

Document Version

Final published version

Citation (APA)

Pons, J. C., Lopez, P. V., Murakami, Y., & Iersel, L. V. (2025). Fence Decompositions and Cherry Covers in Non-binary Phylogenetic Networks. *IEEE Transactions on Computational Biology and Bioinformatics*, 22(5), 2173 - 2183.
<https://doi.org/10.1109/TCBBIO.2025.3587086>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

In case the licence states "Dutch Copyright Act (Article 25fa)", this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

**Green Open Access added to [TU Delft Institutional Repository](#)
as part of the Taverne amendment.**

More information about this copyright law amendment
can be found at <https://www.openaccess.nl>.

Otherwise as indicated in the copyright section:
the publisher is the copyright holder of this work and the
author uses the Dutch legislation to make this work public.

Fence Decompositions and Cherry Covers in Non-Binary Phylogenetic Networks

Joan Carles Pons , Pau Vives López , Yukihiro Murakami , and Leo van Iersel

Abstract—Reticulate evolution can be modelled using phylogenetic networks. Tree-based networks, which are one of the more general classes of phylogenetic networks, have recently gained eminence for its ability to represent evolutionary histories with an underlying tree structure. To better understand tree-based networks, numerous characterizations have been proposed, based on tree embeddings, matchings, and arc partitions. Here, we build a bridge between two arc partition characterizations, namely maximal fence decompositions and cherry covers. Results on cherry covers have been found for general phylogenetic networks. We first show that the number of cherry covers is the same as the number of support trees (underlying tree structure of tree-based networks) for a given semi-binary network. Maximal fence decompositions have only been defined thus far for binary networks (constraints on vertex degrees). We remedy this by generalizing fence decompositions to non-binary networks, and using this, we characterize semi-binary tree-based networks in terms of forbidden structures. Furthermore, we give an explicit enumeration of cherry covers of semi-binary networks, by studying its fence decomposition. Finally, we prove that it is possible to characterize semi-binary tree-child networks, a subclass of tree-based networks, in terms of the number of their cherry covers.

Index Terms—Phylogenetic networks, tree-based networks, fence decomposition, cherry cover.

I. INTRODUCTION

PHYLOGENETIC trees serve as a tool for depicting non-reticulate evolution (e.g. [1]). Their inherent simplicity contrasts with the impossibility of representing more complex evolutionary scenarios involving reticulate events like hybridizations, recombinations, or lateral gene transfers. Phylogenetic networks extend phylogenetic trees, allowing for the representation of these more complex situations [2], [3].

The space of phylogenetic networks is divided into classes to reproduce different biologically significant scenarios or to explore challenging mathematical and computational problems. One can find a review of such classes in [4]. One of the ongoing

discourses in literature is the debate regarding the nature of evolution. It revolves around whether evolution predominantly follows a tree-like pattern, punctuated by occasional horizontal occurrences [5], or if evolution inherently exhibits network-like characteristics without any resemblance to a tree structure [6]. Tree-based networks [7] were introduced to represent the networks of the first type. In graph theory terms, such networks are those that have a rooted spanning tree (called a *support tree*) for which the leaves are those of the network. Suppressing elementary nodes (nodes of indegree and outdegree equal to one) in the support tree results in a *base-tree* as the underlying tree structure for the network. In general, a tree-based network can have more than one support tree, and possibly more than one base tree (different support trees can give rise to the same base trees). From a biological perspective, it is not only relevant to determine the existence of an underlying tree-like signal within a network, but also to assess whether this tree is unique or whether there are many different trees that could equally well serve as base trees. Tree-based networks include several relevant subclasses of phylogenetic networks, including, among others, tree-child networks [8]. Equivalently, a network is tree-child if and only if each embedded tree, that is, is a support tree [9].

In addition to the network classes mentioned above, one typically characterizes phylogenetic networks based on the degree of vertices, that is, the number of arcs that every vertex is allowed to be incident to. This characterization is useful, and oftentimes necessary when there are ambiguities in the order of how evolutionary events have unfolded. These ambiguities are caused by insufficient data (*soft polytomies*) or a simultaneous divergence of multiple species from a single speciation event (*hard polytomies*) [10]; we also encounter the same ambiguities in reticulate evolution [11]. Roughly speaking, one considers *binary* networks when assuming no ambiguities are present, *semi-binary* networks when assuming ambiguities are present for reticulate evolution, and *non-binary* networks when assuming ambiguities are present in both speciations and reticulate evolution. Here, non-binary means that the network is ‘not necessarily binary’, and thus the class of binary networks are contained in the class of semi-binary networks, which is itself contained in the class of non-binary networks. A formal definition for the three classes is given in Section II-A. Note that in the literature, *semi-binary* is sometimes used to describe networks that have ambiguities in speciation events [12].

In this paper, we study and build a bridge between two characterizations of tree-based networks called *cherry covers* and *fence decompositions*, which are both based on arc partitions.

Received 21 July 2024; revised 27 June 2025; accepted 28 June 2025. Date of publication 10 July 2025; date of current version 9 October 2025. The work of Joan Carles Pons and Pau Vives López was supported in part by the “ERDF A way of making Europe” under Grant PID2021-126114NB-C44 and Grant MCIN/AEI/10.13039/501100011033, and in part by the Netherlands Organisation for Scientific Research (NWO) under Grant OCENW.M.21.306 and Grant OCENW.KLEIN.125. (Corresponding author: Joan Carles Pons.)

Joan Carles Pons and Pau Vives López are with the Department of Mathematics and Computer Science, University of the Balearic Islands, E-07122 Palma, Spain (e-mail: joancarles.pons@uib.es; pau.vives@uib.cat).

Yukihiro Murakami and Leo van Iersel are with the Delft Institute of Applied Mathematics, Delft University of Technology, 2600 Delft, AA, Netherlands (e-mail: y.murakami@tudelft.nl; l.j.vaniersel@tudelft.nl).

Digital Object Identifier 10.1109/TCBBIO.2025.3587086

The concept of cherry covers relative to phylogenetic networks was introduced in [13]. Roughly speaking, a cherry cover refers to a partition of the network arcs into sets of sizes two and three, such that each of these sets contains arcs that share endpoints in some way. It was shown that non-binary networks are tree-based if and only if they have a cherry cover. On the other hand, fence decompositions of phylogenetic networks were introduced in [14], [15], and similarly to cherry covers, it outlines a systematic approach to breaking down any network into its “fundamental” substructures. Roughly speaking, it is an arc partition in which every set is an “up-down path” of arcs, meaning that each arc shares at least one and at most two endpoints with other arcs in the set. This partition can be used to characterize binary tree-based networks in terms of forbidden structures, or to count support trees in binary tree-based networks. We give formal definitions and known results on cherry covers and fence decompositions in Sections II-B and II-C, respectively.

Both cherry covers and fence decomposition have been used for the study of tree-based networks; even so, the concepts have not been studied in a framework in which both can be handled together. In Section III, we prove that the number of cherry covers is the same as the number of support trees for semi-binary tree-based networks. In Section IV, we show that the number of cherry covers (and thus also the number of support trees) of a binary tree-based network can be written as a formula involving the elements of its fence decomposition. Notice that the problem of counting the number of support trees in binary tree-based networks have been previously solved in [12], [16] by the association of a bipartite graph to the network and in [15] by the use of the fence decomposition. In Section V we extend the fence decomposition of binary phylogenetic networks introduced in [15] to non-binary networks. This requires the introduction of a new set of fence structures, and in so doing, we extend problems defined for binary networks to the non-binary setting. In Section VI we solve two of these problems. We extend to semi-binary tree-based networks (a) the problem of counting the number of their cherry covers (and therefore the number of their support trees) from the “new” fence decomposition, mimicking the results from Section IV, and (b) the characterisation of semi-binary tree-based networks in terms of forbidden structures. Both generalisations require similar approaches, albeit much more technical and tricky than their binary counterparts. Finally, in Section VII, we prove that the subclass of semi-binary tree-child networks can be characterised in terms of the number of their cherry covers. The paper ends with some concluding remarks in Section VIII.

II. PRELIMINARIES

In this section we give some definitions and results that will be used throughout the manuscript.

A. Phylogenetic Networks

A (non-binary) *phylogenetic network*, or simply a *network*, $N = (V, A)$ on a set X of *taxa*, is a directed acyclic graph (V, A) without parallel arcs such that any node $u \in V$ is either:

- 1) a *root*, with $\text{indeg } u = 0$, $\text{outdeg } u = 1$ (and there can only be one such node), or
- 2) a *leaf*, with $\text{indeg } u = 1$, $\text{outdeg } u = 0$, or
- 3) a *tree node*, with $\text{indeg } u = 1$, $\text{outdeg } u \geq 2$, or
- 4) a *reticulation*, with $\text{indeg } u \geq 2$, $\text{outdeg } u = 1$, together with a bijective function between X and the set of leaves. Vertices in $V \setminus X$ are called *interior nodes*.

The arc incident to the root is called the *root arc*, and arcs that feed into a reticulation node are called *reticulation arcs*. For a given arc $a \in A$, where $a = uv$, we define the *tail* of a as $\text{tail}(a) = u$ and the *head* of a as $\text{head}(a) = v$. A network is *semi-binary* if all tree nodes have outdegree 2 (reticulations with indegree greater than 2 are allowed), and it is *binary* if, in addition to these conditions, all reticulations have indegree 2. We write non-binary to mean *not necessarily binary*. So the class of binary networks are contained in the class of semi-binary networks, which are themselves contained in the class of non-binary networks.

In this paper we focus on some relevant subclasses of phylogenetic networks, mainly tree-based and tree-child networks.

A (non-binary) network N is *tree-based* [7], [12] with *base-tree* T when N can be obtained from T by the following process:

- 1) subdivide some arcs of T generating elementary (indegree and outdegree equal to one) nodes called *attachment points*;
- 2) add new arcs between pairs of attachment points and from nodes in T to attachment nodes, in such a way that N remains acyclic, without parallel arcs, and such that each attachment point maintains indegree or outdegree 1;
- 3) suppress every elementary node.

Removing all arcs added in step (2) gives a subdivision of T , called a *support tree* of N . In other terms, a tree (possibly containing elementary nodes) obtained from N by removing all but one incoming reticulation arcs from each reticulation without creating new leaves is a support tree. Note that the set of vertices of a support tree for N is also the set of vertices of N .

A (non-binary) network is *tree-child* [8] if every interior node has at least one child which is a tree node or a leaf. The class of tree-child networks is included in the class of tree-based networks.

B. Cherry Covers

We adapt here some definitions and results from [13]. Let $N = (V, A)$ be a semi-binary phylogenetic network.

A *cherry shape* in N is a subgraph of N composed by three different nodes $x, y, p \in V$ and two arcs $px, py \in A$. We denote the cherry shape by its set of arcs $\{px, py\}$. We will refer to p as the *internal node* and to x and y as the *terminal nodes*. A *reticulated cherry shape* in N is a subgraph of N composed by four different nodes $x, y, p_x, p_y \in V$ and three arcs $p_x x, p_y p_x, p_y y \in A$, with p_y a tree node and p_x a reticulation. Similarly as in the previous case, we denote the reticulated cherry shape as $\{p_x x, p_y p_x, p_y y\}$, and we call p_x, p_y the internal nodes, x, y the terminal nodes, and $p_y p_x$ the *middle arc*. We refer to both cherry and reticulated cherry shape as *shapes*.

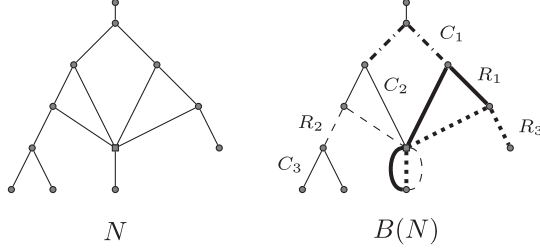


Fig. 1. A semi-binary network N and its bulged version $B(N)$ with a possible cherry cover $\{C_1, C_2, C_3, R_1, R_2, R_3\}$. The different line type indicates the edges of each cherry shape.

A set P of shapes is a *cherry cover* for a binary network $N = (V, A)$, if every arc in A , apart from the root arc, is contained in a single shape of P . To adapt the concept of cherry covers to semi-binary networks, we require the following definition. The *bulged version* of a semi-binary network N , denoted by $B(N)$, is the multigraph obtained from N by replacing the outgoing arc of each reticulation node u with $\text{indeg } u = k$ by $k - 1$ parallel arcs. Notice that the bulged version of a binary network is (isomorphic to) the network itself. Then a *cherry cover* of the bulged version $B(N)$ of a semi-binary network N is a set P of shapes such that each arc of $B(N)$, apart from the root arc, is contained in a single shape of P . See an example in Fig. 1.

Cherry covers can be used to characterize the class of tree-based networks.

Theorem 1 (adapted from Theorem 3.3 of [13]): A semi-binary phylogenetic network N is tree-based if, and only if, $B(N)$ has a cherry cover.

C. Fence Decomposition

We include here another static characterization of binary tree-based networks based on an arc partition, called maximal zig-zag trails [14], [15]. Let N be a binary network. A *zig-zag trail* of length k is a sequence (a_1, a_2, \dots, a_k) of different arcs where $k \geq 1$, where either $\text{head}(a_{2i-1}) = \text{head}(a_{2i})$ and $\text{tail}(a_{2j}) = \text{tail}(a_{2j+1})$ holds for $i \in [\lfloor \frac{k}{2} \rfloor] = \{1, 2, \dots, \lfloor \frac{k}{2} \rfloor\}$ and $j \in [\lfloor \frac{k-1}{2} \rfloor]$, or $\text{tail}(a_{2i-1}) = \text{tail}(a_{2i})$ and $\text{head}(a_{2j}) = \text{head}(a_{2j+1})$ holds for $i \in [\lfloor \frac{k}{2} \rfloor]$ and $j \in [\lfloor \frac{k-1}{2} \rfloor]$. We call a zig-zag trail Z *maximal* if there is no zig-zag trail that contains Z as a subsequence. Depending on the nature of $\text{tail}(a_1)$ and $\text{tail}(a_k)$, we have four possible maximal zig-zag trails.

- **Crowns:** $k \geq 4$ is even and $\text{tail}(a_1) = \text{tail}(a_k)$ or $\text{head}(a_1) = \text{head}(a_k)$. By reordering the arcs, assume henceforth that $\text{head}(a_1) = \text{head}(a_k)$.
- **M-fences:** $k \geq 2$ is even, it is not a crown, and $\text{tail}(a_i)$ is a tree vertex for every $i \in [k]$.
- **N-fences:** $k \geq 1$ is odd and $\text{tail}(a_1)$ or $\text{tail}(a_k)$, but not both, is a reticulation. By reordering the arcs, assume henceforth that $\text{tail}(a_1)$ is a reticulation and $\text{tail}(a_k)$ a tree vertex.
- **W-fences:** $k \geq 2$ is even and both $\text{tail}(a_1)$ and $\text{tail}(a_k)$ are reticulations.

We call a set S of maximal zig-zag trails a *fence decomposition* of N if the elements of S partition all arcs, except for the root arc,

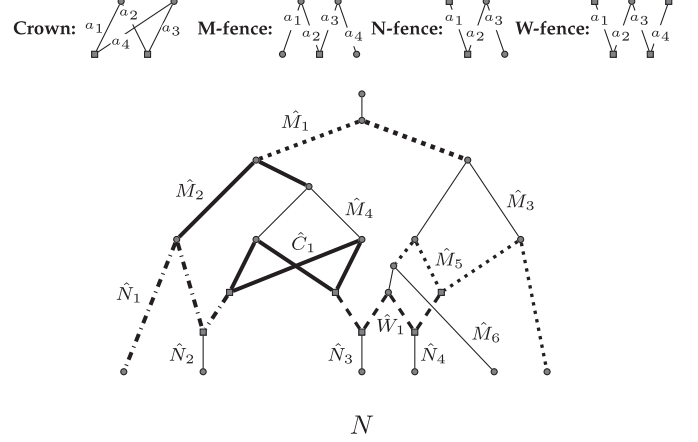


Fig. 2. Representation of the different zig-zag trails present in the binary setting and a network N with fence decomposition $S = \bigcup_{i=1}^6 \{\hat{M}_i\} \cup \bigcup_{i=1}^4 \{\hat{N}_i\} \cup \{\hat{W}_1\} \cup \{\hat{C}_1\}$. The different line types indicates the arcs of each maximal zig-zag trail.

of N . See an example in Fig. 2. One can characterize networks using such a decomposition.

Theorem 2 (adapted from Theorem 4.2 of [15]): Any binary network N has a unique fence decomposition.

One can also obtain a forbidden structure characterization of binary tree-based networks using the decomposition.

Lemma 3 (adapted from Corollary 4.6 of [15]): Let N be a binary network. Then N is tree-based if and only if it has no W -fences.

For non-crown maximal zig-zag trails, we require the notion of *endpoints* to be used in later proofs. Let $A = (a_1, \dots, a_k)$ be a maximal zig-zag trail. We shall write $\text{End}(A)$ to denote the set of endpoints of A . If A is an M -fence, then $\text{End}(A) = \{\text{head}(a_1), \text{head}(a_k)\}$. If A is an N -fence, then $\text{End}(A) = \{\text{tail}(a_1), \text{head}(a_k)\}$. If A is a W -fence, then $\text{End}(A) = \{\text{tail}(a_1), \text{tail}(a_k)\}$.

III. CHERRY COVERS AND SUPPORT TREES IN SEMI-BINARY NETWORKS

Recall that cherry covers in the semi-binary setting were defined over the bulged version of the network, obtained by adding, between each reticulation and its child, as many parallel arcs as the indegree of the reticulation considered minus 1. Over this bulged version, the definition of a cherry cover is given as in the binary case. These additional parallel arcs may present duplicates when counting the number of cherry covers for a semi-binary network. For example, consider two reticulated cherry shapes $\{a_1, b, c\}$ and $\{a_2, b, c\}$, where a_1 and a_2 are parallel arcs. Clearly, one should consider the shapes to be equivalent; in this vein, we say that such two reticulated cherry shapes are *isomorphic*. Additionally, we say that two cherry covers P_1 and P_2 are *isomorphic* if there exists a one-to-one correspondence between their set of reticulated cherry shapes into isomorphic reticulated cherry shapes. Informally, the isomorphism between two cherry covers can be understood as an isomorphism between the parallel arcs added in the bulged version (in a sense that we do not distinguish them).

We denote by \mathcal{P}_N and \mathcal{S}_N the set of all (non-isomorphic) cherry covers and the set of support trees of a semi-binary network N , respectively. Notice that from Theorem 1, $\mathcal{P}_N \neq \emptyset$ if and only if N is tree-based. Moreover, by definition, N is tree-based if and only if $\mathcal{S}_N \neq \emptyset$. In the next proposition we prove that the number of cherry covers coincides with the number of support trees in semi-binary networks.

Proposition 4: Let N be a semi-binary network. Then $|\mathcal{P}_N| = |\mathcal{S}_N|$.

Proof: If N is not tree-based, then $|\mathcal{P}_N| = |\mathcal{S}_N| = 0$. So suppose that N is tree-based.

Let P be a cherry cover of N . Let $\{p_x x, p_y p_x, p_y y\} \in P$ be a reticulated cherry shape. Observe that deleting the middle arc $p_y p_x$ reduces the indegree of p_x by 1 and the outdegree of p_y by 1. The degrees of all other vertices remain unchanged. Therefore, the resulting subgraph has the same leafset as that of N . Consider T , a tree obtained by removing all reticulation arcs of N that are covered as middle arcs of reticulated cherry shapes in P . By the same argument as above, T is a subgraph of N with the same leafset as that of N . Thus, T must be a support tree. So for every cherry cover, we can construct a support tree. This shows that $|\mathcal{P}_N| \leq |\mathcal{S}_N|$.

Let S be a support tree of N , obtained by removing all but one incoming reticulation arcs from each reticulation. We cover each such deleted arcs $\{a_1, \dots, a_k\}$ with the middle arc of a reticulated cherry shape. We claim that we can extend the covering to a cherry cover of N . Since N is semi-binary, $\text{tail}(a_i)$ is a tree vertex of outdegree-2 for each i . So covering a_i as a middle arc also covers the other outgoing arc of $\text{tail}(a_i)$ and also the outgoing arc of $\text{head}(a_i)$. Covering in this way does not cause any overlaps between the reticulated cherry shapes, except possibly at the parallel outgoing arcs of reticulations (in the bulged version of N). Indeed, it is not possible for both outgoing arcs of a tree vertex to be middle arcs, since S is a support tree of N . At this moment, every outgoing parallel arcs of reticulations are covered (in the bulged version of N). Let u be a tree vertex. Either both outgoing arcs of u are covered by a single reticulated cherry shape, or none of its outgoing arcs are covered. In the latter case, we cover the two outgoing arcs using a cherry shape. Repeating this ensures that all outgoing arcs of all vertices are covered by cherry shapes and reticulated cherry shapes. So for every support tree, we can construct a cherry cover. It follows that $|\mathcal{S}_N| \leq |\mathcal{P}_N|$. \square

IV. COVERING FENCES USING CHERRY COVERS

In this section, we show how one can obtain a cherry cover from a fence decomposition in binary tree-based networks. Recall that a network is tree-based if and only if it has a cherry cover (Theorem 1). Additionally, a network is tree-based if and only if it has no W -fences in its unique fence decomposition (Lemma 3). Note that, in a binary network, exactly one reticulation arc for every reticulation is contained in a reticulated cherry shape as a middle arc.

Let N be a network and let \hat{N} be an N -fence of N . In what follows, we shall write $\hat{N} := (a_1, a_2, \dots, a_k)$, and we will let c_{2j-1} denote the arc connecting $\text{head}(a_{2j-1})$ and its

only child for $j \in [\frac{k-1}{2}]$. We refer to a_1 as the *first arc* of the N -fence \hat{N} . The first lemma states that although a tree-based network may have non-unique cherry covers, the reticulated cherry shapes that cover arcs of N -fences are fixed.

Lemma 5 (adapted from Lemma 13 of [17]): Let N be a binary tree-based network that has an N -fence \hat{N} of length at least 3. Then every cherry cover of N contains the reticulated cherry shapes $\{c_{2j-1}, a_{2j}, a_{2j+1}\}$ for $j \in [\frac{k-1}{2}]$.

Note that the arcs of the reticulated cherry shapes do not exactly coincide with the arcs of \hat{N} in Lemma 5. In particular, the arc a_1 of \hat{N} is not covered; we shall see later that a_1 will be covered as an outgoing arc of a reticulation in another fence. Also, observe that the reticulated cherry shapes cover the arcs c_{2j-1} , which are not in \hat{N} . Such arcs are contained as first arcs in other N -fences. They will not be covered in other reticulated cherry shapes, similarly to how a_1 was not covered by the reticulated cherry shapes corresponding to \hat{N} .

We shall show similar results for M -fences and crowns. Let \hat{M} be an M -fence of N . In what follows, we shall write $\hat{M} := (a_1, a_2, \dots, a_k)$, and we will let c_{2j} denote the arc connecting $\text{head}(a_{2j})$ and its child for $j \in [\frac{k-2}{2}]$.

Lemma 6: Let N be a binary tree-based network, and let \hat{M} be an M -fence of length $k \geq 4$. There exists $k/2$ distinct ways of covering the arcs of \hat{M} and the arcs c_{2j} for $j \in [\frac{k-2}{2}]$ using cherry and reticulated cherry shapes.

Proof: In any cherry cover of N , exactly one incoming arc of every reticulation is covered as a middle arc of a reticulated cherry shape. Since the M -fence contains $k/2 - 1$ reticulations, we must have exactly $k/2 - 1$ reticulated cherry shapes that cover the arcs of \hat{M} and the arcs c_{2i} . By construction, the remaining two arcs of \hat{M} must be arcs a_m and a_n for some $m, n \in [k]$. Suppose without loss of generality that $n > m$. We show that $n = m + 1$. If not, then consider the arc a that is different from a_m in \hat{M} where $\text{tail}(a) = \text{tail}(a_m)$. Note that $a = a_{m-1}$ or $a = a_{m+1}$. By assumption, a is covered by a reticulated cherry shape. But since $\text{tail}(a)$ is a tree vertex, and since a is covered as the middle arc of a reticulated cherry shape, a can only be covered in a shape together with a_m . This gives the required contradiction, since a_m is yet to be covered. Thus $n = m + 1$, and we may cover these remaining arcs as a cherry shape, $\{a_m, a_{m+1}\}$.

Therefore every M -fence of length $k \geq 4$ must be covered by a single cherry shape and $k/2 - 1$ reticulated cherry shapes in every cherry cover. Note that the choice of arcs in the cherry shape determines the cherry cover for the M -fence subgraph. We conclude that there are $k/2$ distinct ways of covering the arcs of \hat{M} and the arcs c_{2j} for $j \in [\frac{k-2}{2}]$ using cherry and reticulated cherry shapes. \square

Finally let \hat{C} be a crown of N . In what follows, we shall write $\hat{C} := (a_1, a_2, \dots, a_k)$, and we will let c_{2j} denote the arc connecting $\text{head}(a_{2j})$ and its child for $j \in [\frac{k}{2}]$.

Lemma 7: Let N be a binary tree-based network, and let \hat{C} be a crown of length $k \geq 4$. There exists exactly 2 distinct ways of covering the arcs of \hat{C} and the arcs c_{2j} for $j \in [\frac{k}{2}]$ using reticulated cherry shapes.

Proof: As in the proofs of the previous two lemmas, we use the fact that exactly one incoming arc of every reticulation is covered as a middle arc of a reticulated cherry shape. Each reticulated cherry shape consists of 3 arcs. The crown, together with the arcs c_{2j} , has exactly $3k/2$ arcs in total; the crown contains exactly $k/2$ reticulations. It follows that every arc of \hat{C} and the arcs c_{2j} for each j must be covered by a reticulated cherry shape. There are two ways to do this, by considering either of the covers

$$\left\{ c_{2j}, a_{2j+1}, a_{2j+2} : j \in \left[\frac{k-2}{2} \right] \right\} \cup \{c_k, a_1, a_2\}$$

or

$$\{ \{c_{2j}, a_{2j}, a_{2j-1}\} : j \in [k/2] \}.$$

No other ways of using reticulated cherry shapes to cover the $3k/2$ arcs exist. Therefore, there are exactly 2 distinct ways of covering the arcs of \hat{C} and the arcs c_{2j} for $j \in [k/2]$. \square

Theorem 8: Let N be a binary tree-based network. Let $\hat{M}_1, \hat{M}_2, \dots, \hat{M}_m$ denote the M -fences of N of length $k_i \geq 4$ where $i \in [m]$ and let $\hat{C}_1, \hat{C}_2, \dots, \hat{C}_c$ denote the crowns of N . Then N has $2^{c-m} k_1 k_2 \dots k_m$ distinct cherry covers and support trees.

Proof: We can partition every cherry cover so that each block corresponds to a maximal zig-zag trail, together with outgoing arcs of the reticulations of the trail. This follows from the fact that every arc with a reticulation tail must be covered by a reticulated cherry shape. By Lemmas 5, 6, and 7, we know that each N -fence of length at least 3, M -fence of length k at least 4, and crown (together with outgoing arcs of reticulations in each zig-zag trail) can be covered in 1, $k/2$, and 2 ways. The permutations are counted locally within each zig-zag trail, and they have no effect on one another. It follows that there are

$$2^c \times \prod_{i=1}^m \frac{k_i}{2} = 2^{c-m} k_1 k_2 \dots k_m$$

many possible cherry covers. By Proposition 4, this is also the number of support trees for N . \square

V. EXTENDING FENCES TO NON-BINARY NETWORKS

The goal of this section is to generalize the existence of a unique fence decomposition from binary networks (Theorem 2) to non-binary networks.

We start by extending the definitions of fences and crowns, introduced in Section II-C, to non-binary networks. In extending crowns, we shall glue together a crown to a W -fence, an N -fence, or to another crown. Roughly speaking, one can view non-binary crowns as structures that contain a binary crown. In particular, a binary crown will also be considered to be a non-binary crown. For the W , N , M -fences, we impose that they do not contain any binary crowns. In the binary case, if a zig-zag trail contains a crown, then the degree constraints on the vertices imply that the entire trail must be a crown. Thus, the extra crown-less condition over the W , N , M -fences for the non-binary case is not a more restrictive condition on the binary case. Consequently, this definition is also appropriate for binary structures.

Let N be non-binary; we introduce five more zig-zag trails. Let $(a_1, a_2, \dots, a_k, c_1, c_2, \dots, c_{2k'})$ be a zig-zag trail, of different arcs, where $(c_1, \dots, c_{2k'})$ is a crown. We say that such a path is a

- *Lower (Upper) W-crown:* $k \geq 1$ is odd ($k \geq 2$ is even), (a_1, \dots, a_k) does not contain a crown, $\text{head}(a_k) = \text{head}(c_1)$ ($\text{tail}(a_k) = \text{tail}(c_1)$).
- *Lower (Upper) N-crown:* $k \geq 2$ is even ($k \geq 1$ is odd), (a_1, \dots, a_k) does not contain a crown, $\text{head}(a_k) = \text{head}(c_1)$ ($\text{tail}(a_k) = \text{tail}(c_1)$).
- *Double-crown:* There exists a $j < k$ such that (a_1, a_2, \dots, a_j) forms a crown and (a_{j+1}, \dots, a_k) does not contain a crown.

As in the binary case, we define endpoints of the new structures. Formally, if A is an upper/lower W -crown, then $\text{End}(A) = \{\text{tail}(a_1)\}$, and if A is an upper/lower N -crown, then $\text{End}(A) = \{\text{head}(a_1)\}$. Fig. 3 shows some examples of all the different types of zig-zag trails that we can find in non-binary networks.

We now extend the notion of zig-zag trails to non-binary networks. Specifically, we generalize the concept of maximal zig-zag trails by considering them maximal within one of the previously defined classes, that is, maximal W , N , M -fences, crowns, double-crowns and lower and upper N , W -crowns. This means that, in the non-binary case, a maximal zig-zag trail may be contained within another zig-zag trail, as long as it does not belong to any of the defined classes. Another difference between maximal zig-zag trails in binary and those in non-binary networks is that two maximal zig-zag trails can overlap at arcs in the latter. Let N be a non-binary network, and let A, B be two maximal zig-zag trails of N . We say that A and B *intersect* if they share at least one common arc. Let $\mathcal{F}(N)$ denote the set of all maximal zig-zag trails of N . We call a subset F of $\mathcal{F}(N)$ an *extended zig-zag trail* if for any element $A \in F$, either $F = \{A\}$, or for any other element $B \in F$, there exists a sequence $A = A_1, \dots, A_k = B$ of maximal zig-zag trails in F , such that A_i intersects with A_{i+1} for $i \in [k-1]$. Such a set F is called *maximal* if there is no other extended zig-zag trail G such that $F \subsetneq G$. We call the set of all maximal extended zig-zag trails an *extended fence decomposition*.

Let F be an extended zig-zag trail. We denote by $V(F)$ and $A(F)$ the set of vertices and arcs of F , respectively. By considering an extended fence decomposition D of the network, it is clear that each arc must be covered by exactly one maximal extended zig-zag trail $F \in D$. Let $G \subseteq F$. We say that G *covers* F , or that G is a *covering* of F , if $A(G) = A(F)$. Fig. 4 shows an example of an extended zig-zag trail together with two of its possible coverings.

Additionally, let $F = \{A_1, \dots, A_k, C_1, \dots, C_\ell\}$, where C_i represents all the crowns and double-crowns in F , and A_i represent the other trails. We denote by $\text{End}(F)$ the set of endpoints of each A_i . We call this set the set of *endpoints* of F .

Proposition 9: The extended fence decomposition of a non-binary network $N = (V, A)$ induces a partition of A .

Proof: Let $N = (V, A)$ be a non-binary network and let D be its extended fence decomposition.

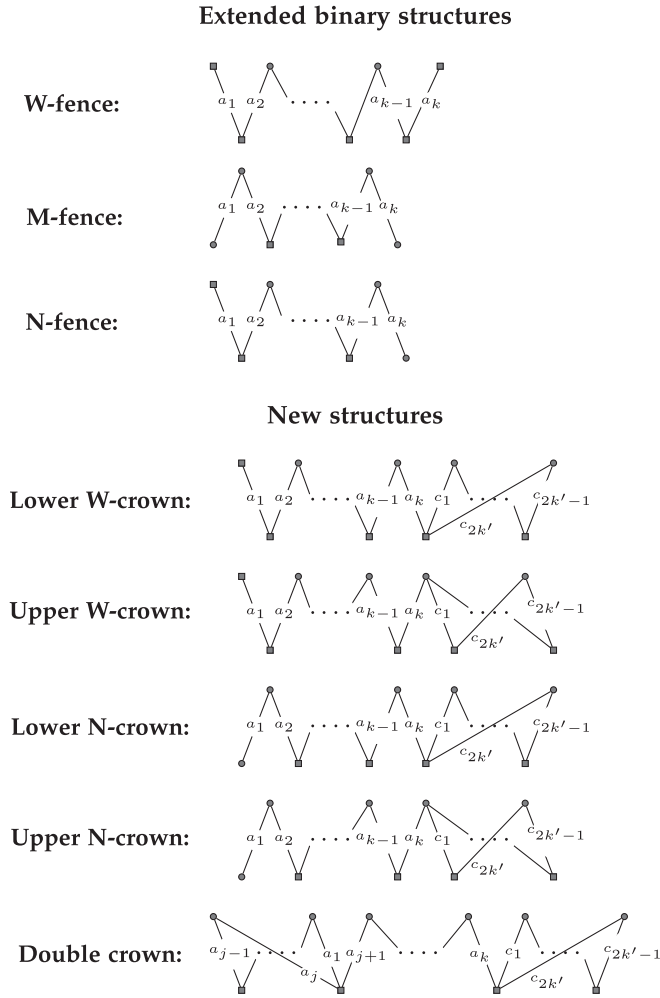


Fig. 3. Example of the new fences introduced in the non-binary case. The upper part illustrates the extension of binary structures to the non-binary case. The additional condition on trails ensures that these structures replicate those from the binary case, preventing crowns from forming within the trail. The lower part presents the new structures. Note that for the lower and upper W -crowns, we have that $\text{tail}(a_1)$ is a reticulation. For the lower and upper N -crowns, we have that $\text{head}(a_1)$ is a tree node. For the double crown, there are four possibilities to connect the two binary crowns. Here, we show one case, where the binary crowns are connected at the reticulations.

First, we show that any two distinct maximal extended zig-zag trails $F, F' \in D$ must be disjoint. Suppose, for contradiction, that there exists a maximal zig-zag trail $B \in F \cap F'$. Let $B' \in F$. By the definition of an extended zig-zag trail, there exists a sequence of maximal zig-zag trails $B = A_1, \dots, A_k = B'$ such that A_i intersects A_{i+1} for all $i \in [k-1]$. Since $A_1 = B \in F'$ and each A_i intersects A_{i+1} , it follows by the maximality of F' that every $A_i \in F'$. In particular, $B' \in F'$. Similarly, for every $C \in F'$, we necessarily have $C \in F$, which implies $F = F'$, contradicting the assumption that F and F' are distinct. Thus, different maximal extended zig-zag trails in D are disjoint.

Next, we prove that each arc of N is contained in exactly one extended zig-zag trail. Let $a = u_0 v_0 \in A$. We will show that a is contained in at least one maximal zig-zag trail. To do this, we construct a maximal zig-zag trail by extending the trail in both directions from u_0 and v_0 until no further extension is possible.

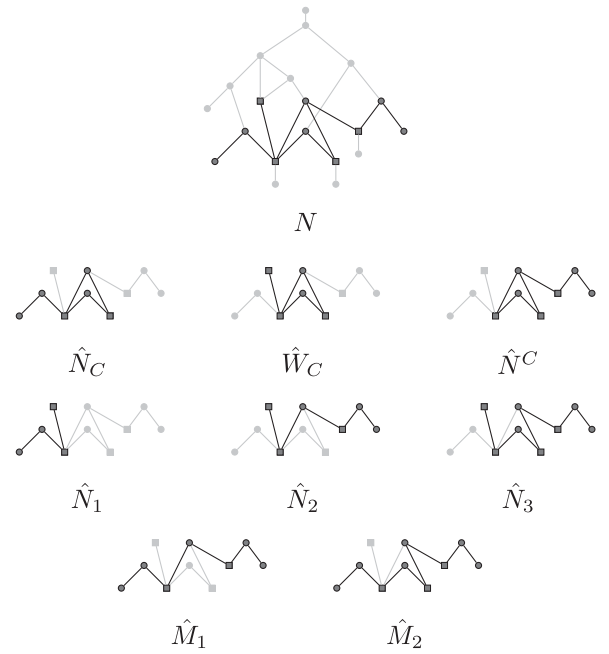


Fig. 4. Example of a non-binary network N and a non-trivial maximal extended zig-zag trail $F = \{\hat{N}_C, \hat{W}_C, \hat{N}^C, \hat{N}_1, \hat{N}_2, \hat{N}_3, \hat{M}_1, \hat{M}_2\}$ of N . The highlighted edges represent different zig-zag trails: \hat{N}_1, \hat{N}_2 and \hat{N}_3 are N -fences; \hat{M}_1 and \hat{M}_2 are M -fences; \hat{N}^C is an upper N -crown; \hat{N}_C is a lower N -crown; and \hat{W}_C is a lower W -crown. The sets of maximal zig-zag trails $G_1 = \{\hat{N}_1, \hat{M}_1, \hat{M}_2\}$ and $G_2 = \{\hat{N}_C, \hat{W}_C, \hat{N}^C\}$ represent two different coverings of F . Observe that \hat{N}_1 and \hat{N}^C do not share any edges, but they belong to the same extended zig-zag trail F since \hat{N}_1 and \hat{M}_1 intersect, and \hat{M}_1 and \hat{N}^C intersect.

First, let us assume that a is not contained in any crown of N . In this case, we consider a zig-zag trail of the form

$$(b_k, b_{k-1}, \dots, b_1, b_0 = a = a_0, a_1, \dots, a_{k'})$$

where $u_0 = \text{tail}(b_1)$ and $v_0 = \text{head}(a_1)$, and such that no further edge can be added to either end of the trail without violating the conditions of a zig-zag trail. Depending on the parity of k and k' , the constructed trail represents a maximal W , M , or N -fence, determined by whether the endpoints of the trail are reticulations or tree nodes.

Now, let us assume that a is contained in some crown. Let $C = (c_1, \dots, c_{2k})$ be a crown with $a = c_i$ for some $i \in [2k]$. If no arc $a_1 \in A \setminus A(C)$ exists such that $\text{head}(a_1) \in \{\text{head}(c_i) : i \in [2k]\}$ or $\text{tail}(a_1) \in \{\text{tail}(c_i) : i \in [2k]\}$, then C itself is a maximal zig-zag trail. Otherwise, assume that exists $a_1 \in A \setminus A(C)$ such that either $\text{head}(a_1) = \text{head}(c_1)$ or $\text{tail}(a_1) = \text{tail}(c_1)$ (rearranging the edges of C , if necessary).

Consider the case where $\text{head}(a_1) = \text{head}(c_1)$. Then, we can extend the trail to a zig-zag trail of the form

$$(a_{k'}, a_{k'-1}, \dots, a_1, c_1, \dots, c_{2k}),$$

where all the edges are distinct, no additional edge can be added to $a_{k'}$, and the trail $(a_{k'}, a_{k'-1}, \dots, a_1)$ contains at most one crown. If it contains no crown, the constructed zig-zag trail is a maximal lower N -crown (if k' is even) or a maximal lower W -crown (if k' is odd). If it contains a crown, let us denote it by

say $(a_i, a_{i-1}, \dots, a_{i-2l})$, then the zig-zag trail

$$(a_i, a_{i-1}, \dots, a_{i-2l}, a_{i-2l-1}, \dots, a_1, c_1, \dots, c_{2k})$$

is a maximal double crown.

A similar argument applies when $\text{tail}(a_1) = \text{tail}(c_1)$, yielding maximal upper N -crowns, upper W -crowns, or double crowns.

Thus, every arc of the network is contained in at least one maximal zig-zag trail, which, as shown earlier, belongs to a unique maximal extended zig-zag trail. This completes the proof. \square

By construction, we have that extended zig-zag trails are connected, so we have zig-zag trails connecting any two nodes of the trail. The following result states that we do not only have paths between any two nodes of the fence, but we have zig-zag trails, which can be extended to maximal zig-zag trails.

Lemma 10: Let N be a non-binary network and let F be a maximal extended zig-zag trail in N and $u, v \in V(F)$. Then there exists a zig-zag trail C (not necessarily maximal) with all its arcs in $A(F)$, such that $\text{End}(C) = \{u, v\}$.

Proof: Let A and B be two maximal zig-zag trails in F such that u is a node of A and v is a node of B . By construction of F , there exists a sequence $A = A_1, A_2, \dots, A_k = B$ of maximal zig-zag trails of F , with $k \geq 2$, such that A_i and A_{i+1} are intersecting for every i . We are going to construct a maximal zig-zag trail with arcs in $A(F)$, including u and v . Let d_i denote one of the arcs in the intersection of A_i and A_{i+1} . Now, we define maximal zig-zag trails C_i as follows:

- 1) $C_1 = A_1$.
- 2) For $1 < i < k$:
 - Denote $C_{i-1} = (c_1^{i-1}, \dots, c_{m_{i-1}}^{i-1})$ and $A_i = (a_1^i, \dots, a_{n_i}^i)$ and let $j \in [m_{i-1}]$ such that $c_j^{i-1} = a_l^i$ for some $l \in [n_i]$. Suppose that $u \in \{\text{tail}(c_h^{i-1}), \text{head}(c_h^{i-1})\}_{h=1}^j$ and $d_i \in \{a_h^i\}_{h=l}^{n_i}$, otherwise, rearrange the trails to be in this situation. Then, we define

$$C_i = (c_1^{i-1}, \dots, c_j^{i-1}, a_{l+1}^i, a_{l+2}^i, \dots, a_{n_i}^i).$$

- 3) Denote $C_{k-1} = (c_1^{k-1}, \dots, c_{m_{k-1}}^{k-1})$. We have $d_{k-1} \in A(C_{k-1}) \cap A(A_k)$, so let us denote $d_k = c_p^{k-1} = a_q^k$. Assume that $u \in \{\text{tail}(c_h^{k-1}), \text{head}(c_h^{k-1})\}_{h=1}^p$ and $v \in \{\text{tail}(a_h^k), \text{head}(a_h^k)\}_{h=q}^{n_k}$, otherwise, rearrange the trails. Thus, define

$$C_k = (c_1^{k-1}, \dots, c_p^{k-1}, a_{q+1}^k, \dots, a_{n_k}^k)$$

Clearly, each C_i is a maximal zig-zag trail containing u that intersects A_{i+1} . Additionally, we can easily see the maximality of C_i from the maximality of C_{i-1} and A_i . Thus, C_k is a maximal zig-zag trail containing u and v with all its arcs in F . From C_k , we can extract a zig-zag trail from u to v with all its arcs in F . \square

It is worth noting that, from the construction given in the proof of Lemma 10, we can deduce that, if two disjoint fences A and B are present within the same extended zig-zag trail F , we can identify a new fence C intersecting A and B , with any two selected vertices from $V(A)$ and $V(B)$. Furthermore, if we select u and v to be endpoints of F , from the proof of Lemma 10,

the resulting extended trail C is maximal. Accordingly, we state the following result.

Corollary 11: Let N be a non-binary network and let F be a maximal extended zig-zag trail in N and $u, v \in \text{End}(F)$. Then, there exists $C \in F$ with $\text{End}(C) = \{u, v\}$.

VI. CHARACTERIZATION OF SEMI-BINARY TREE-BASED NETWORKS

In this section, we extend the results of counting the number of cherry covers (and then of support trees), Theorem 8, and the characterization of tree-based networks in terms of forbidden structures, Lemma 3, from binary to semi-binary tree-based networks.

Proposition 9 states that the extended fence decomposition partitions the set of arcs of any non-binary network. Consequently, this decomposition also applies to semi-binary networks.

Now, our focus is to study the forbidden structures in semi-binary tree-based networks with respect to their extended fence decomposition. Due to the semi-binary property, in particular since the outdegree of tree nodes is 2, it is clear that the network cannot exhibit upper N -crowns or upper W -crowns. Additionally, we prove the absence of W -fences, lower W -crowns and double-crowns.

Lemma 12: Let N be a semi-binary tree-based network, and let D be its extended fence decomposition. For all $F \in D$, F satisfies the following conditions:

- a) F does not contain any W -fences.
- b) F does not contain any lower W -crowns.
- c) F has at most 1 crown. In particular, F does not contain any double-crowns.

Proof: Let us start by proving **a)**. Let N be a semi-binary tree-based network, and suppose there exists a W -fence (a_1, \dots, a_{2l}) in a maximal zig-zag trail $F \in D$. Since N is tree-based, it has some underlying support tree T as the result of iteratively deleting all but one incoming arc of each reticulation, ensuring that no new leaves are added at any step.

Observe that $\text{tail}(a_1)$ and $\text{head}(a_1)$ are two connected reticulations, then $a_1 \in E(T)$. Consequently, $a_2 \notin E(T)$. As $\text{tail}(a_2)$ is a tree node with outdegree 2, we have $a_3 \in E(T)$; otherwise $\text{tail}(a_2)$ corresponds to a (different) leaf on T . Repeating this reasoning, we conclude that $a_{2l-1} \in E(T)$ and $a_{2l} \notin E(T)$. This selection of arcs implies that the reticulation $\text{tail}(a_{2l})$ corresponds to a (different) leaf in the support tree which is a contradiction.

An analogous constructive argument can be used to prove **b)**.

Finally, let us show condition **c)**. Let N be a semi-binary tree-based network, and suppose there exists an extended zig-zag trail $F \in D$ with more than 1 crown. Let (a_1, \dots, a_n) and (b_1, \dots, b_m) be two different crowns. Notice that n and m are even integers. Let us assume that $\text{tail}(a_1) = \text{tail}(a_n)$ and $\text{tail}(b_1) = \text{tail}(b_m)$, if not, we can rename the arcs to make sure that this is the case. We distinguish the cases where both paths either intersect in a reticulation or they do not. Let T be a support tree of N .

Firstly, assume they intersect in a reticulation. Then there exists $i \in [n]$ and $j \in [m]$ such that $\text{head}(a_i) = \text{head}(a_{i+1}) = \text{head}(b_j) = \text{head}(b_{j+1})$. Observe that either $a_i \notin E(T)$ and $a_{i+1} \notin E(T)$, or $b_j \notin E(T)$ and $b_{j+1} \notin E(T)$. Without loss of generality, assume that $a_i \notin E(T)$ and $a_{i+1} \notin E(T)$. Since $a_i \notin E(T)$, by the semi-binary property, we have $a_{i-1} \in E(T)$, implying $a_{i-2} \notin E(T)$, and this process can be iterated. Repeating the same argument for $a_{i+1} \notin E(T)$, we get $a_{i+2} \in E(T)$, $a_{i+3} \notin E(T)$, etc. Since n is even, we end up having $a_1 \notin E(T)$ and $a_n \notin E(T)$. However, due to $\text{tail}(a_1) = \text{tail}(a_n)$, this makes $\text{tail}(a_1)$ a leaf of T , contradicting the fact that it is a support tree of N .

Now, assume that we have two binary crowns (a_1, \dots, a_n) and (b_1, \dots, b_m) that are disjoint in their reticulations. By Lemma 10, there exists a zig-zag trail (c_1, \dots, c_l) that joins both structures, such that $c_i \neq a_j$ and $c_i \neq b_k$ for any possible index combinations i, j, k . Due to the semi-binary condition, $\text{head}(c_1)$ and $\text{head}(c_l)$ are reticulations, then l is an even number. By relabelling arcs, we can consider the zig-zag trail $(a_2, a_3, \dots, a_n, a_1, c_1, \dots, c_l, b_2, b_3, \dots, b_m, b_1)$ where $\text{head}(a_1) = \text{head}(a_2) = \text{head}(c_1)$, $\text{head}(b_1) = \text{head}(b_2) = \text{head}(c_l)$, $\text{tail}(a_1) = \text{tail}(a_n)$ and $\text{tail}(b_1) = \text{tail}(b_m)$.

If $a_1 \notin E(T)$ and $a_2 \notin E(T)$, then applying the same argument as in the previous case we reach a contradiction. Therefore, we either have $a_1 \in E(T)$ or $a_2 \in E(T)$. In that case, $c_1 \notin E(T)$, and since l is even, $c_l \in E(T)$, implying $b_1 \notin E(T)$ and $b_2 \notin E(T)$. Again, as in the previous case this leads to a contradiction.

Finally, since (by definition) double-crowns contain two binary crowns, then F does not contain those structures. \square

Lemma 12 restrict the possible structures that we can find in the fence decomposition of a tree-based network, so we can classify the maximal extended zig-zag trails that we can find in semi-binary tree-based networks.

Let N be a semi-binary tree-based network, D its extended fence decomposition and let $F \in D$. We say that:

- F is a *generalized N -fence* if F does not have crowns and $\text{End}(F)$ contains exactly one reticulation.
- F is a *generalized M -fence* if F does not have crowns and $\text{End}(F)$ contains only tree vertices.
- F is a *generalized crown* if F has one crown and $\text{End}(F)$ is either empty or it only contains tree vertices.

A reinterpretation of Lemma 12 shows that the previous definitions classify the maximal extended zig-zag trails that we can find in semi-binary tree-based networks.

Corollary 13: Let N be a semi-binary tree-based network, D its extended fence decomposition and let $F \in D$. Then, one of the following is satisfied:

- 1) F is a generalized N -fence.
- 2) F is a generalized M -fence.
- 3) F is a generalized crown.

Proof: By Lemma 12 c), we have that F either has one crown or no crowns. Now, suppose that $\text{End}(F)$ has at least 2 reticulations u and v . By Corollary 11, we have a maximal zig-zag trail (a_1, \dots, a_k) that goes from u to v . If this trail does not contain crowns, then it is a maximal W -fence, which,

by Lemma 12 a), we know that are prohibited in tree-based networks. If the trail has a crown, we can consider $k' \in [k]$ to be the smallest integer such that $(a_i, a_{i+1}, \dots, a_{2k'})$ is a crown for some i . Thus, the zig-zag trail $(a_1, \dots, a_{2k'})$ is a maximal W -crown, which, by Lemma 12 b) are also prohibited. If F has a crown and $\text{End}(F)$ has one reticulation, a similar argument leads us to F having a maximal W -crown. In other case, F satisfy (1), (2) or (3).

Now that we have described the structures that appear in semi-binary tree-based networks, our objective is to characterize tree-based networks and to count support trees. Recall that, to construct a support tree T , we have to select one reticulation arc from each reticulated cherry shape of the network, in a way that we do not create new leaves. Since we have to make a selection of arcs, we would not need the whole information of each extended zig-zag trail. Instead, we will work with covers. The following results will give us minimal coverings (w.r.t. the inclusion), for each extended zig-zag trail, that will be useful to count support trees in the latter.

Corollary 14: Let F be a generalized N -fence with $\text{End}(F) = \{u, v_1, \dots, v_k\}$, where u is the only reticulation endpoint and v_i are tree-vertices. Let $G = \bigcup_{i=1}^k \{\hat{N}_i\}$, where \hat{N}_i are maximal N -fences in F going from u to v_i . Then, G is a covering of F .

Proof: By Corollary 11, there exists zig-zag trails \hat{N}_i from u to v_i contained in F . Since F does not have crowns, these trails do not contain crowns and must be unique. Thus, \hat{N}_i are maximal N -fences in F . Now let us prove that $G = \bigcup_{i=1}^k \{\hat{N}_i\}$ covers F . We have to prove that $A(G) = A(F)$. Clearly, we have $A(G) \subseteq A(F)$, so we need to show the other inclusion. Let $a \in A(F)$. By construction of F , there exists a maximal zig-zag trail A such that a is an arc of A , so let us denote $A = (a_1, \dots, a_{i-1}, a, a_{i+1}, \dots, a_m)$. If $\text{tail}(a_1) = u$ and $\text{head}(a_m) = v_j$ or $\text{head}(a_1) = v_j$ and $\text{tail}(a_m) = u$, for some $j \in [k]$, then A is either \hat{N}_j or \hat{N}_j inverted. In both cases we have $a \in A(\hat{N}_j)$. Else, we have $\text{head}(a_1) = v_j$ and $\text{head}(a_m) = v_l$, for some $j, l \in [k]$. In this case, it is easy to see that a must be either an arc in \hat{N}_j or in \hat{N}_l . Therefore, $a \in A(G)$. \square

Corollary 15: Let F be a generalized M -fence with $\text{End}(F) = \{v_1, \dots, v_k\}$, where v_i are tree-vertices. Let $G = \bigcup_{i=1}^k \{\hat{M}_i\}$, where \hat{M}_i are maximal M -fences in F going from v_1 to v_i . Then, G is a covering of F .

Proof: Analogous proof to the one presented in Corollary 14. \square

Corollary 16: Let F be a generalized crown with $\text{End}(F) = \{v_1, \dots, v_k\}$, where v_i are tree-vertices. Let $G = \{C\} \cup \bigcup_{i=1}^k \{\hat{N}_i^c\}$, where C is the only crown of F and \hat{N}_i^c are maximal lower N -crowns in F starting at v_i . Then, G is a covering of F .

Proof: Analogous proof to the one presented in Corollary 14. \square

Now, we have concrete coverings for each extended zig-zag trail that appears in semi-binary tree-based networks. For each extended zig-zag trail F , and G constructed in Corollaries 14, 15 and 16, depending on the nature of F , we will refer to G as an *efficient covering* of F . We will use efficient coverings to construct support trees.

For a covering $G = \bigcup_{i=1}^k \{A_i\}$, with $A_i = (a_1^i, a_2^i, \dots, a_{m_i}^i)$, we will denote by d_j^i the child, if it exists, of $\text{head}(a_j^i)$ for $j \in [m_i]$. Observe that, in some cases, we may have $d_i = d_{i+1}$, depending on the parity of m_i . When this occurs, we treat the situation by using only one of the two labels, choosing appropriately based on the parity of the index. The multiplicity of each arc $c_j^i := \text{head}(a_j^i)d_j^i$, in the bulged version, is equal to the indegree of $\text{head}(a_j^i)$ minus one. In all cherry covers, each of these parallel arcs will be covered in a reticulated cherry shape as the lowest arc. Furthermore, no two parallel arcs appear in the same reticulated cherry shape. The symmetry of each parallel arc means that the order in which they are covered is irrelevant in a cherry cover. Thus, we allow for the arc c_j^i to appear in multiple reticulated cherry shapes. It is to be understood, that each appearance of such arcs c_j^i refers to different versions of the same parallel arc.

Consider an efficient covering $G = \bigcup_{i=1}^k \{A_i\}$ of a generalized M or N fence F , where each of the $A_i = (a_1^i, a_2^i, \dots, a_{m_i}^i)$ start at the same node, so that $a_1^1 = a_1^2 = \dots = a_1^k$. Since every pair of maximal fences A_i has different endpoints, no maximal fence is properly contained in another maximal fence. This means that every maximal fence A_i ‘splits’ off from the other fences at some point. In particular, each fence A_i splits at some point from the previous fences A_j with $j < i$. Since F does not have any crowns, once the fence splits, it ends in a branch structure that does not intersect the previous trails again. We define the *split number* of a zig-zag trail A_i as $s_G^{A_i} = \min\{j \in \{1, \dots, m_i\} \mid a_j^i \notin A(\bigcup_{l=1}^{i-1} \{A_l\})\}$.

Lemma 17: Let N be a semi-binary tree-based network. Let F be a generalized N -fence of N and $G = \bigcup_{i=1}^k \{\hat{N}_i\}$ its efficient covering, with $\hat{N}_i = (a_1^i, a_2^i, \dots, a_{k_i}^i)$. Then, for $i \in [k]$ and $j \in [\frac{k_i-1}{2}]$, every cherry cover of N contains the reticulated cherry shapes $\{c_{2j-1}^i, a_{2j}^i, a_{2j+1}^i\}$.

Proof: The arc c_1^i must be covered as the bottom arc of a reticulated cherry shape. Since a_1^i is incident to two reticulations, it cannot be covered as a middle arc. This means that all other incoming arcs of $\text{tail}(c_1^i)$ must be covered as a middle arc, together with c_1^i , i.e., the shape must be of the form $\{c_1^i, a_2^i, a_3^i\}$. Since the reticulated cherry shapes contain a_3^i , this means in particular that $\text{tail}(c_3^i)$ has one incoming arc a_3^i that is already covered. Then, all other incoming arcs of $\text{tail}(c_3^i)$ must be covered as a middle arc. That is, we must cover the parallel arcs c_3^i with reticulated cherry shapes $\{c_3^i, a_4^i, a_5^i\}$. This argument can be repeated for all arcs c_{2j-1}^i for $i \in [k]$ and $j \in [\frac{k_i-1}{2}]$ to obtain a covering of F and all outgoing arcs, and we are done. \square

We define the *length* of a generalized M -fence as the number of arcs in F , denoted by k_F . For an efficient covering $G = \bigcup_{i=1}^k \{\hat{M}_i\}$ of F , let k_i denote the length of each \hat{M}_i and $s_G^{\hat{M}_i}$ its split number with respect to G . It is clear that $k_F = k_1 + \sum_{i=2}^k (k_i - (s_G^{\hat{M}_i} - 1))$.

Lemma 18: Let N be a semi-binary tree-based network, and let F be a generalized M -fence of length $k_F \geq 4$. Let $G = \bigcup_{i=1}^k \{\hat{M}_i\}$ be an efficient covering of F for some designated vertex v , with $\hat{M}_i = (a_1^i, \dots, a_{k_i}^i)$ such that $\text{head}(a_1^i) = v$, for all i . There exists exactly $k_F/2$ distinct ways of covering the arcs

of F and the arcs c_{2j}^i for $i \in [k]$ and $j \in [\frac{k_i-2}{2}]$ using cherry and reticulated cherry shapes.

Proof: The generalized M -fence F contains $k_F/2 - k$ reticulations and $k_F - (k+1)$ reticulation arcs, so we must have exactly $k_F/2 - 1$ reticulated cherry shapes covering the arcs of F and the outgoing arcs of the reticulations. Thus, $k_F - 2$ arcs of F are covered by reticulated cherry shapes. The remaining two arcs of F must be arcs a_m^i and a_n^j for some $i, j \in [k]$ and $m \in [k_i]$ and $n \in [k_j]$. Without loss of generality assume that $n > m$. By a similar argument to the one given in Lemma 6, we can see that $i = j$ and $n = m + 1$.

Therefore, every generalized M -fence of length $k_F \geq 4$ must be covered by a single cherry shape and $k_F/2 - 1$ reticulated cherry shapes in every cherry cover. In particular, note that the choice of arcs for the cherry shape determines the rest of the reticulated cherry shapes that cover the rest of the arcs of F . Since there are $k_F/2$ potential locations for cherry shapes, we conclude that there are $k_F/2$ distinct ways of covering the arcs of F and the arcs $\text{head}(a_{2j}^i)c_{2j}^i$ for $i \in [k]$ and $j \in [\frac{k_i-2}{2}]$ using cherry and reticulated cherry shapes. \square

Lemma 19: Let N be a semi-binary tree-based network, and let F be a generalized crown. Let $G = \bigcup_{i=1}^k \{\hat{N}_c^i\} \cup \{C\}$ be an efficient covering of F , where $C = (b_1, \dots, b_{2m_0})$ is a crown and $\hat{N}_c^i = (a_1^i, \dots, a_{m_i}^i, b_{\sigma_i(1)}, \dots, b_{\sigma_i(2m_0)})$ are the lower N -crowns, for some cyclic permutations $\sigma_i : [2m_0] \rightarrow [2m_0]$. Also, let $c_{2\ell}$ denote the arc between $\text{head}(b_{2\ell})$ and its child, and c_{2j}^i the arc between $\text{head}(a_{2j}^i)$ and its child. Then, there exist exactly 2 distinct ways of covering the arcs of F , the arcs $c_{2\ell}$ for $\ell \in [m_0]$ and the arcs c_{2j}^i for $j \in [\frac{m_i}{2}]$ and $i \in [k]$ using reticulated cherry shapes.

Proof: By the semi-binary property and the fact that for each reticulation, only one of its incoming arcs is not covered as the middle arc of a reticulated cherry shape, we have that the arcs of C and the arcs $c_2, c_4, \dots, c_{2m_0}$ can be covered in two different ways in the bulged version of N , as we had in the binary case.

Once the arcs of the crown are covered, the only arcs left to cover are the arcs $a_1^i, \dots, a_{m_i}^i$ and the arcs c_{2j}^i for $j \in [\frac{m_i}{2}]$ and $i \in [k]$. But since \hat{N}_c^i are lower N -crowns, we have $\text{head}(a_{m_i}^i) = \text{head}(b_{\sigma_i(1)}) = \text{head}(b_{\sigma_i(2m_0)})$, and we have that either $b_{\sigma_i(1)}$ or $b_{\sigma_i(2m_0)}$ are covered in a reticulated cherry shape containing $c_{\sigma_i(2m_0)}$. Thus, the cherry cover contains the reticulated cherry shape $\{c_{\sigma_i(2m_0)}, a_{m_i}^i, a_{m_i-1}^i\}$. Now we have $a_{m_i-1}^i$ covered in a shape that does not contain $c_{m_i-2}^i$. Because of this, every reticulation arc incident to $\text{tail}(c_{m_i-2}^i)$ that is not $a_{m_i-1}^i$ must be covered as a reticulated cherry shape together with $c_{m_i-2}^i$. Therefore, we have a reticulated cherry shape of the form $\{c_{m_i-2}^i, a_{m_i-2}^i, a_{m_i-3}^i\}$. Note that m_i is even, since \hat{N}_c^i is a lower N -crown. Repeating the argument, we have reticulated cherry shapes of the form $\{c_{m_i-2j}^i, a_{m_i-2j}^i, a_{m_i-2j-1}^i\}$ for $1 \leq j \leq \frac{m_i}{2}$, and this holds for every $i \in [k]$. Observe that with this argument we covered all the arcs of F , as well as the arcs $c_{2\ell}$ for $\ell \in [m_0]$ and the arcs c_{2j}^i for $j \in [\frac{m_i}{2}]$ and $i \in [k]$.

Summing up, we have 2 possibilities of covering the arcs of C , and once they are covered, the rest of the shapes that cover

the arcs of F are fixed. It follows that every cherry cover of N covers the arcs of F in one of two possible ways. \square

Theorem 20: Let N be a semi-binary tree-based network. Let A_1, A_2, \dots, A_m denote the generalized M -fences of N of length $k_i \geq 4$ where $i \in [m]$ and let B_1, B_2, \dots, B_c denote the generalized crowns of N . Then N has $2^{c-m} k_1 k_2 \dots k_m$ distinct cherry covers and support trees.

Proof: Let P be a cherry cover. Let D be the extended fence decomposition of N . By Corollary 13, we know that the only extended zig-zag trails in D are generalized N and M -fences and generalized crowns. This decomposition gives us a partition of the arcs of N into disjoint blocks, each one of them of three different types:

- 1) Generalized N -fences without the only arc connecting two reticulations, and adding the outgoing arcs of the rest of each reticulation.
- 2) Generalized M -fences together with the outgoing arcs of each reticulation.
- 3) Generalized crowns without the arcs connecting reticulations and adding the outgoing arcs of the reticulations.

By Lemma 17, we know that the type 1) structures can be covered in 1 way. By Lemma 18, for a generalized M -fence of length k , we have that type 2) structures can be covered in $k/2$ ways. And, using Lemma 19, type 3) structures can be covered in 2 different ways. It follows that there are

$$2^c \times \prod_{i=1}^m \frac{k_i}{2}$$

many possible cherry covers. \square

Theorem 21: Let N be a semi-binary network, and let D be its extended fence decomposition. Then, N is tree-based if and only if, for all zig-zag fences $F \in D$ in N , the following holds:

- i) F does not contain any W -fence.
- ii) F does not contain any lower W -crown.
- iii) F contains at most 1 binary crown.

Proof: The only if part follows directly from Lemma 12. For only if part, let us assume that N does not have any maximal extended zig-zag trail F satisfying (i), (ii) or (iii). Thus, each of the maximal extended zig-zag trails are generalized N -fences, generalized M -fences or generalized crowns. Let us construct a cherry cover of N . We can partition the arcs of N in its set of maximal extended zig-zag trails, which gives another partition in blocks of the three different types:

- 1) Generalized N -fences without the only arc connecting two reticulations, and adding the outgoing arcs of the rest of each reticulation.
- 2) Generalized M -fences together with the outgoing arcs of each reticulation.
- 3) Generalized crowns without the arcs connecting reticulations and adding the outgoing arcs of the reticulations.

If we can cover each block with cherry shapes and reticulated cherry shapes, that would give us a cherry cover for the whole network.

Therefore, let us consider, for each block of type 1), the set of shapes and reticulated cherry shapes defined in Lemma 17, and for the structures of type 2) and 3), one of the sets defined in the

proof of Lemmas 18 and 19, respectively. Clearly, this gives us a proper covering of the blocks in cherry shapes and reticulated cherry shapes, so the union of all such sets gives us a covering for the whole network in disjoint shapes. This is a cherry cover for N , and by Theorem 1, N is tree-based. \square

VII. CHARACTERIZATION OF SEMI-BINARY TREE-CHILD NETWORKS

Recall that \mathcal{P}_N denotes the set of all cherry covers of a network N . In this section, we characterize semi-binary tree-child networks [8] in terms of $|\mathcal{P}_N|$.

Lemma 22: Let N be a semi-binary network with k reticulations with indegrees r_1, \dots, r_k , respectively. Then, $|\mathcal{P}_N| \leq \prod_{i=1}^k r_i$.

Proof: Observe that, in a cherry cover, once we select the reticulated cherry shapes, the rest of the cherry cover is fixed. Moreover, in the semi-binary case, reticulated cherry shapes are determined by its middle arc. So we need to count the possible selections for this middle arc. For a reticulation vertex u with indeg $u = d$, there must be $d - 1$ reticulation arcs incident to u covered as the middle arc of a reticulated cherry shape, and only 1 arc without this property. Now, if we denote by $\{u_i\}_{i=1}^k$ and $\{r_i\}_{i=1}^k$, the set of reticulations and its indegrees, respectively, we have r_i possibilities to select the only arc, incident to u_i , that would not be covered as a middle arc of a reticulated cherry shape in a cherry cover. Thus, we have $\prod_{i=1}^k r_i$ possible ways to select such arcs, and therefore at most this many cherry covers. Note that for some combination of reticulation arc selections, it is possible for a cherry cover to not exist. \square

Proposition 23: Let N be a semi-binary network with r reticulations. Let r_1, \dots, r_k denote the indegrees of each reticulation. Then, N is tree-child if, and only if, $|\mathcal{P}_N| = \prod_{i=1}^k r_i$.

Proof: First, assume that N is tree-child. Observe that in this case the extended fence decomposition of N only contains M -fences of size 2 and k generalized M -fences of size $2r_i$. By Theorem 20, the result follows.

Now, assume that $|\mathcal{P}_N| = \prod_{i=1}^k r_i$ and suppose that N is not a tree-child network. Then, there exists a non-leaf vertex u such that all its children are reticulations. By the semi-binary property, u either has 1 child (if it is a reticulation) or 2 children (if it is a tree node). Assume that u is a reticulation, and let v be its only reticulation child and h the only child of v .

Consider the graph N' obtained by: subdividing the arc uv with a vertex u' ; adding a new vertex v' ; and adding the arc $u'v'$. It is clear that the result of this operation gives a semi-binary network. Observe that, by the construction of N' , any cherry cover of N determines, over the bulged version of N' , a unique covering of all the arcs but $u'v'$ and $u'v$. Therefore, for any cherry cover $P \in \mathcal{P}_N$, if we add the cherry shape $\{u'v', u'v\}$, we get a cherry cover for the bulged version of N' , so we have $P \cup \{\{u'v', u'v\}\} \in \mathcal{P}_{N'}$. Thus, $|\mathcal{P}_{N'}| \geq |\mathcal{P}_N| = \prod_{i=1}^k r_i$.

Also, observe that over the bulged version of N , we have many parallel arcs between v and h , let a be one of them. Let P be a cherry cover of N . Then, P must cover the arc a in some reticulated cherry shape, let us denote by $\{a, wv, wl\}$, where w is a parent of v such that it is tree

node and l is the child of w different from v . Observe that $w \neq u$ since u is a reticulation. Consider the set $P' = (P \setminus \{\{a, uv, wl\}\}) \cup \{\{a, u'v, u'v'\}\} \cup \{\{wv, wl\}\}$. It is clear that $P' \setminus \{\{a, u'v, u'v'\}\}$ is not a cherry cover of N , since it does not cover the arc a , and P' does not contain intersecting shapes. Therefore, P' represents a different cherry cover than the ones that we could build from \mathcal{P}_N . Noting that the indegrees of the reticulations have not changed, this means that $|\mathcal{P}_{N'}| \geq \prod_{i=1}^k r_i + 1$, contradicting Lemma 22.

If u is a tree node and v and w represent its children, let us construct N'' from N by subdividing the arc uv with a node u' , adding a new vertex v' , and adding an arc $u'v'$. A similar argument as in the previous case leads us to a contradiction. \square

VIII. CONCLUSION

To summarize our contributions, we have shown that the number of cherry covers coincide with the number of support trees for a semi-binary network (Proposition 4), and we showed that for binary tree-based networks, there is a natural bridge between its fence decomposition and its cherry covers (Theorem 8). We have extended the notion of a fence decomposition to the non-binary setting, by introducing new types of fences in Section V. Using this new characterization, every non-binary network can be decomposed into its extended fence decomposition (Proposition 9). With this decomposition, we proved the generalized version of Theorem 8, by showing that one can count the number of cherry covers (and therefore the number of support trees) of a semi-binary tree-based network (Theorem 20). We also gave a new characterization of semi-binary tree-based networks based on forbidden structures within the extended fence decomposition (Theorem 21). Finally, we gave a new characterization of semi-binary tree-child networks based on the number of cherry covers (Proposition 23).

We now discuss potential future directions regarding cherry covers and extended fence decompositions, starting with open problems on generalizations to the non-binary setting. For example, can we show that for non-binary tree-based networks, the number of cherry covers coincide with the number of support trees? At first glance, this does not seem true. To give some justification, recall that within a cherry cover for a non-binary network, each arc may be covered more than once (for the formal definition, see Definition 2.10 of [13]). Consider a non-binary tree T on three leaves $\{x, y, z\}$ with a common parent t . Including the root ρ , this is a tree on 5 vertices $\{\rho, t, x, y, z\}$ and 4 arcs $\{\rho t, tx, ty, tz\}$. By definition, $\{\{tx, ty\}, \{ty, tz\}\}$ and $\{\{tx, ty\}, \{tx, tz\}\}$ are both cherry covers of T . Since T is a tree, it has only one support tree, namely the tree itself. Thus, naively, the equality $|S_T| = |\mathcal{P}_T|$ does not hold. However, we note that there is exactly one *maximal cherry cover* $\{\{tx, ty\}, \{tx, tz\}, \{ty, tz\}\}$ of T , which has maximal size over all possible cherry covers. Therefore, we conjecture the following.

Conjecture 24: Let N be a non-binary network. Let S_N and \mathcal{P}_N denote the set of support trees and maximal cherry covers for N , respectively. Then $|S_T| = |\mathcal{P}_T|$.

In another direction, one can attempt for a forbidden structure characterization for non-binary tree-based networks. Here, the

analysis of the extended fence decompositions quickly becomes more complex, as upper W -crowns and upper N -crowns are not immediately forbidden due to the relaxation of degree constraints.

REFERENCES

- [1] M. Nei and S. Kumar, *Molecular Evolution and Phylogenetics*. Oxford, U.K.: Oxford Univ. Press, 2000.
- [2] D. H. Huson, R. Rupp, and C. Scornavacca, *Phylogenetic Networks: Concepts, Algorithms and Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2010.
- [3] E. Bapteste et al., "Networks: Expanding evolutionary thinking," *Trends Genet.*, vol. 29, no. 8, pp. 439–441, 2013.
- [4] S. Kong, J. C. Pons, L. Kubatko, and K. Wicke, "Classes of explicit phylogenetic networks and their biological and mathematical significance," *J. Math. Biol.*, vol. 84, no. 6, 2022, Art. no. 47.
- [5] V. Daubin, N. A. Moran, and H. Ochman, "Phylogenetics and the cohesion of bacterial genomes," *Science*, vol. 301, no. 5634, pp. 829–832, 2003.
- [6] T. Dagan and W. Martin, "The tree of one percent," *Genome Biol.*, vol. 7, pp. 1–7, 2006.
- [7] A. R. Francis and M. Steel, "Which phylogenetic networks are merely trees with additional arcs?," *Systematic biol.*, vol. 64, no. 5, pp. 768–777, 2015.
- [8] G. Cardona, F. Rosselló, and G. Valiente, "Comparison of tree-child phylogenetic networks," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 6, no. 4, pp. 552–569, 2008.
- [9] C. Semple, "Phylogenetic networks with every embedded phylogenetic tree a base tree," *Bull. Math. Biol.*, vol. 78, no. 1, pp. 132–137, 2016.
- [10] E. Sayyari and S. Mirarab, "Testing for polytomies in phylogenetic species trees using quartet frequencies," *Genes*, vol. 9, no. 3, 2018, Art. no. 132.
- [11] J. Ottenburghs, "Multispecies hybridization in birds," *Avian Res.*, vol. 10, no. 1, 2019, Art. no. 20.
- [12] L. Jetten and L. van Iersel, "Nonbinary tree-based phylogenetic networks," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 15, no. 1, pp. 205–217, Jan./Feb. 2018.
- [13] L. van Iersel, R. Janssen, M. Jones, Y. Murakami, and N. Zeh, "A unifying characterization of tree-based networks and orchard networks using cherry covers," *Adv. Appl. Math.*, vol. 129, 2021, Art. no. 102222.
- [14] L. Zhang, "On tree-based phylogenetic networks," *J. Comput. Biol.*, vol. 23, no. 7, pp. 553–565, 2016.
- [15] M. Hayamizu, "A structure theorem for rooted binary phylogenetic networks and its implications for tree-based networks," *SIAM J. Discrete Math.*, vol. 35, no. 4, pp. 2490–2516, 2021.
- [16] J. C. Pons, C. Semple, and M. Steel, "Tree-based networks: Characterisations, metrics, and support trees," *J. Math. Biol.*, vol. 78, pp. 899–918, 2019.
- [17] L. van Iersel, M. Jones, E. Julien, and Y. Murakami, "Making orchard by adding leaves," 2023, *arXiv:2305.03106*.

Joan Carles Pons received the PhD degree in mathematics from the University of the Balearic Islands (UIB), in 2016. He is currently an associate professor with the Department of Mathematics and Computer Science, UIB and co-PI of the Computational Biology and Bioinformatics Research Group, UIB.

Pau Vives López received the graduated degree in mathematics, in 2022. He is currently working toward the PhD degree with the University of the Balearic Islands, and member of the Computational Biology and Bioinformatics Research Group with the University of the Balearic Islands.

Yukihiro Murakami received the bachelor's and master's degrees in mathematics from Corpus Christi College, Oxford, in 2017, and the PhD degree from the Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, in 2021. He is an assistant professor with the Delft University of Technology.

Leo van Iersel received the MSc degree in applied mathematics from Twente University, the Netherlands, in 2004, and the PhD degree from the Eindhoven University of Technology, in 2009. He is an associate professor with the Delft University of Technology.