Journal of Graph Algorithms and Applications
http://jgaa.info/ vol. 26, no. 1, pp. 81–90 (2022)
DOI: 10.7155/jgaa.00583

Non-binary universal tree-based networks

Mareike Fischer[1] *Michelle Galla*[1] *Kristina Wicke*[2]

[1]Institute of Mathematics and Computer Science, University of Greifswald, Greifswald, Germany
[2]Department of Mathematics, The Ohio State University, Columbus (OH), USA

Submitted:	August 2021	Reviewed:	November 2021	Revised:	December 2021
Accepted:	March 2022	Final:	March 2022	Published:	April 2022
Article type:	Regular paper		Communicated by:	P. Ferragina	

Abstract. A tree-based network N on X is called universal if every phylogenetic tree on X is a base tree for N. Recently, binary universal tree-based networks have attracted great attention in the literature and their existence has been analyzed in various studies. In this note, we extend the analysis to non-binary networks and show that there exist both a rooted and an unrooted non-binary universal tree-based network with n leaves for all positive integers n.

Keywords: phylogenetic tree, phylogenetic network, tree-based network, universal tree-based network

1 Introduction and preliminaries

Phylogenetic networks are a generalization of phylogenetic trees allowing for the representation of reticulate evolutionary events such as hybridization and horizontal gene transfer. Even though the existence of reticulate evolutionary events is widely accepted, it has been argued that evolution is still fundamentally tree-like with some occasional events of e.g. horizontal gene transfer. This has led to the introduction of so-called tree-based networks, i.e. networks that can be obtained from a tree by adding additional edges [5]. In the following, we consider the existence of one particular class of tree-based networks, namely universal tree-based networks. However, we start by introducing some definitions and notations.

Let $X = \{1, \ldots, n\}$ be a non-empty finite set (e.g. of taxa or species).

An *unrooted phylogenetic network* N on X is a connected, simple graph $G = (V, E)$ with $X \subseteq V$ and no degree-2 vertices, where the set of degree-1 vertices (referred to as the *leaves* of the network) is bijectively labeled and thus identified with X. Such an unrooted network is called *binary* if every non-leaf vertex has degree 3. We call an unrooted network *non-binary* if it is not necessarily binary.

E-mail addresses: email@mareikefischer.de (Mareike Fischer) michelle.galla@gmx.de (Michelle Galla) wicke.6@osu.edu (Kristina Wicke)

A *rooted phylogenetic network* N on X is a rooted directed acyclic graph $G = (V, E)$ with no parallel edges satisfying the following properties:

(i) the root has indegree 0 and outdegree 2 or more;

(ii) all vertices with outdegree 0 have indegree 1, and the set of vertices with outdegree 0 is identified with X (and the vertices in X are again called *leaves*);

(iii) all other vertices either have indegree 1 and outdegree 2 or more (in which case they are called *tree vertices*) or indegree 2 or more and outdegree 1 (in which case they are called *reticulations* or *reticulation vertices*).

If the root has outdegree 2, all reticulations have indegree 2 and all tree-vertices have outdegree 2, the network is called *binary*. We call a rooted network *non-binary* if it is not necessarily binary. For technical reasons, if $|X| = 1$, we allow an (un)rooted non-binary network to consist of a single leaf (which in case of a rooted non-binary network is then at the same time considered to be the root). Moreover, note than an *(un)rooted non-binary phylogenetic tree* on X is an (un)rooted non-binary phylogenetic network whose underlying graph structure is a tree. Furthermore, we refer to non-phylogenetic, i.e. non-leaf-labeled, trees as *tree shapes*.

A non-binary (un)rooted phylogenetic network N on X is called *tree-based* if there is a spanning tree $T = (V, E')$ for N (with $E' \subseteq E$) whose leaf set is equal to X. T is then called a *support tree* (for N). Moreover, the tree T' that can be obtained from T by suppressing potential degree-2 vertices is called a *base tree* (for N). Note that the existence of a support tree T for N implies the existence of a base tree T' for N. Moreover, note that from a graph-theoretical point of view, the support tree T is a *subdivision* of the base tree T' (as T can be obtained from T' by subdividing edges of T with degree-2 vertices). Following [10, p. 357], we will also sometimes say that the support tree T provides an *embedding* of T' in N.

This view is also reflected by the following more constructive definition of tree-basedness by [8] (rooted case) and [7] (unrooted case). First, a rooted non-binary phylogenetic network N on X is tree-based with non-binary base tree T' if N can be obtained from T' via the following steps:

(i) Subdivide the edges of T' by *attachment points* of in- and outdegree 1.

(ii) Add edges, called *linking edges*, between pairs of attachment points or from tree vertices to attachment points, so that N remains acylic and so that each attachment point has degree at least 3, but indegree or outdegree 1.

(iii) Suppress every attachment point that is not incident to a linking edge.

Note that in case of rooted *binary* tree-based networks, the base tree T' has to be binary and in Step (ii) of the above construction, linking arcs can only be added between pairs of attachment points such that N remains binary and acyclic. Similarly, an unrooted non-binary phylogenetic network N on X is tree-based with non-binary base tree T' if N can be obtained from T' via subdividing the edges of T' by attachment points of degree 2, adding linking edges between attachment points, between an attachment point and an original vertex of the tree, or between two original vertices of the tree, and finally suppressing every attachment point that is not incident to a linking edge. Again, in case of unrooted *binary* tree-based networks, T' has to be an unrooted binary tree and linking edges can only be inserted between pairs of attachment points in order to keep the network binary.

Now, an (un)rooted non-binary tree-based network N on X is called *universal* if every non-binary (un)rooted phylogenetic tree on X is a base tree for N.

Universal tree-based networks have recently been analyzed in different studies. First, [5], who introduced the class of tree-based networks, showed that there is a rooted binary universal tree-based network on X for $n = 3$ and asked whether such a network exists in general. This was answered affirmatively by [6] and, independently, by [11], who gave explicit constructions for such networks for all positive integers n. While the construction in [6] contains $\Theta(n!)$ reticulations, the construction in [11] contains $\Theta(n^2)$ reticulations. More recently, [3] showed that a rooted binary universal tree-based network with n leaves has $\Omega(n \log(n))$ reticulations and gave a construction of a rooted binary universal tree-based network with $O(n \log(n))$ reticulations. Moreover, [4] have recently shown that the existence of a rooted binary universal tree-based network implies the existence of an unrooted binary universal tree-based network. Note, however, that so far all considerations of universal tree-based networks in the literature have been concerned with binary networks. In this note, we extend some of these findings to non-binary networks and constructively show that there exist both a rooted and an unrooted non-binary universal tree-based network with n leaves for all positive integers n.

2 Results

2.1 Rooted universal tree-based networks

Theorem 1. *For all positive integers n, there exists a rooted non-binary universal tree-based network with n leaves.*

In the proof of Theorem 1 we will present a construction of a rooted non-binary tree-based network N on X for each n. Following the constructions in [3, 6, 11], this construction consists of two parts: the upper part, which contains the root, is a non-binary network on n leaves that has every rooted non-binary tree shape on n leaves as a base tree; the lower part, which contains the leaves, reorders the leaves of these tree shapes, in order to enable any permutation of leaves and thus, to enable every rooted non-binary phylogenetic tree on X to be a base tree for this network. For the latter, one can for example use a so-called *Beneš network* (cf. [1, 2]) as in [3]. Alternatively, one can simply use a complete bipartite graph $K_{n,n}$ for this purpose (cf. Figure 1(a) and (b)).

Thus, in the following we will only show that the upper part of our construction has every non-binary tree shape as a base tree. Analogously to [3] it then follows that the combination of the upper part with a Beneš network or a complete bipartite graph yields a non-binary universal tree-based network.

Proof (Theorem 1). For all positive integers n, we now give a construction of a rooted non-binary phylogenetic network U_n on n leaves that has every non-binary tree shape on n leaves as base tree. We begin by describing the construction of U_n. First of all, for $n = 1$, U_n consists of a single vertex. Now, let $n \geq 2$. Then, in order to construct U_n, we start with a rooted star tree T^* with root ρ on n leaves where a rooted star tree is a rooted tree shape such that all leaves are adjacent to the root and:

1. Add attachment points to the edges of T^* as follows:

   – For leaf 1 and n, add $n-2$ attachment points on the edges $(\rho, 1)$ and (ρ, n), respectively, and label them $t_1^1, t_1^2, \ldots, t_1^{n-2}$ and $t_n^1, t_n^2, \ldots, t_n^{n-2}$, respectively (starting the labeling at the attachment point closest to the root; note that these vertices will be tree vertices in the final network);

- For each leaf $l = 2, 3, \ldots, n-2$, add $2n-4$ attachment points on the edge (ρ, l) and label them $r_l^1, t_l^1, r_l^2, t_l^2, \ldots, r_l^{n-2}, t_l^{n-2}$ (again, starting the labeling at the attachment point closest to the root; note that all vertices labeled with "r" will be reticulation vertices in the final network and all vertices labeled with "t" will be tree vertices);
- For leaf $n-1$, add $2n-5$ attachment points on edge $(\rho, n-1)$ and label them $r_{n-1}^1, t_{n-1}^1, r_{n-1}^2, t_{n-1}^2, \ldots, t_{n-1}^{n-3}, r_{n-1}^{n-2}$ (again, starting the labeling at the attachment point closest to the root; note that there is no attachment point t_{n-1}^{n-2}, however, all vertices labeled with "r" will be reticulation vertices in the final network and all vertices labeled with "t" will be tree vertices).

2. Add the following edges between attachment points:

- (t_i^k, r_{i+1}^k) for $i = 2, \ldots, n-2$ and $k = 1, \ldots, n-2$ (horizontal edges between tree and reticulation vertices);
- (t_i^k, r_j^{k+1}) for $i = 2, \ldots, n-2$, $j = i+1, \ldots, n-1$ and $k = 1, \ldots, n-3$ (diagonal edges between tree vertices and reticulation vertices from left to right);
- (t_i^k, r_j^{k+1}) for $i = 3, \ldots, n-1$, $j = 2, \ldots, i-1$ and $k = 1, \ldots, n-3$ (diagonal edges between tree vertices and reticulation vertices from right to left);
- (t_i^k, r_j^k) for $i \in \{1, n\}$, $j = 2, \ldots, n-1$ and $k = 1, \ldots, n-2$ (diagonal edges between tree vertices on the paths from ρ to leaves 1 and n, respectively, and reticulation vertices on the paths from ρ to leaves $2, \ldots, n-1$).

Figure 1(e) shows the resulting construction for $n = 3$ and Figure 1(i) shows the construction for $n = 5$.

We now use induction on n to show that – ignoring the leaf labels – every non-binary tree shape on n leaves is base tree of U_n. Since there is exactly one such tree shape for $n = 1$ (consisting of a single vertex) and $n = 2$, the base case holds for all $n \leq 2$. Now, suppose that the claim holds for up to $n-1$ leaves and consider the network U_n on n leaves (with $n \geq 3$).

Note that as the basic structure of U_n is a star tree on n leaves, the star tree on n trivially is a base tree of U_n. Therefore, we will now show that any other rooted non-binary tree shape on n leaves is also a base tree of U_n.

Thus, let T_n be an arbitrary rooted non-binary tree shape (that is not a star tree) with n leaves and root ρ. We will now show that T_n is a base tree of U_n by constructing an explicit embedding of T_n into U_n. i.e., we construct a subdivision of T_n covering all vertices of U_n such that the leaf sets of T_n and U_n coincide. As $n \geq 3$, we know that T_n contains at least one cherry $[u, v]$, i.e. a pair of leaves u and v who share a common parent (cf. [9, Proposition 1.2.5]), say w. As, by assumption, T_n is not a star tree, we may assume that $w \neq \rho$. Suppose that w has $k \geq 2$, children in total (including u and v). Moreover, without loss of generality we may assume that the children of w are labeled $1, \ldots, k$ when enumerating all leaves and are positioned at the outermost left of the tree when drawing it in the plane. We now delete all children of w (which implies that w is now a leaf) and retrieve a tree shape T_{n-k+1} with $n-k+1$ leaves. As $n-k+1 < n$, by induction T_{n-k+1} is a base tree of U_{n-k+1}. Let $V(T_{n-k+1})$ and $E(T_{n-k+1})$ denote the vertex set, respectively edge set, of the underlying embedding of T_{n-k+1} into U_{n-k+1}. In the following, we will first show that T_{n-k+1} can also be embedded in U_n; we will then re-introduce the deleted children of w and show that this yields a base tree of U_n.

Note that by construction T_{n-k+1} contains leaves labeled with w, $k+1$, $k+2, \ldots, n$. Before we embed T_{n-k+1} into U_n, we relabel some vertices of U_{n-k+1}. To be precise, we rename vertices (if they exist) as follows (as an example see Figure 1(e)):

$$w \hookrightarrow 1;$$
$$t^l_{n-k+1} \hookrightarrow t^l_n \text{ for } l = 1, \dots, n-k-1;$$
$$r^l_j \hookrightarrow r^l_{j+k-1} \text{ and } t^l_j \hookrightarrow t^l_{j+k-1} \text{for } j = 2, \dots, n-k \text{ and } l = 1, \dots, n-k-1.$$

We now sequentially extend the network U_{n-k+1} to U_n by introducing additional vertices and edges.

First of all, we add attachment points on existing edges (cf. Figure 1(f)), where for technical reasons $t^0_1 = t^0_n = \rho$.

- edge $(t^{n-k-1}_1, 1)$: $k-1$ attachment points $t^{n-k}_1, t^{n-k+1}_1, \dots, t^{n-2}_1$;
- edge (t^{n-k-1}_n, n): $k-1$ attachment points $t^{n-k}_n, t^{n-k+1}_n, \dots, t^{n-2}_n$;
- edge $(r^{n-k-1}_{n-1}, n-1)$ (if it exists): $2(k-1) = 2k-2$ attachment points $t^{n-k-1}_{n-1}, r^{n-k}_{n-1}, t^{n-k}_{n-1}, \dots, r^{n-2}_{n-1}$;
- edge $(t^{n-k-1}_{j+k-1}, j+k-1)$ for $j = 2, \dots, n-k-1$ (if it exists): $2(k-1) = 2k-2$ attachment points $r^{n-k}_{j+k-1}, t^{n-k}_{j+k-1}, r^{n-k+1}_{j+k-1}, t^{n-k+1}_{j+k-1}, \dots r^{n-2}_{j+k-1}, t^{n-2}_{j+k-1}$.

We add all newly introduced edges to $E(T_{n-k+1})$, i.e. we extend the embedding of T_{n-k+1} to cover all newly introduced attachment points (as an example see Figure 1(f)).

We then add $k-1$ edges connecting the root to leaves $2, \dots, k$ and on each of these edges add $2n-4$ attachment points called $r^1_i, t^1_i, \dots, r^{n-2}_i, t^{n-2}_i$ for $i = 2, \dots, k$ (as an example see Figure 1(g)). If $k = n-1$, we only add $2n-5$ attachment points on edge $(\rho, n-1)$. In particular, we do not add the attachment point t^{n-2}_{n-k}.

In order to complete the construction of U_n, we add all required edges between newly introduced vertices, i.e. we complete the construction of U_n according to the construction principle presented at the beginning of the proof (see page 84; as an example see Figure 1(h)).

We now re-introduce the children of w to the embedding of T_{n-k+1} in order to obtain an embedding of T_n, i.e. we re-introduce the leaves $2, \dots, k$ (note that we do not re-introduce leaf 1, as this was already re-introduced in a previous step). We do this in the following way (cf. Algorithm 1):

First, any existing edge between a tree vertex t^i_1 and a reticulation vertex r^i_{k+1} for $i < n-k$ in the embedding of T_{n-k+1} is replaced by a path between these two vertices, respectively, in order to cover the newly introduced vertices in U_n (Lines 3–6 in Algorithm 1). As an example, the edge (t^1_1, r^1_4) in the embedding of T_3 into U_5 depicted in Figure 1(h) is replaced by the path $\big((t^1_1, r^1_2), (r^1_2, t^1_2), (t^1_2, r^1_3), (r^1_3, t^1_3), (t^1_3, r^1_4)\big)$ (see Figure 1(i)). Once all such existing edges have been replaced by paths, the children of w are added to the embedding of T_{n-k+1} to obtain an embedding of T_n (Lines 7–9 in Algorithm 1). Essentially, for each such child of w, say l, an edge between the tree vertex t^i_1 closest to the root and currently having out-degree 1 (where i is formally determined by Algorithm 1) and the reticulation vertex r^i_l is introduced (first bullet point in Line 8 of Algorithm 1). Then, r^i_l is connected to leaf l via a "straight path" from r^i_l to l, i.e., via a sequence of edges indexed by l (second bullet point in Line 8 of Algorithm 1). Here, for the last edge added, i.e. for the edge incident to leaf l, a case distinction, merely an artifact of the construction of U_n, is necessary, as the parent of leaf l can either be a reticulation vertex (if $l = n-1$) or a tree vertex (in all other cases). In summary, these operations transform the embedding of T_{n-k+1} into an

Algorithm 1:

$i = 1$;
while $i < n - k$ **do**
  if *edge* (t_1^i, r_{k+1}^i) *is in* $E(T_{n-k+1})$ **then**
    remove edge (t_1^i, r_{k+1}^i) from $E(T_{n-k+1})$;
    add edges $(t_1^i, r_2^i), (r_2^i, t_2^i), (t_2^i, r_3^i), (r_3^i, t_3^i), \ldots, (r_k^i, t_k^i), (t_k^i, r_{k+1}^i)$;
    $i = i + 1$;
  else
    add the following edges to $E(T_{n-k+1})$:
    - $(t_1^i, r_2^i), (t_1^i, r_3^i), \ldots, (t_1^i, r_k^i)$;
    - $(r_j^i, t_j^i), (t_j^i, r_j^{i+1}), (r_j^{i+1}, t_j^{i+1}), \ldots, (r_j^{n-2}, t_j^{n-2}), (t_j^{n-2}, j)$ for $j = 2, \ldots, k-1$;

    if $k = n - 1$ **then**
      add the following edges to $E(T_{n-k+1})$:
      - $(r_k^i, t_k^i), (t_k^i, r_k^{i+1}), (r_k^{i+1}, t_k^{i+1}), \ldots, (t_k^{n-3}, r_k^{n-2}), (r_k^{n-2}, k)$;

    else
      add the following edges to $E(T_{n-k+1})$:
      - $(r_k^i, t_k^i), (t_k^i, r_k^{i+1}), (r_k^{i+1}, t_k^{i+1}), \ldots, (r_k^{n-2}, t_k^{n-2}), (t_k^{n-2}, k)$;

    end
  end
end

embedding of T_n in such a way that all vertices of U_n are also vertices of the embedding of T_n and the leaf sets of U_n and T_n coincide. Thus, T_n is a base tree of U_n (as an example see Figure 1(i)). As T_n was an arbitrary rooted non-binary tree shape (that is not the star tree) on n leaves and as the star tree is trivially a base tree for U_n this completes the proof. □

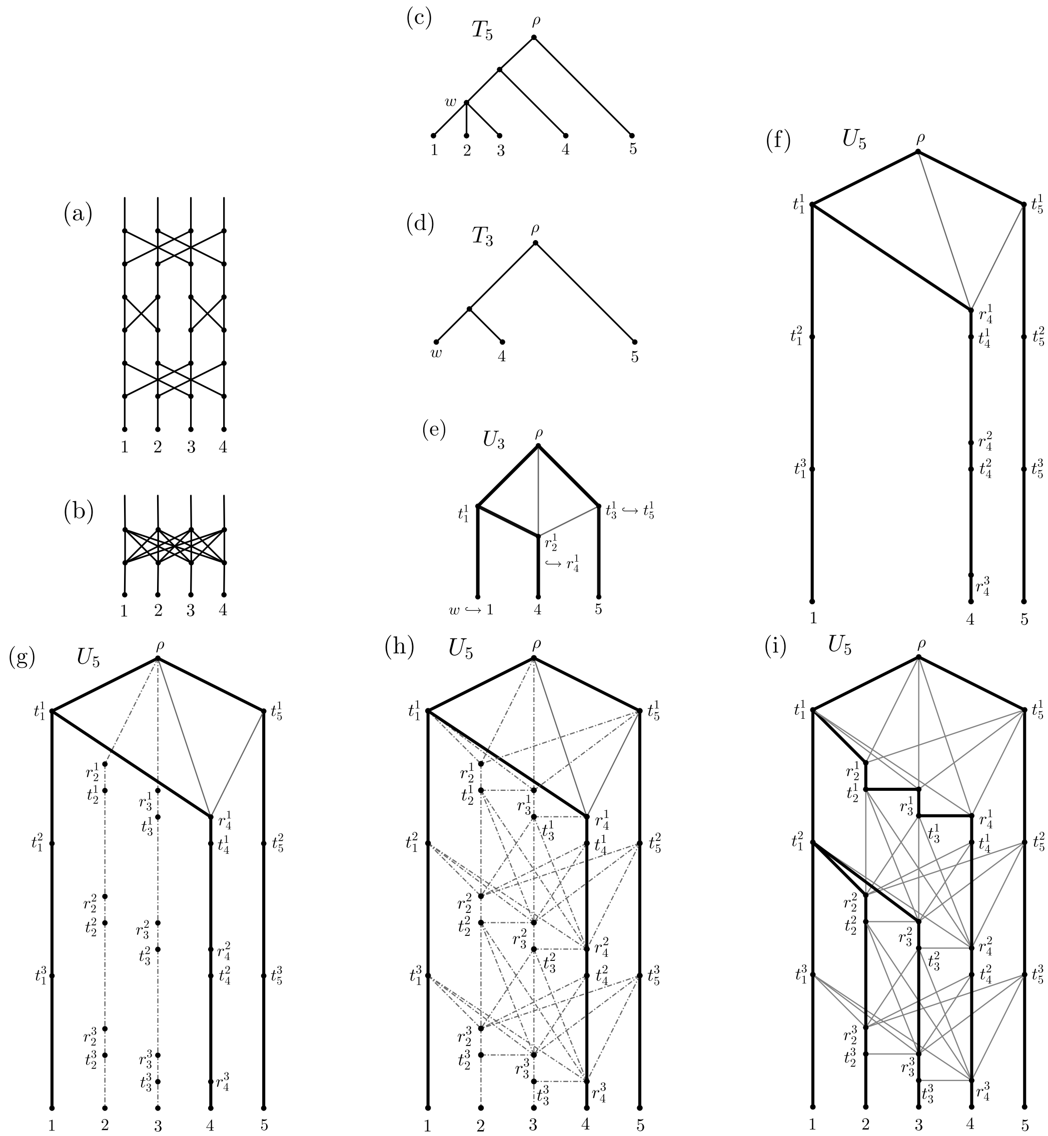

Figure 1: Illustration of the construction and concepts used in the proof of Theorem 1. (a) and (b) show the lower part in the construction of a rooted non-binary universal tree-based network on 4 leaves in form of the Beneš network of size four (Figure taken from [3]) (a) or the complete bipartite graph $K_{4,4}$ (b) (all edges are directed down the page). Moreover, T_5 is a non-binary tree shape on 5 leaves (c). We consider vertex w and delete its children, which yields tree shape T_3 on 3 leaves (d). By the inductive hypothesis T_3 is a base tree of U_3; an embedding is depicted in bold (e). After relabeling vertices, which is depicted by hooked arrows in (e), 2 attachment points are added on the edges $(t_1^1, 1)$ and $(t_5^1, 5)$, respectively, and 4 attachment points are added on the edge $(r_4^1, 4)$ (f). All new edges created in this step, e.g. (t_1^1, t_1^2), are added to the embedding of T_3. Then, 2 edges connecting the root to leaves 2 and 3 are added. These edges are subdivided by introducing 6 attachment points on each edge (g). Then, the construction of U_5 is completed by introducing all missing edges between tree vertices and reticulation vertices and between pairs of reticulation vertices (h). In the last step, the embedding of T_3 is transformed back to an embedding of T_5 (i): Firstly, the edge (t_1^1, r_4^1) (depicted in bold in (h)) is replaced by the edges $(t_1^1, r_2^1), (r_2^1, t_2^1), (t_2^1, r_3^1), (r_3^1, t_3^1)$ and (t_3^1, r_4^1) (depicted in bold in (i)). In the last step the edges $(t_1^2, r_2^2), (t_1^2, r_3^2), (r_2^2, t_2^2), (t_2^2, r_2^3), (r_2^3, t_2^3), (t_2^3, 2), (r_3^2, t_3^2), (t_3^2, r_3^3), (r_3^3, t_3^3)$ and $(t_3^3, 3)$ are added to the embedding of T_5 into U_5. Note that in U_5 all horizontal edges, i.e. edges of type (t_i^k, r_{i+1}^k) for $i = 2, 3$ and $k = 1, 2, 3$, are directed left to right; all other edges are directed away from the root. Similarly, all edges in T_3, T_5 and U_3 are directed away from the root.

2.2 Unrooted universal tree-based networks

Even though Theorem 1 states the existence of a rooted non-binary universal tree-based network on n leaves for all positive integers n, we can use the same construction to show the following statement for unrooted networks.

Corollary 1. *For all positive integers n, there exists an unrooted non-binary universal tree-based network on n leaves.*

Proof. First, for $n = 1$ and $n = 2$, an unrooted non-binary universal tree-based network trivially exists: For $n = 1$, the only unrooted non-binary tree shape is a single vertex, which at the same time is the only unrooted non-binary tree-based network; for $n = 2$, the only unrooted non-binary tree shape is an edge between the two leaves. Thus, any unrooted non-binary tree-based network on 2 leaves can be considered an unrooted non-binary universal tree-based network for $n = 2$. Now, for $n \geq 3$, consider the construction of the rooted non-binary universal tree-based network in the proof of Theorem 1. By ignoring the designation of the vertex ρ as root of the network and the orientation of edges, this construction yields an unrooted non-binary universal tree-based network with n leaves, which completes the proof. □

3 Discussion

In this note, we have constructively shown that there exist both a rooted and an unrooted non-binary universal tree-based network with n leaves for all positive integers n. Like the rooted binary universal tree-based networks in [3, 6, 11] the rooted non-binary universal tree-based network constructed in this paper is *stack-free* (i.e. it has no two reticulations one of which is a parent of the other), but unlike the constructions in [3, 6, 11] it is not *temporal* (*time-consistent*) (where a rooted non-binary network N^r is called temporal if there is a mapping $t : V(N^r) \rightarrow \mathbb{N}$ such that if (u, v) is a tree edge, then $t(u) < t(v)$, while if (u, v) is a reticulation edge, then $t(u) = t(v)$). To see this, consider vertices ρ, t_1^1 and r_2^1 in U_3 (Figure 1(e)) or U_5 (Figure 1(i)). As (ρ, t_1^1) is a tree edge, in a temporal rooted non-binary network we would have $t(\rho) < t(t_1^1)$. However, as (t_1^1, r_2^1) is a reticulation edge, we would also have $t(t_1^1) = t(r_2^1)$ and similarly, as (ρ, r_2^1) is also a reticulation edge, we would have $t(\rho) = t(r_2^1)$. In particular by the previous case, $t(\rho) = t(r_2^1) = t(t_1^1)$, which contradicts the fact that $t(\rho) < t(t_1^1)$. It would thus be of interest for future research to investigate whether there also exists a rooted non-binary universal tree-based network on n leaves that is temporal for all positive integers n.

Moreover, it might be of interest to infer a theoretical lower bound for the number of reticulations needed for a rooted non-binary tree-based network to be universal (as done in [3] for the rooted binary case) and to investigate if there exists a rooted non-binary universal tree-based network that achieves this minimum. Note that the construction given in this manuscript requires $(n-2)^2$ reticulations in the upper part U_n and n reticulations in the lower part (if the complete bipartite graph $K_{n,n}$ is used). Thus, the construction given here has $O(n^2)$ reticulations. However, our construction is possibly more complex than necessary. For example, for $n = 4$, it is possible to construct another rooted non-binary tree-based network U_4' that has every non-binary rooted tree shape with four leaves as a base tree and only uses two reticulations (whereas U_4 uses four reticulations). The network U_4', which could replace U_4 in our construction, is shown in Figure 2. It would thus be of interest to study whether this construction can be generalized to $n > 4$ and analyze if it is minimal in the number of reticulations.

We remark that in the case of rooted binary universal tree-based networks on n leaves, [3] obtained the theoretical lower bound of $\Omega(n\log(n))$ reticulations required by equating $\frac{1}{\sqrt{2}}\left(\frac{2}{e}\right)^n n^{n-1}$ (an asymptotic approximation to the number $(2n-3)!!$ of rooted binary phylogenetic trees with n leaves (see, e.g., [10, p. 16])) with 2^r and solving for r, where r denotes the number of reticulations (since if N is a rooted binary phylogenetic network with n leaves and r reticulations it embeds at most 2^r distinct rooted binary phylogenetic trees, as each embedding of such a tree is obtained by choosing exactly one of the two reticulation edges directed into each reticulation vertex). Now, in the non-binary case, this approach will result in the trivial bound of one reticulation required unless further restrictions are imposed on the indegrees of reticulation vertices. To see this, first denote by $nbr(n)$ the number of rooted non-binary phylogenetic trees with n leaves (note that no closed formula for $nbr(n)$ is known but exponential generating functions have been established (see, e.g., [10, p. 17]). Next, consider a phylogenetic network N with one reticulation vertex of indegree $nbr(n)$. Then N could in theory embed up to $nbr(n)$ distinct rooted non-binary phylogenetic trees, each one corresponding to a different choice of one of the $nbr(n)$ reticulation edges present in N. However, if for instance the maximum indegree of any reticulation vertex is fixed to some integer $d < nbr(n)$, non-trivial bounds may be obtainable.

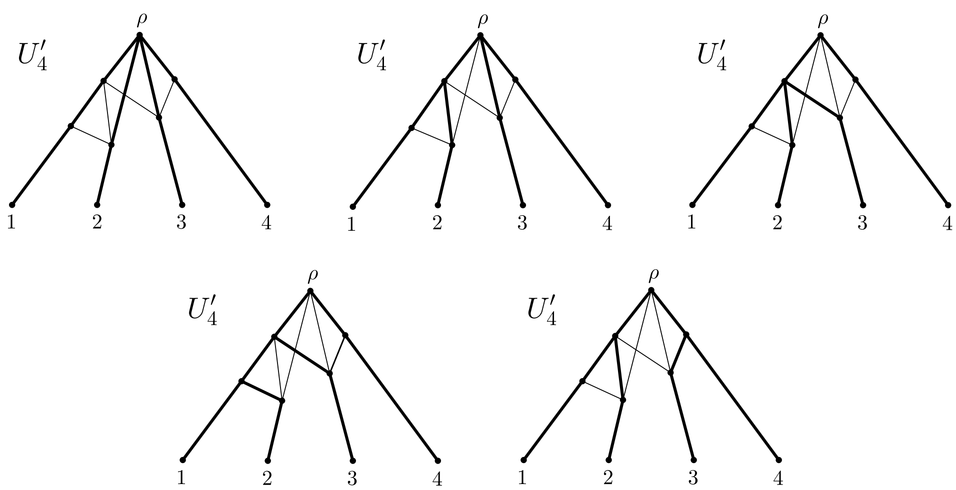

Figure 2: Rooted non-binary tree-based network U_4' that has every rooted non-binary tree shape with 4 leaves as a base tree (an embedding is highlighted in bold, respectively). U_4' could thus replace U_4 as described in the proof of Theorem 1 in our construction of a rooted non-binary universal tree-based network on 4 leaves.

Acknowledgements

We thank an anonymous referee of an earlier version of this manuscript for suggesting the simpler complete bipartite graph as an alternative to the Beneš network for the lower part of our construction. In addition, we thank two anonymous referees for valuable suggestions to improve the quality of this manuscript in its present form. Moreover, we thank Lina Herbst for helpful discussions on the general topic and for comments on an earlier version of this manuscript. Last but not least, Mareike Fischer thanks the joint research project ***DIG-IT!*** supported by the European Social Fund (ESF), reference: ESF/14-BM-A55-0017/19, and the Ministry of Education, Science and Culture of Mecklenburg-Vorpommern, Germany.

References

[1] V. E. Beneš. Optimal rearrangeable multistage connecting networks. *Bell System Technical Journal*, 43(4):1641–1656, Jul 1964. doi:10.1002/j.1538-7305.1964.tb04103.x.

[2] V. E. Beneš. Permutation groups, complexes, and rearrangeable connecting networks. *Bell System Technical Journal*, 43(4):1619–1640, Jul 1964. doi:10.1002/j.1538-7305.1964.tb04102.x.

[3] M. Bordewich and C. Semple. A universal tree-based network with the minimum number of reticulations. *Discrete Applied Mathematics*, 250:357–362, Dec. 2018. doi:10.1016/j.dam.2018.05.010.

[4] A. Francis, K. T. Huber, and V. Moulton. Tree-based unrooted phylogenetic networks. *Bulletin of Mathematical Biology*, 80(2):404–416, Feb 2018. doi:10.1007/s11538-017-0381-3.

[5] A. Francis and M. Steel. Which phylogenetic networks are merely trees with additional arcs? *Systematic Biology*, 64(5):768–777, 2015. doi:10.1093/sysbio/syv037.

[6] M. Hayamizu. On the existence of infinitely many universal tree-based networks. *Journal of Theoretical Biology*, 396:204–206, May 2016. doi:10.1016/j.jtbi.2016.02.023.

[7] M. Hendriksen. Tree-based unrooted nonbinary phylogenetic networks. *Mathematical Biosciences*, 302:131–138, Aug 2018. doi:10.1016/j.mbs.2018.06.005.

[8] L. Jetten and L. van Iersel. Nonbinary tree-based phylogenetic networks. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 15(1):205–217, Jan 2018. doi:10.1109/tcbb.2016.2615918.

[9] C. Semple and M. Steel. *Phylogenetics (Oxford Lecture Series in Mathematics and Its Applications)*. Oxford University Press, 2003.

[10] M. Steel. *Phylogeny: Discrete and random processes in evolution.* Society for Industrial and Applied Mathematics, Philadelphia PA, 2016.

[11] L. Zhang. On tree-based phylogenetic networks. *Journal of Computational Biology*, 23(7):553–565, Jul 2016. doi:10.1089/cmb.2015.0228.