $\{u', x\}$ and $\{v', y\}$ and creating a new edge between the subdividing vertices.

Now suppose that (x, y) is also a 2-chain in N_e. We say that a network has an *xy-ladder* of *length* k if there exist disjoint paths $(x, u_1, \ldots, u_k)$ and $(y, v_1, \ldots, v_k)$ such that u_i and v_i are adjacent for $1 \leq i \leq k$. Let $p \geq 1$ be the maximum length of an xy-ladder in N. Take any such ladder and observe that there exists at least one edge g that is not incident to any vertex of the ladder. Then the maximum length of an xy-ladder is p in N_g and is $p-1$ in N_e. Hence, we can again uniquely reconstruct N from N_e by subdividing the edges $\{u', x\}$ and $\{v', y\}$ and creating a new edge between the subdividing vertices. ∎

The following corollary can be proved in a similar way to Corollaries 5.2 and 5.3.

Corollary 9.3.

(i) Any simple binary level-k phylogenetic network on X with $k \geq 2$ and $|X| \geq 3k - 2$ is edge-reconstructible.

(ii) Any binary phylogenetic network $N = (V, E)$ on X with $|X| \geq \max\{3(|E| - |V|) + 1, 5\}$ is edge-reconstructible.

10. Discussion. In this paper we have introduced the concept of leaf-reconstructible phylogenetic networks. We have shown that several large classes of phylogenetic networks are leaf-reconstructible, and used our results to show that level-3 networks are defined by their quarnets. We conjecture that all unrooted phylogenetic networks with 5 or more leaves are leaf-reconstructible. We expect that this could be a difficult conjecture to settle, as with other variants of the graph reconstruction conjecture.

In another direction, it could be of interest to also consider leaf-reconstructibility of nonbinary networks. In Theorem 4.1, we showed that nonbinary phylogenetic trees are leaf-reconstructible, and in Theorem 4.3 that even all decomposable nonbinary phylogenetic networks are leaf-reconstructible, but what about non-decomposable nonbinary networks? The following related question could also be worth considering: If every nonbinary phylogenetic network with at least five leaves is leaf-reconstructible, then is every graph reconstructible?

In Section 9, we considered edge-reconstructibility, a variant of the leaf-reconstructibility problem. Another variant that should be considered is leaf-reconstructibility for directed phylogenetic networks. This is an important class of networks, in which the networks are directed acyclic graphs, with a single root and leaves labeled by the set X. In [9] certain examples of directed phylogenetic networks are presented which indicate that such networks may not be leaf-reconstructible, but it remains an open problem whether or not this is the case (note that not all digraphs are reconstructible [17]).

In the longer term, it would be interesting to consider leaf-reconstructibility of networks that arise in biological settings. Indeed, even if not every network is leaf-reconstructible, it may be that counter-examples are somewhat unlikely to occur as evolutionary histories (e.g. if they are highly symmetric).

One way to approach this could be to consider random networks. As we have seen in Corollary 5.4, for any fixed k, almost all level-k phylogenetic networks are leaf-reconstructible. It would be interesting to know whether or not almost all phylogenetic networks on a fixed leaf-set are leaf-reconstructible. In this context, it is worth noting that almost every graph has reconstructing number three [2]. We have shown that decomposable and binary level-4 networks with at least five leaves have reconstruction number at most 2. So, do almost all (binary) phylogenetic networks have reconstruction number at most 2?

Finally, it would be interesting to consider leaf-reconstructibilty of networks that are generated according to some model of molecular evolution (see e.g. [4] for a review of such models). This would be somewhat analogous to recent ground-breaking work on reconstructibility of pedigrees in a stochastic setting [19, 20], and could focus on models such as those presented in, for example, [13].

REFERENCES

[1] K.J. Asciak, M.A. Francalanza, J. Lauri and W. Myrvold, A survey of some open questions in reconstruction numbers, *Ars Combinatoria*, **97**: 443–456, 2010.
[2] B. Bollobás, Almost every graph has reconstruction number three, *Journal of Graph Theory*, **14** (1): 1–4, 1990.
[3] J.A. Bondy and R.L. Hemminger, Graph reconstruction – a survey, *Journal of Graph Theory*, **1**: 227–268, 1977.
[4] J. Felsenstein, *Inferring phylogenies*, Sunderland: Sinauer Associates, 2004.
[5] P. Gambette and K.T. Huber, On encodings of phylogenetic networks of bounded level, *Journal of Mathematical Biology*, **65** (1): 157–180, 2012.
[6] P. Gambette, V. Berry, and C. Paul, Quartets and unrooted phylogenetic networks, *Journal of bioinformatics and computational biology*, **10**(4): 1250004, 2012.
[7] D. Gusfield, *ReCombinatorics: the algorithmics of ancestral recombination graphs and explicit phylogenetic networks*, MIT Press, 2014.
[8] D. Huson, R. Rupp and C. Scornavacca. *Phylogenetic networks: concepts, algorithms and applications* Cambridge University Press, 2010.
[9] K.T. Huber, L.J.J. van Iersel, V. Moulton and T. Wu, How much information is needed to infer reticulate evolutionary histories? *Systematic biology*, **64**:102–111, 2015.
[10] K.T. Huber and V. Moulton, Encoding and constructing 1-nested phylogenetic networks with trinets, *Algorithmica*, **66** (3): 714–738, 2013.
[11] F. Harary and E. Palmer, The reconstruction of a tree from its maximal subtrees, *Canadian Journal of Mathematics*, **18**: 803–810, 1966.
[12] L.J.J. van Iersel and V. Moulton, Trinets encode tree-child and level-2 phylogenetic networks, *Journal of Mathematical Biology*, **68**(7): 1707–1729, 2014.
[13] G. Jin, L. Nakhleh, S. Snir and T. Tuller, Maximum likelihood of phylogenetic networks, *Bioinformatics*, **22**: 2604–2611, 2006.
[14] P.J. Kelly, A congruence theorem for trees, *Pacific Journal of Mathematics*, **7**: 961–968, 1957.
[15] V. Krishnamoorthy and K. Parthasarathy, On the reconstruction conjecture for separable graphs, *Journal of the Australian Mathematical Society (Series A)*, **30**: 307–320, 1981.
[16] C. Semple and M. Steel, *Phylogenetics*, Oxford Lecture Series in Mathematics and its Applications, **24**, Oxford University Press, 2013.
[17] P.K. Stockmeyer, The falsity of the reconstruction conjecture for tournaments, *Journal of Graph Theory*, **1**: 19–25, 1977.
[18] B. Thatte, Combinatorics of pedigrees I: counterexamples to a reconstruction question, *SIAM Journal on Discrete Mathematics*, **22**(3): 961-970, 2008.
[19] B. Thatte, Reconstructing pedigrees: some identifiability questions for a recombination-mutation model, *Journal of mathematical biology*, **66** (1-2): 37–74, 2013.
[20] B. Thatte ad M. Steel, Reconstructing pedigrees: A stochastic perspective, *Journal of Theoretical Biology*, **251** (3): 440–449, 2008.
[21] Y. Yongzhi, The reconstruction conjecture is true if all 2-connected graphs are reconstructible, *Journal of Graph Theory*, **12** (2): 237–243, 1988.