EXTRACTING DIAMONDS: IDENTIFIABILITY OF 4-NODE CYCLES IN LEVEL-1 PHYLOGENETIC NETWORKS UNDER A PSEUDOLIKELIHOOD COALESCENT MODEL

George Tiley*
Kew Royal Botanic Gardens
Richmond, UK
g.tiley@kew.org

Claudia Solís-Lemus*
University of Wisconsin-Madison
Madison, WI
solislemus@wisc.edu

Abstract

Here, we investigate the (generic) identifiability of the position of the hybrid node in a 4-node hybridization cycle in a semi-directed level-1 phylogenetic network. While generic identifiability is easily attained under non-restrictive assumptions such as $t \in (0, \infty)$ for all branches and $\gamma \in (0, 1)$ for the inheritance probability of the hybrid edges, simulations show that accurate detection of these cycles can be complicated by inadequate sampling, small sample size or gene tree estimation error. We identify practical advice for evolutionary biologists on best sampling strategies to improve the detection of this type of hybridization cycle.

Keywords: Phylogenetic networks, Reticulation, Identifiability, Coalescent model

1 Introduction

The increasing evidence of reticulate processes in the Tree of Life has inspired the development of novel mathematical and statistical methods to infer phylogenetic networks from genomic data [24, 10, 26, 27]. A phylogenetic network is an extension of the tree graph by the inclusion of hybrid nodes that allow for two incoming parental edges which represent genetic material being transferred from two different sources. Methods to infer phylogenetic networks include classes from combinatorial [17, 18, 49, 34, 45], distance-based [47, 33, 48], or model-based under Markov models that use genetic sequences or markers as input [16, 38, 31] or under the multispecies coalescent model [39] that uses gene trees as input [52, 42, 54].

*Equal contribution; co-corresponding authors

Model-based approaches provide accurate representations of the reticulate evolutionary process among species represented by the estimated phylogenetic network. However, the accurate estimation of this parameter requires proof of identifiability. Indeed, multiple studies have tackled the question of whether phylogenetic networks are identifiable under different models [35, 42, 15, 30, 6, 2, 44, 16, 48], and here, we extend the identifiability proofs in [42, 44] to further validate the inference of level-1 phylogenetic networks under a pseudolikelihood model built on the multispecies network coalescent model. Since identifiability can be unattainable in many cases, in practice, one aims to prove *generic identifiability* which means that the network parameter is almost surely identifiable.

Let $\mathcal{N}$ be a semi-directed level-1 phylogenetic network with h hybridization events and n taxa. We denote this the *network parameter* of the model. Let $\mathcal{G} = \{G_1, G_2, \ldots, G_g\}$ be g estimated gene trees as input for the model. We can define a likelihood as the product of the probability under the multispecies network coalescent model [51] of each gene tree given the network parameter (with its numerical parameters as branch lengths and inheritance probabilities). For the case of a species tree as the model parameter, it has been proven that the distribution of gene trees identifies (or generically identifies) the tree topology, branch lengths and even root position under the multispecies tree coalescent model [9, 5, 4, 40]. The same is not known for the full likelihood model on species networks, but generic identifiability proofs have been derived for the case of a quartet-based pseudolikelihood model [42, 6, 44] that assumes probabilities of each observed quartet gene tree under its corresponding species quarnet are independent. These results show that the presence of hybridization events in n-taxon level-1 semi-directed phylogenetic networks is generically identifiable from the observed distribution of gene trees under a pseudolikelihood model as described in the next theorem.

Theorem 1 (Generic identifiability of a hybridization event [42, 44]) *Let $\mathcal{N}$ be a semi-directed level-1 n-taxon phylogenetic network with one hybridization event. Let k be the number of nodes in the cycle defined by the hybridization event. Then, the hybridization event is generically identifiable if i) $k > 3$, ii) $t \in (0, \infty)$ for all branch lengths, and iii) $\gamma \in (0, 1)$ for the inheritance probability corresponding to the hybridization event.*

The proof is based on techniques from algebraic geometry, and the level-1 assumption extends this result to any number of hybridization events [42, 44]. Namely, we compare the set of polynomial equations defined by the network on the observed quartet tree frequency patterns to those defined by a tree without the hybridization event (the major tree). We never, however, compare the set of polynomial equations defined by two different phylogenetic networks. Indeed, our results in [42, 44] only allow us to conclude that a level-1 n-taxon network is generically identifiable when compared to its major n-taxon tree, but we cannot conclude anything regarding network-to-network identifiability.

Here, we begin to explore the question of network identifiability by focusing on level-1 phylogenetic networks with hybridization cycles of 4 nodes (4-node cycles). These types of cycles (denoted *diamonds* [37]) are relevant because in practice they seem to lead to flat pseudolikelihood scores. That is, 4-node cycles that only differ on the placement of the hybrid node in the cycle can have similar pseudolikelihood scores.

In this paper, we prove the following theorem regarding the generic identifiability of the placement of the hybrid node in the 4-node hybridization cycle in a level-1 network:

Theorem 2 *Let $\mathcal{N}$ be a semi-directed level-1 n-taxon phylogenetic network with one hybridization event that creates a 4-node hybridization cycle. Then, the placement of the hybrid node in the cycle is generically identifiable if i) $n > 5$, ii) $t \in (0, \infty)$ for all branch lengths, and iii) $\gamma \in (0, 1)$ for the inheritance probability corresponding to the hybridization event.*

One of the main implications of Theorem 2 is that semi-directed network (correct direction of hybrid edges) is generically identifiable from input unrooted gene trees, not only the unrooted network (placement of the cycle). This implies that the direction of the gene flow event (hybrid edge) can be distinguished with the pseudolikelihood model, providing an accurate evolutionary interpretation of the origin and target of reticulations. Our proof follows the same algebraic geometry techniques as in [42, 44] such as the definition of the set of polynomial equations under the coalescent model and the search for unique (or finitely many) solutions.

The organization of the manuscript is as follows. In Section 2, we establish the models more precisely and the approach used to prove generic identifiability. In Section 3, we define the theorems and proofs for the generic identifiability of the placement of the hybrid node in the 4-node hybridization cycle. In Sections 4 and 5, we illustrate the mathematical findings (and the practical considerations when dealing with limited data) with a simulation study. Finally, we present some discussion in Section 6.

2 Model Preliminaries

2.1 Semi-directed level-1 explicit phylogenetic networks

Our main parameter of interest is the topology $\mathcal{N}$ of a phylogenetic network (Definition 1) along with the numerical parameters of the vector of branch lengths ($\boldsymbol{t}$) and a vector of inheritance probabilities ($\boldsymbol{\gamma}$). The inheritance probabilities describe the proportion of genes inherited by a hybrid node from one of its hybrid parent (see Figure 1).

Definition 1 *A rooted explicit phylogenetic network [25] $\mathcal{N}$ on taxon set X is a connected directed acyclic graph with vertices $V = \{r\} \cup V_L \cup V_H \cup V_T$, edges $E = E_H \cup E_T$ and a bijective leaf-labeling function $f : V_L \to X$ with the following characteristics:*

- *The root r has indegree 0 and outdegree 2*
- *Any leaf $v \in V_L$ has indegree 1 and outdegree 0*
- *Any tree node $v \in V_T$ has indegree 1 and outdegree 2*
- *Any hybrid node $v \in V_H$ has indegree 2 and outdegree 1*
- *A tree edge $e \in E_T$ is an edge whose child is a tree node*
- *A hybrid edge $e \in E_H$ is an edge whose child is a hybrid node*

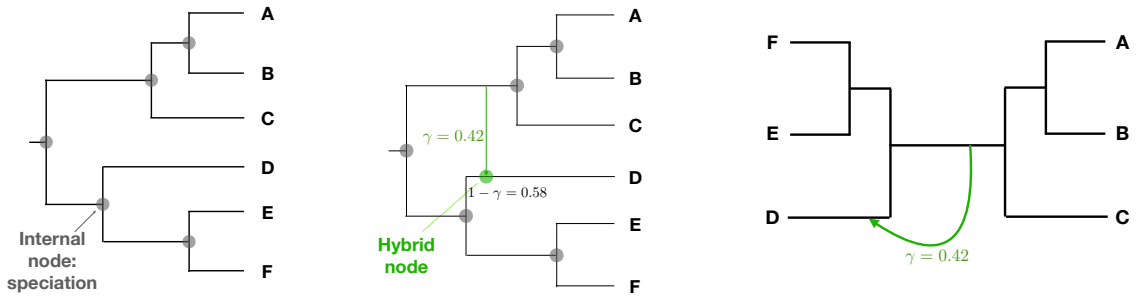

Figure 1: Left: Rooted phylogenetic tree of 6 taxa. Internal nodes (gray circles) represent speciation events and are usually omitted on phylogenetic trees. Center: Rooted phylogenetic network on 6 taxa with $h = 1$ hybridization event represented by the green edge. The hybrid node (in green) represents a reticulation event. The hybrid node has two parent edges: minor hybrid edge in green labeled $\gamma = 0.42$ and major hybrid edge in black labeled $1 - \gamma = 0.58$. The reticulation event can represent different biological processes: hybridization, horizontal gene transfer or introgression. Right: Semi-directed network for the same biological scenario in center. Although the root location is unknown, its position is constrained by the direction of the hybrid edges. For example, the taxon D cannot be an outgroup.

- *A hybrid edge $e \in E_H$ has an inheritance probability parameter $\gamma_e < 1$ which represents the proportion of the genetic material that the child hybrid node received from this parent edge. For a tree edge e, $\gamma_e = 1$*

In a rooted explicit network, every internal node represents a biological process: speciation for tree nodes and hybridization for hybrid nodes. However, other types of phylogenetic networks, such as unrooted networks [25] and semi-directed networks [42, 16] are also useful representation of evolutionary relationships. Unrooted phylogenetic networks are typically obtained by suppressing the root node and the direction of all edges. In semi-directed networks, on the other hand, the root node is suppressed and we ignore the direction of all tree edges, but we maintain the direction of hybrid edges, thus keeping information on which nodes are hybrids. The placement of the root is then constrained, because the direction of the two hybrid edges to a given hybrid node inform the direction of time at this node: the third edge must be a tree edge directed away from the hybrid node and leading to all the hybrid's descendants. Therefore the root cannot be placed on any descendant of any hybrid node, although it might be placed on some hybrid edges. See Figure 1 for the example of a rooted explicit phylogenetic network (center), and its semi-directed version (right).

Each hybridization event creates a blob (or cycle) which represents a subgraph with at least two nodes and no cut-edges. A cut-edge is any edge in the network whose removal disconnects the network. Traditionally, the term *cycle* is used for directed (rooted) networks, and the term *blob* is used for undirected (unrooted) networks. We make no distinction here as we will focus on *semi-directed* networks, and thus, some edges in the cycle (blob) will be directed, and some will not. We also note that for every hybridization event, there are two parent hybrid edges connected to the hybrid node: 1) major hybrid edge with inheritance probability $\gamma > 0.5$, and 2) a minor hybrid edge with inheritance probability $\gamma < 0.5$. Both edges are parametrized with the same γ.

Here, we focus on the case of semi-directed networks. In addition, we assume the following characteristics of the semi-directed network under study.

Assumption 1 *Let $\mathcal{N}$ have n species and h hybridization events (that is, $|V_L| = n$ and $|V_h| = h$).*

Assumption 2 *We further assume that $\mathcal{N}$ is of level-1 [25]. That is, we assume that any given edge can be part of at most one cycle which means that there is no overlap between any two cycles ([25, 41].*

Thus, our parameters of interest are $(\mathcal{N}, \boldsymbol{t}, \boldsymbol{\gamma})$ where $\mathcal{N}$ is an explicit semi-directed level-1 phylogenetic network that links the n species under study, and has h hybridization events. This network has two vectors of numerical parameters: 1) branch lengths $\boldsymbol{t} \in [0, \infty)^{n_e}$ for n_e branches in the network, and 2) inheritance probabilities $\boldsymbol{\gamma} \in [0, 1]^{n_h}$ for n_h minor hybrid edges.

2.2 Phylogenetic networks as algebraic varieties

We represent every semi-directed level-1 network $\mathcal{N}$ as a set of polynomial concordance factor (CF) equations as in [42, 44].

First, each network can be decomposed into 4-taxon subnetworks (*quarnets* in [21]). That is, for a given network $\mathcal{N}$ with $n \geq 4$ taxa, we consider all 4-taxon subsets $\mathcal{S} = \{s = \{a, b, c, d\} : a, b, c, d \in X\}$ to define the theoretical CFs expected under the coalescent model for each 4-taxon subset. These theoretical CFs are already derived for a species tree in [3], and for a species network in [42]. In both cases, the CFs do not depend on the position of the root. For the tree, the major CF is defined for the quartet that agrees with the species tree. That is, if the species tree has the split $ab|cd$ with internal edge t, then the major CF would be $CF_{ab|cd} = 1 - 2/3\exp(-t)$. The CF for the minor resolutions (in disagreement with the species tree $ab|cd$) would then be $CF_{ac|bd} = CF_{ad|bc} = 1/3\exp(-t)$ [22].

For the case of a 4-taxon network, the theoretical CFs are weighted averages of CFs on trees. For example, Figure 2 (center) shows a semi-directed 4-taxon network and the two displayed quartet trees depending on which hybrid edge is used (major hybrid edge used corresponds to the quartet on top, and minor hybrid edge used corresponds to the quartet on the bottom). The CFs on this 4-taxon network are then given by the weighted averages of CF of the two displayed quartet trees with weights: $1 - \gamma$ and γ. Namely,

- $CF_{ab|cd} = (1 - \gamma)(1 - 2/3\exp(-t_{13} - t_{23} - t_2)) + \gamma(1/3\exp(-t_2))$ for the major resolution, and
- $CF_{ac|bd} = CF_{ad|bc} = (1 - \gamma)(1/3\exp(-t_{13} - t_{23} - t_2)) + \gamma(1/3\exp(-t_2))$ for the minor resolutions.

Definition 2 *Let $\mathcal{N}$ be n-taxon semi-directed level-1 explicit phylogenetic network with h hybridizations. This network defines a set of $3\binom{n}{4}$ CF equations under the coalescent model with parameters $\boldsymbol{t}$ and $\boldsymbol{\gamma}$. Denote this system of equations as $CF(\mathcal{N}, \boldsymbol{t}, \boldsymbol{\gamma})$. If we change the variables to $z_i = \exp(-t_i)$ for all internal branch lengths, then $CF(\mathcal{N}, \boldsymbol{z}, \boldsymbol{\gamma})$ is a system of polynomial equations.*

We show in the Appendix the system of polynomial equations for the four phylogenetic networks considered in this study: $\mathcal{N}_{down}$, $\mathcal{N}_{left}$, $\mathcal{N}_{right}$, and $\mathcal{N}_{up}$ (Figure 3). Since each network is obtained by the rotation of the hybrid node in

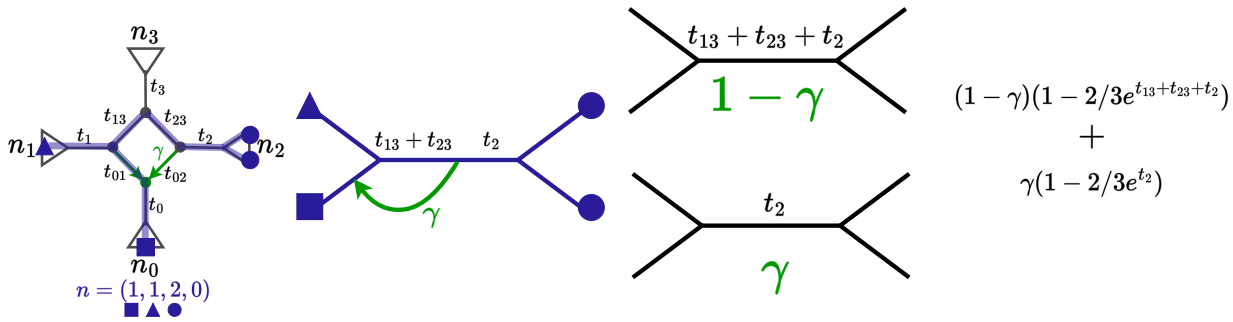

Figure 2: Left: Semi-directed network with a given 4-taxon subset $(1, 1, 2, 0)$ highlighted in blue which corresponds to taking one individual in n_0 (red square), one individual in n_1 (red triangle) and two individuals in n_2 (red circles). To obtain the CF equations for the resulting quartet (center in blue), we split it as two quartet trees weighted by $1 - \gamma$ and γ (center in black). Right: Major CF equation for the quartet as weighted average of CF equations on quartet trees. Because the two quartet trees have the same topology, the two minor CF equations of the quartet are equal (not shown).

the hybridization cycle, there is a simple mapping of 4-taxon subsets that allows us to define the CF equations for all networks from the original $\mathcal{N}_{down}$. This process is also described in the Appendix.

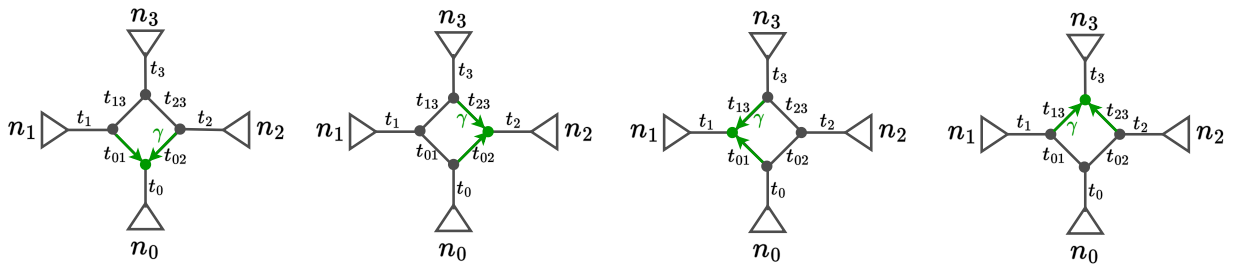

Figure 3: Semi-directed networks with one hybridization cycle with 4 nodes (diamond), but different position of the hybrid node in the cycle. We denote these networks $\mathcal{N}_{down}$, $\mathcal{N}_{right}$, $\mathcal{N}_{left}$, and $\mathcal{N}_{up}$ respectively from left to right.

2.3 Generic identifiability

To show that the position of the hybrid node in the hybridization cycle is generically identifiable, we find if both systems of polynomial equations share solutions in the parameter space: $CF(\mathcal{N}, \boldsymbol{t}, \boldsymbol{\gamma}) = CF(\mathcal{N}', \boldsymbol{t}', \boldsymbol{\gamma}')$.

Definition 3 (Generic identifiability of hybrid node position) *Let $\mathcal{N}$ be an n-taxon semi-directed level-1 explicit phylogenetic network with h hybridizations. We focus on one hybridization that has 4 nodes in the hybridization cycle. Let $\mathcal{N}'$ be a network with the hybrid node rotated inside the 4-node hybridization cycle. Let $CF(\mathcal{N}, \boldsymbol{t}, \boldsymbol{\gamma})$ be the system of polynomial CF equations defined by $\mathcal{N}$, and let $CF(\mathcal{N}', \boldsymbol{t}', \boldsymbol{\gamma}')$ be the system of polynomial CF equations defined by $\mathcal{N}'$. We say the position of the hybrid node in the 4-node hybridization cycle in $\mathcal{N}$ is* ***identifiable*** *if the system of $CF(\mathcal{N}, \boldsymbol{t}, \boldsymbol{\gamma}) = CF(\mathcal{N}', \boldsymbol{t}', \boldsymbol{\gamma}')$ does not have solutions in any set of numerical parameters $(\boldsymbol{t}, \boldsymbol{\gamma}, \boldsymbol{t}', \boldsymbol{\gamma}')$. We say the position of the hybrid node in the 4-node hybridization cycle in $\mathcal{N}$ is* ***generically identifiable*** *if the solution set of the system $CF(\mathcal{N}, \boldsymbol{t}, \boldsymbol{\gamma}) = CF(\mathcal{N}', \boldsymbol{t}', \boldsymbol{\gamma}')$ has Lebesgue measure zero.*

3 Identifiability of 4-node cycles in 6-taxon semi-directed level-1 phylogenetic networks

In this section, we define three theorems comparing the generic identifiability of $\mathcal{N}_{down}$ versus $\mathcal{N}_{right}$ (Theorem 3), $\mathcal{N}_{down}$ versus $\mathcal{N}_{left}$ (Theorem 4), and $\mathcal{N}_{down}$ versus $\mathcal{N}_{up}$ (Theorem 5) when all networks only have six taxa and one hybridization event. We generalize to n taxa and h hybridization events in Remark 2.

Theorem 3 *Let $\mathcal{N}_{down}$ be a semi-directed level-1 6-taxon phylogenetic network with one hybridization event producing a hybridization cycle with 4 nodes. Without loss of generality, let the taxa be partitioned among clades as $n_0 = 1, n_1 = 2, n_2 = 1, n_3 = 2$ (Figure 3). Let the hybrid node be ancestral to the clade n_0. Let $\mathcal{N}_{right}$ be a semi-directed level-1 6-taxon phylogenetic network with one hybridization event producing a hybridization cycle with 4 nodes such that the unrooted version of $\mathcal{N}_{right}$ agrees with the unrooted version of $\mathcal{N}_{down}$. Let the hybrid node in the hybridization cycle in $\mathcal{N}_{right}$ be ancestral to the clade n_2. Then, $\mathcal{N}_{down}$ and $\mathcal{N}_{right}$ are identifiable if $t_1 < \infty, t_{13} > 0, t_3 < \infty$, and $\gamma \in (0, 1)$.*

Let $CF(\mathcal{N}_{down}, \boldsymbol{z}, \boldsymbol{\gamma})$ be the system of CF polynomial equations defined by $\mathcal{N}_{down}$ and let $CF(\mathcal{N}_{right}, \boldsymbol{z}', \boldsymbol{\gamma}')$ be the system of CF polynomial equations defined by $\mathcal{N}_{right}$. Both systems of equations can be found in the Appendix.

Let $\mathcal{P} = \{p(\boldsymbol{z}, \boldsymbol{\gamma}) - q(\boldsymbol{z}', \boldsymbol{\gamma}') : p(\boldsymbol{z}, \boldsymbol{\gamma}) \in CF(\mathcal{N}_{down}, \boldsymbol{z}, \boldsymbol{\gamma}), q(\boldsymbol{z}', \boldsymbol{\gamma}') \in CF(\mathcal{N}_{right}, \boldsymbol{z}', \boldsymbol{\gamma}')\}$ be the set of polynomial equations resulting from matching $CF(\mathcal{N}_{down}, \boldsymbol{z}, \boldsymbol{\gamma})$ to $CF(\mathcal{N}_{right}, \boldsymbol{z}', \boldsymbol{\gamma}')$ for every 4-taxon subset.

Using Macaulay2 [13], we compute the Gröbner basis of $\mathcal{P}$ on the $(\boldsymbol{z}, \boldsymbol{\gamma})$ variables by any elimination order. All Macaulay2 scripts are available at `https://github.com/gtiley/diamond-identifiability`.

The resulting ideal is given by

$$z_1(z_{13} - 1)^2 z_3 \gamma(1 - \gamma) = 0$$
$$z_1(z_{13} - 1) z_3 (z_{23} - 1)\gamma(1 - \gamma) = 0$$

which represent the conditions that the $(\boldsymbol{z}, \boldsymbol{\gamma})$ variables need to satisfy for the polynomial set $\mathcal{P}$ to vanish to zero.

Thus, $\mathcal{N}_{down}$ and $\mathcal{N}_{right}$ are not identifiable in the subset of parameter space corresponding to $\{z_1 = 0\} \cup \{z_{13} = 1\} \cup \{z_3 = 0\} \cup \{\gamma = 0\} \cup \{\gamma = 1\}$.

Theorem 4 *Let $\mathcal{N}_{down}$ be a semi-directed level-1 6-taxon phylogenetic network with one hybridization event producing a hybridization cycle with 4 nodes. Without loss of generality, let the taxa be partitioned among clades as $n_0 = 1, n_1 = 2, n_2 = 1, n_3 = 2$ (Figure 3). Let the hybrid node be ancestral to the clade n_0. Let $\mathcal{N}_{left}$ be a semi-directed level-1 6-taxon phylogenetic network with one hybridization event producing a hybridization cycle with 4 nodes such that the unrooted version of $\mathcal{N}_{left}$ agrees with the unrooted version of $\mathcal{N}_{down}$. Let the hybrid node in the hybridization cycle in $\mathcal{N}_{left}$ be ancestral to the clade n_1. Then, $\mathcal{N}_{down}$ and $\mathcal{N}_{left}$ are generically identifiable if $t_1 < \infty, t_{13} > 0, t_3 < \infty$, and $\gamma \in (0, 1)$.*

The structure of the proof is the same as for Theorem 3, but we repeat it for completeness. Let $CF(\mathcal{N}_{down}, \boldsymbol{z}, \boldsymbol{\gamma})$ be the system of CF polynomial equations defined by $\mathcal{N}_{down}$ and let $CF(\mathcal{N}_{left}, \boldsymbol{z}', \boldsymbol{\gamma}')$ be the system of CF polynomial equations defined by $\mathcal{N}_{left}$. Both systems of equations can be found in the Appendix.

Let $\mathcal{P} = \{p(\boldsymbol{z}, \boldsymbol{\gamma}) - q(\boldsymbol{z}', \boldsymbol{\gamma}') : p(\boldsymbol{z}, \boldsymbol{\gamma}) \in CF(\mathcal{N}_{down}, \boldsymbol{z}, \boldsymbol{\gamma}), q(\boldsymbol{z}', \boldsymbol{\gamma}') \in CF(\mathcal{N}_{left}, \boldsymbol{z}', \boldsymbol{\gamma}')\}$ be the set of polynomial equations resulting from matching $CF(\mathcal{N}_{down}, \boldsymbol{z}, \boldsymbol{\gamma})$ to $CF(\mathcal{N}_{left}, \boldsymbol{z}', \boldsymbol{\gamma}')$.

Using Macaulay2 [13], we compute the Gröbner basis of $\mathcal{P}$ on the $(\boldsymbol{z}, \boldsymbol{\gamma})$ variables by any elimination order. All Macaulay2 scripts are available at `https://github.com/gtiley/diamond-identifiability`.

Let $\mathbb{C}[(\boldsymbol{z}, \boldsymbol{\gamma}, \boldsymbol{z}', \boldsymbol{\gamma}') : \boldsymbol{z} \in [0, \infty)^{n-3}, \boldsymbol{\gamma} \in [0, 1]^h, \boldsymbol{z}' \in [0, \infty)^{n-3}, \boldsymbol{\gamma}' \in [0, 1]^h]$ be the set of all polynomials on the $(\boldsymbol{z}, \boldsymbol{\gamma}, \boldsymbol{z}', \boldsymbol{\gamma}')$ variables.

The resulting ideal is given by

$$\{(z_1)(z_{13} - 1)(z_3)(\gamma_z)(\gamma_z - 1)p_i(\boldsymbol{z}, \boldsymbol{\gamma}) : p_i \in \mathbb{C}[(\boldsymbol{z}, \boldsymbol{\gamma}, \boldsymbol{z}', \boldsymbol{\gamma}')]; \deg(p_i) = 7; i = 1, \ldots, 8\}$$

which represent the conditions that the $(\boldsymbol{z}, \boldsymbol{\gamma})$ variables need to satisfy for the polynomial set $\mathcal{P}$ to vanish to zero.

As the polynomials $\{p_i(\boldsymbol{z}, \boldsymbol{\gamma})\}_{i=1}^{8}$ have Lebesgue measure zero, $\mathcal{N}_{down}$ and $\mathcal{N}_{left}$ are not generically identifiable in the subset of parameter space corresponding to $\{z_1 = 0\} \cup \{z_{13} = 1\} \cup \{z_3 = 0\} \cup \{\gamma = 0\} \cup \{\gamma = 1\}$.

Theorem 5 *Let $\mathcal{N}_{down}$ be a semi-directed level-1 6-taxon phylogenetic network with one hybridization event producing a hybridization cycle with 4 nodes. Without loss of generality, let the taxa be partitioned among clades as $n_0 = 1, n_1 = 2, n_2 = 1, n_3 = 2$ (Figure 3). Let the hybrid node be ancestral to the clade n_0. Let $\mathcal{N}_{up}$ be a semi-directed level-1 6-taxon phylogenetic network with one hybridization event producing a hybridization cycle with 4 nodes such that the unrooted version of $\mathcal{N}_{up}$ agrees with the unrooted version of $\mathcal{N}_{down}$. Let the hybrid node in the hybridization cycle in $\mathcal{N}_{up}$ be ancestral to the clade n_3. Then, $\mathcal{N}_{down}$ and $\mathcal{N}_{up}$ are generically identifiable if $t_1 < \infty, t_{13} > 0, t_3 < \infty$, and $\gamma \in (0, 1)$.*

The structure of the proof is the same as for Theorem 4, but we repeat it for completeness. Let $CF(\mathcal{N}_{down}, \boldsymbol{z}, \boldsymbol{\gamma})$ be the system of CF polynomial equations defined by $\mathcal{N}_{down}$ and let $CF(\mathcal{N}_{up}, \boldsymbol{z}', \boldsymbol{\gamma}')$ be the system of CF polynomial equations defined by $\mathcal{N}_{up}$. Both systems of equations can be found in the Appendix.

Let $\mathcal{P} = \{p(\boldsymbol{z}, \boldsymbol{\gamma}) - q(\boldsymbol{z}', \boldsymbol{\gamma}') : p(\boldsymbol{z}, \boldsymbol{\gamma}) \in CF(\mathcal{N}_{down}, \boldsymbol{z}, \boldsymbol{\gamma}), q(\boldsymbol{z}', \boldsymbol{\gamma}') \in CF(\mathcal{N}_{up}, \boldsymbol{z}', \boldsymbol{\gamma}')\}$ be the set of polynomial equations resulting from matching $CF(\mathcal{N}_{down}, \boldsymbol{z}, \boldsymbol{\gamma})$ to $CF(\mathcal{N}_{up}, \boldsymbol{z}', \boldsymbol{\gamma}')$.

Using Macaulay2 [13], we compute the Gröbner basis of $\mathcal{P}$ on the $(\boldsymbol{z}, \boldsymbol{\gamma})$ variables by any elimination order. All Macaulay2 scripts are available at `https://github.com/gtiley/diamond-identifiability`.

Let $\mathbb{C}[(\boldsymbol{z}, \boldsymbol{\gamma}, \boldsymbol{z}', \boldsymbol{\gamma}') : \boldsymbol{z} \in [0, \infty)^{n-3}, \boldsymbol{\gamma} \in [0, 1]^h, \boldsymbol{z}' \in [0, \infty)^{n-3}, \boldsymbol{\gamma}' \in [0, 1]^h]$ be the set of all polynomials on the $(\boldsymbol{z}, \boldsymbol{\gamma}, \boldsymbol{z}', \boldsymbol{\gamma}')$ variables.

The resulting ideal is given by

$$\{(z_1)(z_{13}-1)(z_3)(\gamma_z)(\gamma_z-1)p_i(\boldsymbol{z},\boldsymbol{\gamma}) : p_i \in \mathbb{C}[(\boldsymbol{z},\boldsymbol{\gamma},\boldsymbol{z}',\boldsymbol{\gamma}')]; \deg(p_i)=10; i=1,\ldots,12\}$$

which represent the conditions that the $(\boldsymbol{z},\boldsymbol{\gamma})$ variables need to satisfy for the polynomial set $\mathcal{P}$ to vanish to zero.

As the polynomials $\{p_i(\boldsymbol{z},\boldsymbol{\gamma})\}_{i=1}^{12}$ have Lebesgue measure zero, $\mathcal{N}_{down}$ and $\mathcal{N}_{up}$ are not generically identifiable in the subset of parameter space corresponding to $\{z_1=0\}\cup\{z_{13}=1\}\cup\{z_3=0\}\cup\{\gamma=0\}\cup\{\gamma=1\}$.

Remark 1 *We note that by assuming that $t \in (0,\infty)$ for all branch lengths, and $\gamma \in (0,1)$ for the inheritance probabilities, we can guarantee generic identifiability of the placement of the hybrid node in the 4-node hybridization cycle.*

Remark 2 *The identifiability of the position of the hybrid node in 4-node cycles in n-taxon level-1 phylogenetic networks is obtained by noticing that we need at most 2 taxa per clade to define all the CF polynomial equations (Lemma 1 in [44]), and thus, if the hybridization cycles are identifiable with only one taxon in some clades, the addition of a second taxon will only reduce the set in the parameter space where the two networks are not (generically) identifiable.*

4 Simulations

4.1 Description of experiments

We simulated gene trees under the multispecies network coalescent (MNSC) using BPP v4.1.4 [12]. The level-1 network under investigation contains eight taxa and one 4-node hybridization cycle (diamond) (Figure 4a). To simulate gene trees under the MNSC, we require a rooted evolutionary history, and thus, we investigate three potential root placements: 1) a balanced root included in the cycle (r_1 in Figure 4b), 2) a ladderized root with the root in the cycle (r_2 in Figure 4c), and 3) a ladderized root such that the cycle occurs in the ingroup (r_3 in Figure 4d). The root height is not the same between the balanced root and ladderized rootings, but the age of episodic gene flow is $1\frac{1}{2}$ units of θ, population size measured in expected nucleotide diversity. The inheritance probability (γ) was 0.5 for all simulations. Simulations were carried out for nine diamonds for the three rooting scenarios, such that we changed if lineages were represented by a clade of two taxa or a single taxon Figure 4e.

Parameters for sequence and gene tree simulation under the MNSC was a θ of 0.01 constant over all nodes. The HKY model of nucleotide substitution [19] with a transition-transversion rate-ratio of 3 and equilibrium frequencies at $\pi_A = \pi_T = 0.3$ and $\pi_C = \pi_G = 0.2$ was selected. A low degree of among site rate variation was incorporated using a gamma distribution [50] with a shape and rate of 0.6. A strict clock was used among branches and there is no among-locus rate variation. Simulated sequences have 1000 base pairs. One-hundred simulations were performed for each of the 27 rooted networks, using 100, 500, 1000, and 5000 loci. Scripts for reproducing simulation experiments are available at `https://github.com/gtiley/diamond-identifiability`.

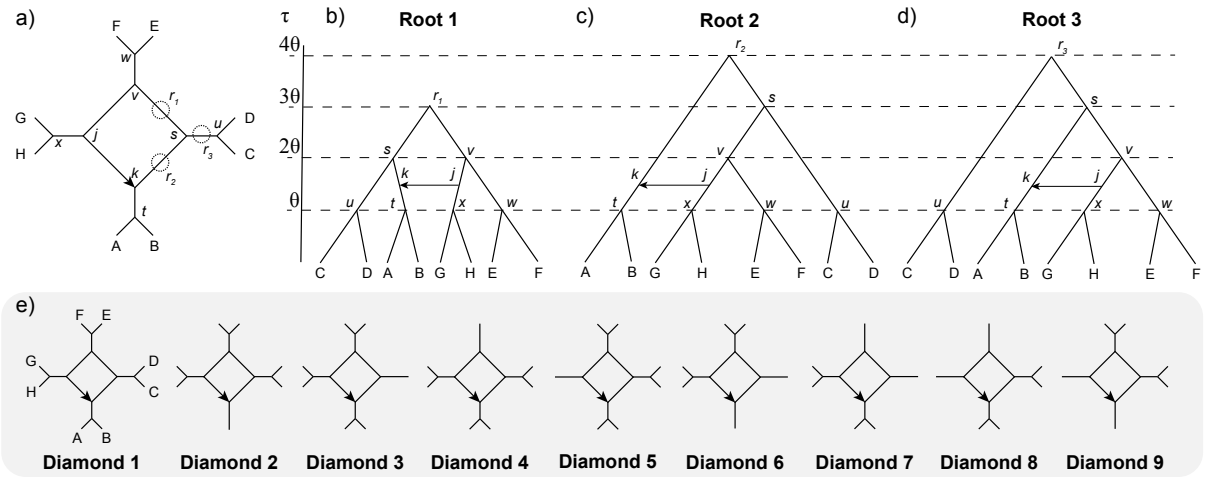

Figure 4: Semi-directed level-1 network topology with a diamond hybridization cycle. a) The semi-directed diamond has tip labels in capital letters and node labels in lower case italicized letters. The hybrid node is labeled as k. The location of the three possible roots are indicated by r_1, r_2, r_3. b-d) Node heights for the three rooted networks in units of θ population size. The direction of episodic gene flow is shown with an arrow. e) For each root, nine networks were investigated, the first including all tips, the next four reducing one cherry to a single branch in an anticlockwise fashion, the next four reducing two cherries to two lineages with a single taxon.

4.2 Network estimation

Networks were estimated from the simulated gene trees using pseudolikelihood methods that utilize the gene trees topology only. We used two methods: 1) the SNaQ function [42] implemented through the PhyloNetworks v0.15.0 Julia package [43] in Julia v1.6.5 and 2) the `InferNetwork_MPL` function [53] implemented through PhyloNet v3.8.3 [46]. Notably, SNaQ uses unrooted quartets as data while `InferNetwork_MPL` uses rooted triples (more on rooting error later). Both analyses used ten independent runs with the maximum number of allowed reticulation events set to zero, one, or two.

Three summary statistics were used to evaluate the performance of the estimations. First, we calculated the proportion of simulations where only one reticulation was detected based on a two-point pseudolikehood score difference. Although the pseudolikelihood does not have straight-forward implementation of model selection such as the Akaike Information Criterion [1], we used differences in pseudolikelihood scores as some operational criteria to evaluate detection of reticulation edges among the replicates in lieu of slope heuristics [7] or goodness-of-fit tests [8]. Second, we verified if the estimated network had the same topology as the true network using the `hardwiredClusterDistance` function [25] in PhyloNetworks. While the absolute distance between networks is difficult to interpret, a distance of 0 means the network topologies are identical. Third, we checked if the true minor hybrid edge is present in the estimated network with only one estimated hybridization event. This is done by drawing bootstrap support from the estimated network onto the true network with the `hybridBootstrapSupport` function from PhyloNetworks. A bootstrap support of 100% shows that the true hybrid edge is present in the estimated network, even if some other aspects of the network are incorrect. Scripts for calculating these summary statistics are available at `https://github.com/gtiley/diamond-identifiability`.

4.3 Incorporating gene tree error into simulation experiments

Biological data is often messy with many potential sources of gene tree estimation error. For example, deep paralogy can lead to gene trees that conflict with species trees [32] and lead to rooting errors when using an outgroup [20, 23]. Such gene tree errors can be prevalent in groups with histories of large-scale gene duplication and loss such as plants [28] and insects [29]. These rooting errors may negatively effect the performance of inference methods, especially those based on rooted triples. Therefore, for each simulation condition, we randomly re-rooted 10% and 30% of the gene trees. Because rooting errors may also be accompanied by other topological errors, we estimated networks from the simulated data incorporating a random re-rooting and a random nearest neighbor interchange (NNI) move to 10% and 30% of gene trees. Scripts for introducing random roots and NNI moves to trees are available at `https://github.com/gtiley/diamond-identifiability`.

5 Results

5.1 Simulation results: Error-free gene trees

The pseudolikelihood methods implemented in PhyloNetworks and PhyloNet generally perform well across all simulation conditions. SNaQ is capable of correctly identifying the presence of one reticulation event, recovering the correct network, and recovering the reticulation node in all simulations as the number of gene trees increased (Supplementary Figures 9, 10, 11). For diamonds d_6 and d_9 when the rooting is ladderized and in the cycle (r_2) or ladderized and out of the cycle (r_3), SNaQ sometimes missed the correct network or reticulation node when the number of gene trees was low, but converged to the correct network as gene trees increased from 100 to 5000. `InferNetwork_MPL` was more efficient with respect to the number of gene trees for d_6 or d_9 and performed well for r_1 (Supplementary Figure 12) and r_2 (Supplementary Figure 13), but sometimes struggled with r_3 (Supplementary Figure 14). Across all diamonds for r_3, `InferNetwork_MPL` did not always detect one reticulation edge, such that two reticulation edges were preferred over none (Supplementary Data). Even when considering the estimated networks that only allowed one reticulation event, the network was not always correct; although, the correct reticulation node is almost always recovered (Supplementary Figure 14). If the recovery of the reticulation node is the most favorable criteria of an estimator, both SNaQ and `InferNetwork_MPL` perform well with error free gene trees, with SNaQ requiring more gene trees in cases where the the hybrid node had a single taxon for a descendant instead of a cherry (Figure 5).

5.2 Simulation results: Gene trees with error

SNaQ tolerated rooting errors well when they were present in both 10% (Supplementary Figures 15, 16, 17) and 30% (Supplementary Figures 18, 19, 20) of gene trees. The patterns for d_6 and d_9 with r_2 and r_3 from error-free gene trees were observed again, which was expected since SNaQ used unrooted quartets as data. `InferNetwork_MPL` was largely robust to root error in 10% of the gene trees for r_1 and r_2, with the exception of sometimes detecting more than one reticulation edge (Supplementary Figures 21 and 22). However, it was difficult to recover the correct network for r_3 and

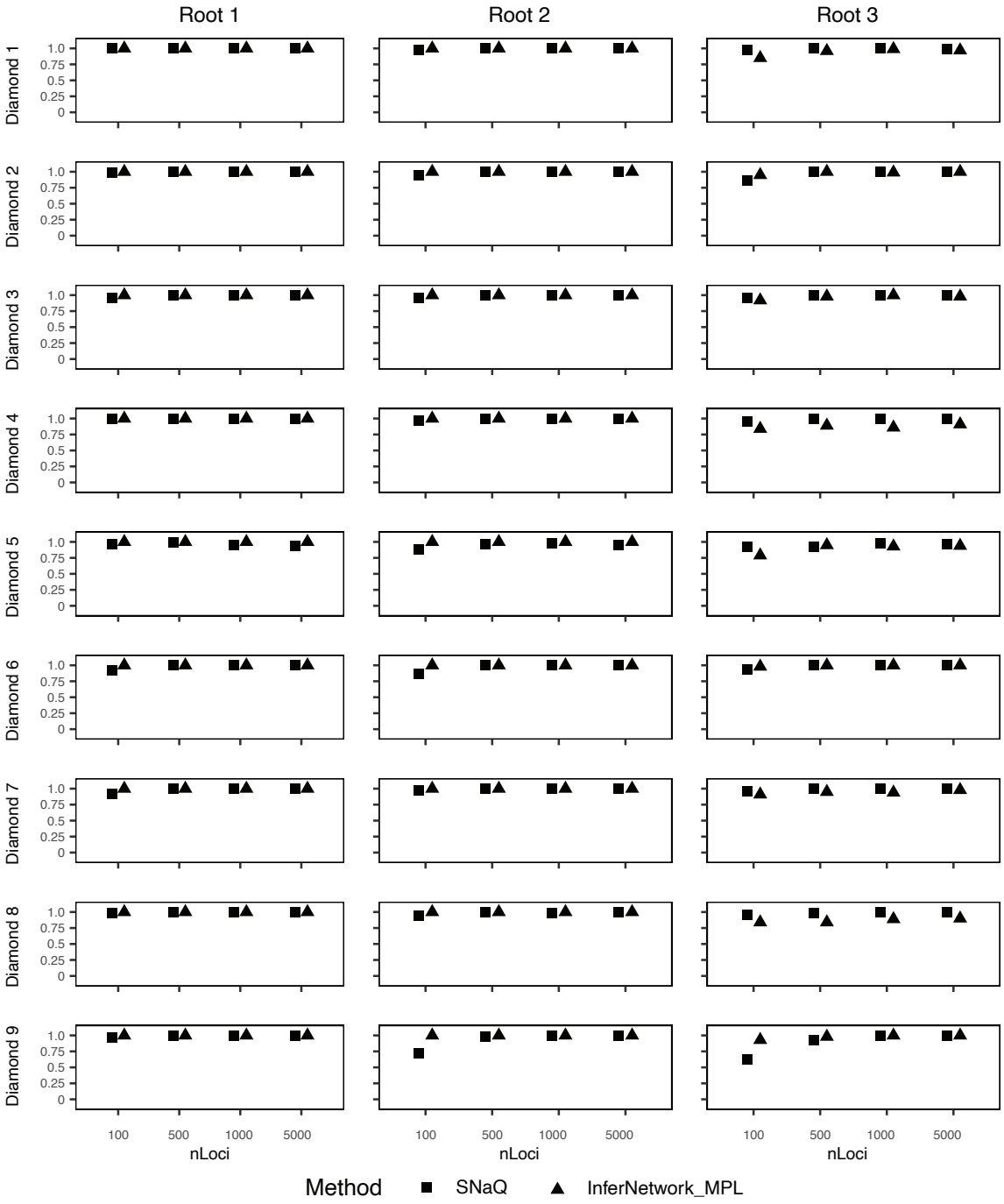

Figure 5: Proportion of 100 simulation replicates where the true reticulation edge is recovered by SNaQ (squares) or `InferNetwork_MPL` (triangles). X-axis corresponds to the number of gene trees in the input sample. The diamond and root numbers correspond to Figure 4.

the proportion of incorrect networks sometimes increased with the number of gene trees (Supplementary Figure 23). The lowest proportions of correct network estimations are observed for diamonds d_4, d_7, and d_9, which are not all cases where the hybrid node has a single taxon as a descendant. The overall network being incorrect does not always mean that the estimated reticulation node is wrong, but the decreased performance compared to simulations without gene tree error was drastic (Supplementary Material Table 15). As the proportion of gene tree errors increased from 10% to 30%, similar patterns were observed for `InferNetwork_MPL`, but with a higher proportion of network estimation errors (Supplementary Figure 24, 25, 26). For example, the reticulation node was never recovered for d_9 with r_3 and 5000 gene trees (Supplementary Figure 26).

Introducing an NNI move to the gene trees with random roots had little effect on SNaQ's ability to correctly detect one reticulation node, estimate the correct network, and recover the reticulation node. SNaQ nearly converges to the correct answer in all simulation conditions as the number of gene trees increased, regardless of whether errors were present in 10% (Supplementary Figures 27, 28, 29) or 30% (Supplementary Figures 30, 31, 32) of gene trees. The NNI error did not greatly affect `InferNetwork_MPL` compared to the rooting error alone. `InferNetwork_MPL` is robust to gene tree error with wrong roots and the NNI move in 10% of gene trees for r_1 and r_2 (Supplementary Figures 33, 34). There are many estimation errors for r_3, with the reticulation node detected about half of the time with 5000 gene trees for d_4 and d_9; however, the reticulation node is almost always recovered despite the estimated network being wrong for d_6 (Supplementary Figure 35). When 30% of the gene trees had rooting and NNI errors, `InferNetwork_MPL` still performed well for r_1 and r_2 while the proportion of incorrect estimates increaseed for r_3 (Supplementary Figures 36, 37, 38). Even with the more conservative assumption of 10% gene tree error, SNaQ was more robust than `InferNetwork_MPL`, at least when errors cause rooting problems (Figure 6). The estimation performance was overall sensitive to the root placement of the simulated network and sampling of a single taxon versus cherries as hybrid node descendants despite the diamond remaining constant across simulations.

6 Discussion

6.1 Mathematical insights

We have shown that the placement of the hybrid node in a 4-node hybridization cycle is generically identifiable in semi-directed level-1 phylogenetic networks. This result has important biological implications such as correct rooting of semi-directed network which provide more information about the actual speciation process from the origin of the clade. More work is needed to understand the identifiability of larger (and smaller) cycles. We decided to focus on the case of 4-node hybridization cycle because of empirical evidence suggesting flat pseudolikelihood on these hybridization events. However, smaller cycles are also biologically interesting as they describe hybridization events between more closely related populations. It has already been proven that 2-node cycles are not identifiable, but 3-node cycles are whenever we have sufficient sampling from the hybrid and sister clades [42, 44]. Thus, we can explore the identifiability of the

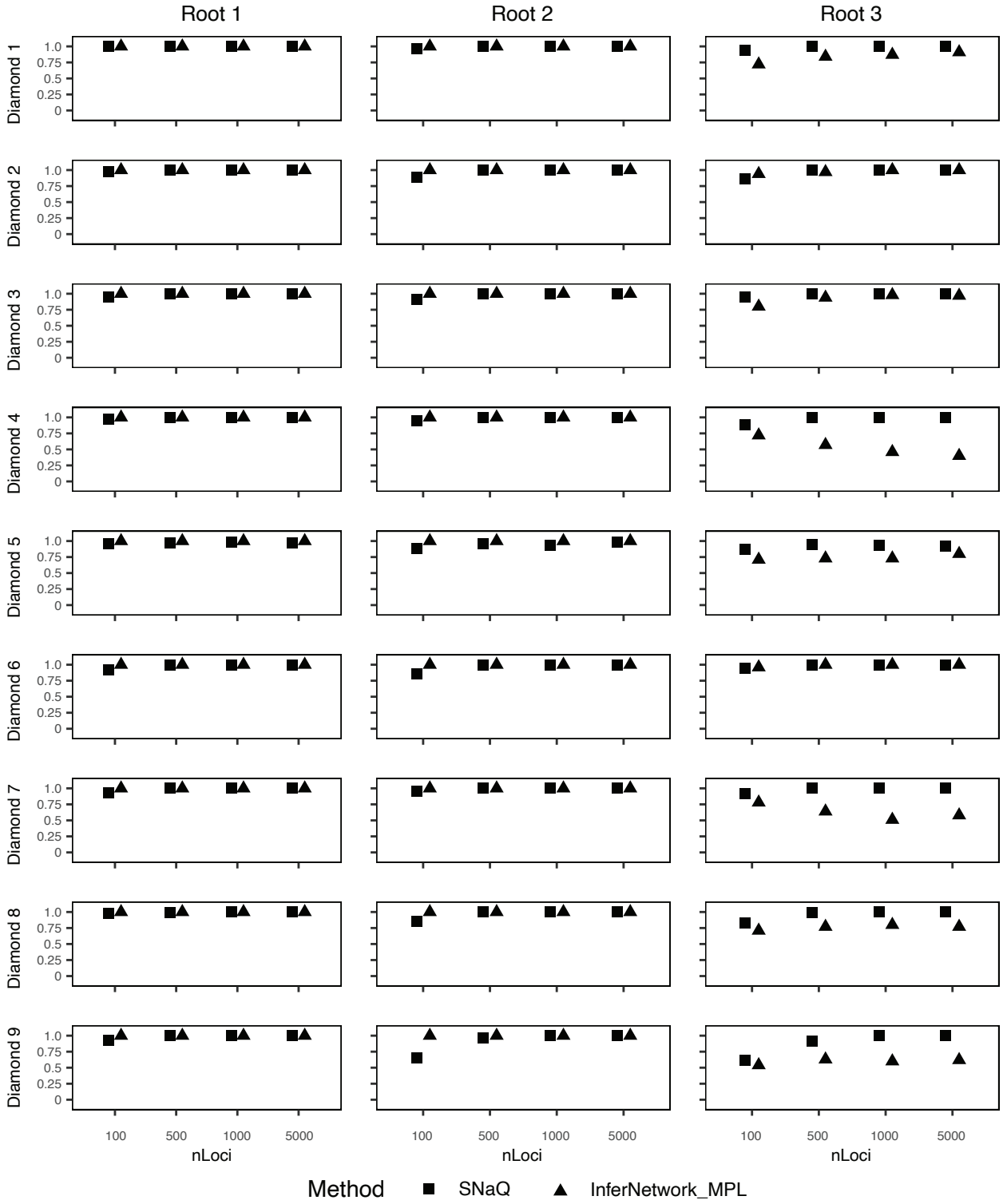

Figure 6: Proportion of 100 simulation replicates where the reticulation node was recovered by SNaQ (squares) or `InferNetwork_MPL` (triangles). X-axis corresponds to the number of gene trees in the input sample. The diamond and root numbers correspond to Figure 4. Error was introduced into 10% of gene trees for each round of simulation by randomly rerooting the gene tree and performing a random NNI move.

placement of the hybrid node in the hybridization 3-node cycle as future work. In particular, we want to answer the following question:

Question 1 *Let $\mathcal{N}_1, \mathcal{N}_2, \mathcal{N}_3$ be the three rotations of the n-taxon semi-directed level-1 phylogenetic networks with one hybridization event producing a hybridization cycle of 3 nodes. Are $\mathcal{N}_1, \mathcal{N}_2, \mathcal{N}_3$ (generically) identifiable?*

Furthermore, all existing identifiability studies on semi-directed phylogenetic networks are restricted to a level-1 network [42, 44, 6, 15, 16]. Questions on the identifiability of higher level networks remain open.

6.2 Biological insights from simulations and best practices

Pseudolikelihood methods can perform well for estimating a level-1 network from gene tree distributions in the absence of gene tree error, consistent with previous simulations and empirical analyses [53, 42]. PhyloNet's `InferNetwork_MPL` was more efficient than PhyloNetwork's SNaQ for diamonds d_2, d_6, and d_9, capable of recovering the correct network with as few as 100 gene trees where SNaQ would require at least 500 (Figure 5). However, `InferNetwork_MPL` was not always perfect even with error-free gene trees such as d_8 for r_3. In addition, SNaQ was far more robust to the types of gene tree errors explored here than `InferNetwork_MPL`, capable of detecting one reticulation event and recovering the correct network in most conditions when provided a sufficient number of gene trees (Figure 6). `InferNetwork_MPL` could still detect the presence of a reticulation edge more often than not depending on the diamond analysed with r_3, but the estimated network was frequently incorrect and if one were using unknown empirical data, there is a risk of estimating additional false positive reticulation nodes in the presence of gene tree rooting errors (Supplementary Data).

The diamonds and rootings where network estimation errors were most prevalent have implications for empirical analyses. SNaQ errors primarily occurred for d_2, d_6, and d_9, all diamonds where the hybrid node descendant was a single taxon (Figure 4e). This implies that sampling two or more individuals for the lineage of interest could provide an increase in statistical power over a single taxon, at least for SNaQ and methods that use unrooted gene trees. Interestingly, network estimation errors were most prevalent when the cycle was restricted to an ingroup (r_3). `InferNetwork_MPL` especially struggled with d_4, d_7, and d_9 in the presence of errors, as it appears more data would not be helpful. Precisely why these diamonds performed poorly is unclear, but it could be an artifact of how errors were introduced to simulations. For example, d_4 and d_7 both have cherries for the descendants of the hybrid node. Since the outgroup errors were done randomly, this created more chances for a descendant of the hybrid node to be the outgroup. Outgroup lineages are typically sampled at an appropriate evoltionary distance to polarize phylogenetic relationships or site patterns in the context of *D*-statistics [14] [11] [36], but our results show that rooting errors and potentially other sources of topological errors (e.g. deep paralogy, assembly or genotyping, biases in starting material such as museum tissue vs fresh) could be misleading for estimating level-1 networks with rooted triples. An unrooted quartet-based method such as SNaQ should be more robust to gene tree error, but rooted triple-based methods should perform well and can be beneficial for a low number of gene trees (e.g. 100) if gene tree quality can be reliably evaluated.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 101026923, awarded to GPT. This work was also funded by the National Science Foundation [DEB-2144367 to CSL].

References

[1] Hirotugu Akaike. A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723, 1974.

[2] Elizabeth S Allman, Hector Baños, and John A Rhodes. Nanuq: a method for inferring species networks from gene trees under the coalescent model. *Algorithms for Molecular Biology*, 14:1–25, 2019.

[3] Elizabeth S. Allman, James H. Degnan, and John A. Rhodes. Identifying the rooted species tree from the distribution of unrooted gene trees under the coalescent. *Journal of Mathematical Biology*, 62(6):833–862, 2011.

[4] Elizabeth S Allman, Sonia Petrović, John A Rhodes, and Seth Sullivant. Identifiability of two-tree mixtures for group-based models. *IEEE/ACM transactions on computational biology and bioinformatics*, 8(3):710–722, 2010.

[5] Elizabeth S Allman and John A Rhodes. The identifiability of tree topology for phylogenetic models, including covarion and mixture models. *Journal of Computational Biology*, 13(5):1101–1113, 2006.

[6] Hector Baños. Identifying species network features from gene tree quartets under the coalescent model. *Bulletin of mathematical biology*, 81:494–534, 2019.

[7] Jean-Patrick Baudry, Cathy Maugis, and Bertrand Michel. Slope heuristics: overview and implementation. *Statistics and Computing*, 22(2):455–470, 2012.

[8] Ruoyi Cai and Cécile Ané. Assessing the fit of the multi-species network coalescent to multi-locus data. *Bioinformatics*, 37(5):634–641, 2021.

[9] Joseph T Chang. Full reconstruction of markov models on evolutionary trees: identifiability and consistency. *Mathematical biosciences*, 137(1):51–73, 1996.

[10] James H Degnan. Modeling hybridization under the network multispecies coalescent. *Systematic biology*, 67(5):786–799, 2018.

[11] Eric Y Durand, Nick Patterson, David Reich, and Montgomery Slatkin. Testing for ancient admixture between closely related populations. *Molecular biology and evolution*, 28(8):2239–2252, 2011.

[12] Tomáš Flouri, Xiyun Jiao, Bruce Rannala, and Ziheng Yang. A bayesian implementation of the multispecies coalescent model with introgression for phylogenomic analysis. *Molecular biology and evolution*, 37(4):1211–1223, 2020.

[13] DR Grayson and ME Stillman. Macaulay2, a software system for research in algebraic geometry. Available at `http://www.math.uiuc.edu/Macaulay2/`.

[14] Richard E Green, Johannes Krause, Adrian W Briggs, Tomislav Maricic, Udo Stenzel, Martin Kircher, Nick Patterson, Heng Li, Weiwei Zhai, Markus Hsi-Yang Fritz, et al. A draft sequence of the neandertal genome. *science*, 328(5979):710–722, 2010.

[15] Elizabeth Gross and Colby Long. Distinguishing phylogenetic networks. *SIAM Journal on Applied Algebra and Geometry*, 2(1):72–93, 2018.

[16] Elizabeth Gross, Leo van Iersel, Remie Janssen, Mark Jones, Colby Long, and Yukihiro Murakami. Distinguishing level-1 phylogenetic networks on the basis of data generated by markov processes. July 2020.

[17] Stefan Grünewald, Kristoffer Forslund, Andreas Dress, and Vincent Moulton. Qnet: an agglomerative method for the construction of phylogenetic networks from weighted quartets. *Molecular Biology and Evolution*, 24(2):532–538, 2007.

[18] Stefan Grunewald, Andreas Spillner, Sarah Bastkowski, Anja Bogershausen, and Vincent Moulton. Superq: computing supernetworks from quartets. *IEEE/ACM transactions on computational biology and bioinformatics*, 10(1):151–160, 2013.

[19] Masami Hasegawa, Hirohisa Kishino, and Taka-aki Yano. Dating of the human-ape splitting by a molecular clock of mitochondrial dna. *Journal of molecular evolution*, 22(2):160–174, 1985.

[20] BR Holland, D Penny, and MD Hendy. Outgroup misplacement and phylogenetic inaccuracy under a molecular clock—a simulation study. *Systematic biology*, 52(2):229–238, 2003.

[21] Katharina T. Huber, Vincent Moulton, Charles Semple, and Taoyang Wu. Quarnet inference rules for level-1 networks. *Bulletin of Mathematical Biology*, 80(8):2137–2153, 2018.

[22] Richard R Hudson. Testing the constant-rate neutral allele model with protein sequence data. *Evolution*, pages 203–217, 1983.

[23] John P Huelsenbeck, Jonathan P Bollback, and Amy M Levine. Inferring the root of a phylogenetic tree. *Systematic biology*, 51(1):32–43, 2002.

[24] Daniel H Huson and David Bryant. Application of phylogenetic networks in evolutionary studies. *Molecular biology and evolution*, 23(2):254–267, 2006.

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

[26] Xiyun Jiao, Tomáš Flouri, and Ziheng Yang. Multispecies coalescent and its applications to infer species phylogenies and cross-species gene flow. *National Science Review*, 8(12):nwab127, 2021.

[27] Sungsik Kong, Joan Carles Pons, Laura Kubatko, and Kristina Wicke. Classes of explicit phylogenetic networks and their biological and mathematical significance. *Journal of Mathematical Biology*, 84(6):47, 2022.

[28] James H Leebens-Mack, Michael S Barker, Eric J Carpenter, Michael K Deyholos, Matthew A Gitzendanner, Sean W Graham, Zheng Grosse, Ivo andLi, Michael Melkonian, Siavash Mirarab, et al. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature*, 574(7780):679–685, 2019.

[29] Zheng Li, George P Tiley, Sally R Galuska, Chris R Reardon, Thomas I Kidder, Rebecca J Rundell, and Michael S Barker. Multiple large-scale gene and genome duplications during the evolution of hexapods. *Proceedings of the National Academy of Sciences*, 115(18):4713–4718, 2018.

[30] Colby Long and Laura Kubatko. The effect of gene flow on coalescent-based species-tree inference. *Systematic biology*, 67(5):770–785, 2018.

[31] Sarah Lutteropp, Céline Scornavacca, Alexey M Kozlov, Benoit Morel, and Alexandros Stamatakis. Netrax: accurate and fast maximum likelihood phylogenetic network inference. *Bioinformatics*, 38(15):3725–3733, 2022.

[32] Wayne P Maddison. Gene trees in species trees. *Systematic biology*, 46(3):523–536, 1997.

[33] Alexey Markin, Tavis K Anderson, Venkata Sai Krishna Teja Vadali, and Oliver Eulenstein. Robinson-foulds reticulation networks. In *Proceedings of the 10th ACM international conference on bioinformatics, computational biology and health informatics*, pages 77–86, 2019.

[34] James Oldman, Taoyang Wu, Leo Van Iersel, and Vincent Moulton. Trilonet: piecing together small networks to reconstruct reticulate evolutionary histories. *Molecular biology and evolution*, 33(8):2151–2162, 2016.

[35] Fabio Pardi and Celine Scornavacca. Reconstructible phylogenetic networks: do not distinguish the indistinguishable. *PLoS computational biology*, 11(4):e1004135, 2015.

[36] James B Pease and Matthew W Hahn. Detection and polarization of introgression in a five-taxon phylogeny. *Systematic biology*, 64(4):651–662, 2015.

[37] Joseph Pickrell and Jonathan Pritchard. Inference of population splits and mixtures from genome-wide allele frequency data. *Nature Precedings*, pages 1–1, 2012.

[38] Charles-Elie Rabier, Vincent Berry, Marnus Stoltz, Joao D Santos, Wensheng Wang, Jean-Christophe Glaszmann, Fabio Pardi, and Celine Scornavacca. On the inference of complex phylogenetic networks by markov chain monte-carlo. *PLoS Computational Biology*, 17(9):e1008380, 2021.

[39] Bruce Rannala and Ziheng Yang. Bayes estimation of species divergence times and ancestral population sizes using dna sequences from multiple loci. *Genetics*, 164(4):1645–1656, 2003.

[40] John A Rhodes and Seth Sullivant. Identifiability of large phylogenetic mixture models. *Bulletin of mathematical biology*, 74:212–231, 2012.

[41] Francesc Rosselló and Gabriel Valiente. All that glisters is not galled. *Mathematical biosciences*, 221(1):54–59, 2009.

[42] Claudia Solís-Lemus and Cécile Ané. Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. *PLoS Genet.*, 12(3):e1005896, March 2016.

[43] Claudia Solís-Lemus, Paul Bastide, and Cécile Ané. Phylonetworks: a package for phylogenetic networks. *Molecular biology and evolution*, 34(12):3292–3298, 2017.

[44] Claudia Solis-Lemus, Arrigo Coen, and Cecile Ane. On the identifiability of phylogenetic networks under a pseudolikelihood model. October 2020.

[45] Ming Tan, Haixia Long, Bo Liao, Zhi Cao, Dawei Yuan, Geng Tian, Jujuan Zhuang, and Jialiang Yang. Qs-net: Reconstructing phylogenetic networks based on quartet and sextet. *Frontiers in Genetics*, 10:607, 2019.

[46] Dingqiao Wen, Yun Yu, Jiafan Zhu, and Luay Nakhleh. Inferring phylogenetic networks using phylonet. *Systematic biology*, 67(4):735–740, 2018.

[47] Matthieu Willems, Nadia Tahiri, and Vladimir Makarenkov. A new efficient algorithm for inferring explicit hybridization networks following the neighbor-joining principle. *Journal of bioinformatics and computational biology*, 12(05):1450024, 2014.

[48] Jingcheng Xu and Cécile Ané. Identifiability of local and global features of phylogenetic networks from average distances. *Journal of Mathematical Biology*, 86(1):12, 2023.

[49] Jialiang Yang, Stefan Grünewald, and Xiu-Feng Wan. Quartet-net: a quartet-based method to reconstruct phylogenetic networks. *Molecular biology and evolution*, 30(5):1206–1217, 2013.

[50] Ziheng Yang. Maximum likelihood phylogenetic estimation from dna sequences with variable rates over sites: approximate methods. *Journal of Molecular evolution*, 39(3):306–314, 1994.

[51] Yun Yu, James H. Degnan, and Luay Nakhleh. The probability of a gene tree topology within a phylogenetic network with applications to hybridization detection. *PLoS genetics*, 8(4):e1002660, jan 2012.

[52] Yun Yu, Jianrong Dong, Kevin J Liu, and Luay Nakhleh. Maximum Likelihood Inference of Reticulate Evolutionary Histories. *PNAS*, 111(46):16448–16453, 2014.

[53] Yun Yu and Luay Nakhleh. A maximum pseudo-likelihood approach for phylogenetic networks. *BMC genomics*, 16(10):1–10, 2015.

[54] Chi Zhang, Huw A Ogilvie, Alexei J Drummond, and Tanja Stadler. Bayesian inference of species networks from multilocus sequence data. *Molecular biology and evolution*, 35(2):504–517, 2018.

A Phylogenetic networks as algebraic varieties

A.1 CF equations for $\mathcal{N}_{down}$: $CF(\mathcal{N}_{down}, \boldsymbol{z}, \boldsymbol{\gamma})$

n	CF Formula
$(0,0,2,2)$	$1-\frac{2}{3}z_2z_{23}z_3$ $\frac{1}{3}z_2z_{23}z_3$ $\frac{1}{3}z_2z_{23}z_3$
$(0,1,2,1)$	$1-\frac{2}{3}z_{23}z_2$ $\frac{1}{3}z_{23}z_2$ $\frac{1}{3}z_{23}z_2$
$(0,1,1,2)$	$1-\frac{2}{3}z_3$ $\frac{1}{3}z_3$ $\frac{1}{3}z_3$
$(0,2,2,0)$	$1-\frac{2}{3}z_2z_{23}z_{13}z_1$ $\frac{1}{3}z_2z_{23}z_{13}z_1$ $\frac{1}{3}z_2z_{23}z_{13}z_1$
$(0,2,1,1)$	$1-\frac{2}{3}z_{13}z_1$ $\frac{1}{3}z_{13}z_1$ $\frac{1}{3}z_{13}z_1$
$(0,2,0,2)$	$1-\frac{2}{3}z_3z_{13}z_1$ $\frac{1}{3}z_3z_{13}z_1$ $\frac{1}{3}z_3z_{13}z_1$
$(1,0,2,1)$	$(1-\gamma)\left(1-\frac{2}{3}z_{23}z_2\right)+\gamma\left(1-\frac{2}{3}z_2\right)$ $(1-\gamma)\frac{1}{3}z_{23}z_2+\gamma\frac{1}{3}z_2$ $(1-\gamma)\frac{1}{3}z_{23}z_2+\gamma\frac{1}{3}z_2$
$(1,0,1,2)$	$(1-\gamma)\left(1-\frac{2}{3}z_3\right)+\gamma\left(1-\frac{2}{3}z_{23}z_3\right)$ $(1-\gamma)\frac{1}{3}z_3+\gamma\frac{1}{3}z_{23}z_3$ $(1-\gamma)\frac{1}{3}z_3+\gamma\frac{1}{3}z_{23}z_3$
$(1,1,2,0)$	$(1-\gamma)\left(1-\frac{2}{3}z_{13}z_{23}z_2\right)+\gamma\left(1-\frac{2}{3}z_2\right)$ $(1-\gamma)\frac{1}{3}z_{13}z_{23}z_2+\gamma\frac{1}{3}z_2$ $(1-\gamma)\frac{1}{3}z_{13}z_{23}z_2+\gamma\frac{1}{3}z_2$

$(1,1,0,2)$	$(1-\gamma)\left(1-\frac{2}{3}z_{13}z_3\right)+\gamma\left(1-\frac{2}{3}z_3\right)$ $(1-\gamma)\frac{1}{3}z_{13}z_3+\gamma\frac{1}{3}z_3$ $(1-\gamma)\frac{1}{3}z_{13}z_3+\gamma\frac{1}{3}z_3$
$(1,2,1,0)$	$(1-\gamma)\left(1-\frac{2}{3}z_1\right)+\gamma\left(1-\frac{2}{3}z_{23}z_{13}z_1\right)$ $(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{23}z_{13}z_1$ $(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{23}z_{13}z_1$
$(1,2,0,1)$	$(1-\gamma)\left(1-\frac{2}{3}z_1\right)+\gamma\left(1-\frac{2}{3}z_{13}z_1\right)$ $(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{13}z_1$ $(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{13}z_1$
$(2,0,2,0)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_2z_0z_{01}z_{13}z_{23}\right)+2\gamma(1-\gamma)\left(1-\frac{2}{3}z_2z_0\right)+\gamma^2\left(1-\frac{2}{3}z_2z_0z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_2z_0z_{01}z_{13}z_{23}+2\gamma(1-\gamma)\frac{1}{3}z_2z_0+\gamma^2\frac{1}{3}z_2z_0z_{02}$ $(1-\gamma)^2\frac{1}{3}z_2z_0z_{01}z_{13}z_{23}+2\gamma(1-\gamma)\frac{1}{3}z_2z_0+\gamma^2\frac{1}{3}z_2z_0z_{02}$
$(2,0,1,1)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{13}z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{23}\right)+\gamma^2\left(1-\frac{2}{3}z_0z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_0z_{13}z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}\right)+\gamma^2\frac{1}{3}z_0z_{02}$ $(1-\gamma)^2\frac{1}{3}z_0z_{13}z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}\right)+\gamma^2\frac{1}{3}z_0z_{02}$
$(2,0,0,2)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_3z_0z_{13}z_{01}\right)+2\gamma(1-\gamma)\left(1-\frac{2}{3}z_3z_0\right)+\gamma^2\left(1-\frac{2}{3}z_3z_0z_{23}z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_3z_0z_{13}z_{01}+2\gamma(1-\gamma)\frac{1}{3}z_3z_0+\gamma^2\frac{1}{3}z_3z_0z_{23}z_{02}$ $(1-\gamma)^2\frac{1}{3}z_3z_0z_{13}z_{01}+2\gamma(1-\gamma)\frac{1}{3}z_3z_0+\gamma^2\frac{1}{3}z_3z_0z_{23}z_{02}$
$(2,1,1,0)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{23}z_{13}\right)+\gamma^2\left(1-\frac{2}{3}z_0z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}$
$(2,1,0,1)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{13}\right)+\gamma^2\left(1-\frac{2}{3}z_0z_{02}z_{23}\right)$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}z_{23}$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}z_{23}$
$(2,2,0,0)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_1z_0z_{01}\right)+2\gamma(1-\gamma)\left(1-\frac{2}{3}z_1z_0\right)+\gamma^2\left(1-\frac{2}{3}z_1z_0z_{02}z_{23}z_{13}\right)$ $(1-\gamma)^2\frac{1}{3}z_1z_0z_{01}+2\gamma(1-\gamma)\frac{1}{3}z_1z_0+\gamma^2\frac{1}{3}z_1z_0z_{02}z_{23}z_{13}$ $(1-\gamma)^2\frac{1}{3}z_1z_0z_{01}+2\gamma(1-\gamma)\frac{1}{3}z_1z_0+\gamma^2\frac{1}{3}z_1z_0z_{02}z_{23}z_{13}$
$(1,1,1,1)$	$(1-\gamma)\left(1-\frac{2}{3}z_{13}\right)+\gamma\frac{1}{3}z_{23}$

	$(1-\gamma)\frac{1}{3}z_{13}+\gamma\left(1-\frac{2}{3}z_{23}\right)$ $(1-\gamma)\frac{1}{3}z_{13}+\gamma\frac{1}{3}z_{23}$

Table 1: CF equations for $\mathcal{N}_{down}$: $CF(\mathcal{N}_{down}, \boldsymbol{z}, \boldsymbol{\gamma})$

A.2 Definition of CF equations for network obtained by the rotation of hybrid in hybridization cycle

We start by noticing that there is a correspondence between a quartet in the network $\mathcal{N}_{down}$ and a quartet in any other network. For example, in Figure 7, we want to match the equations corresponding to $n = (0, 1, 2, 1)$ in $\mathcal{N}_{down}$ and $\mathcal{N}_{right}$. We can see that the equations for the quartet $n = (0, 1, 2, 1)$ in $\mathcal{N}_{right}$ can be obtained from the equations in quartet $n = (2, 0, 1, 1)$ in $\mathcal{N}_{down}$.

Therefore, in order to get the equations for $\mathcal{N}_{right}$, we need to identify to which quartets they correspond in the $\mathcal{N}_{down}$.

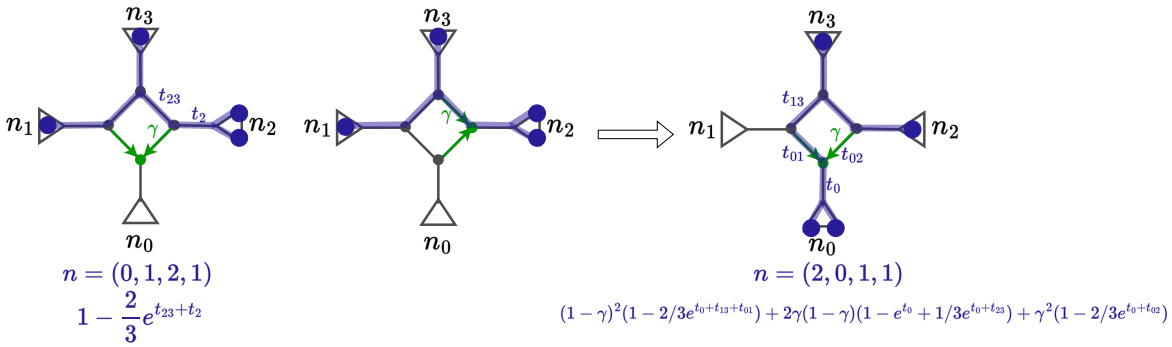

Figure 7: Correspondence between the quartet $n = (0, 1, 2, 1)$ in $\mathcal{N}_{down}$ (left), the same quartet in $\mathcal{N}_{right}$ (center) and the quartet $n = (2, 0, 1, 1)$ in $\mathcal{N}_{down}$ (right).

Note that only the last quartet $(1, 1, 1, 1)$ is not a mere rotation. We have to compute the CFs for the specific network (Figure 8).

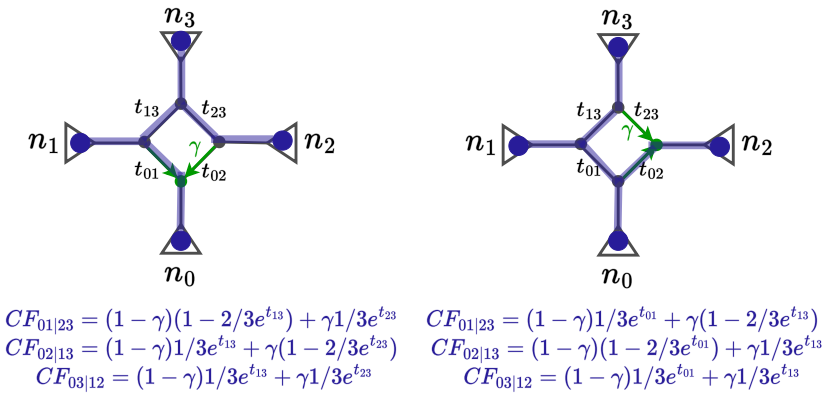

Figure 8: CF equations for quartet $(1, 1, 1, 1)$ comparing the $\mathcal{N}_{down}$ and $\mathcal{N}_{right}$ networks.

A.3 CF equations for $\mathcal{N}_{right}$: $CF(\mathcal{N}_{right}, \boldsymbol{z}, \boldsymbol{\gamma})$

$n \in \mathcal{N}_{right}$	$n \in \mathcal{N}_{down}$	CF Formula
$(0,0,2,2)$	$(2,0,2,0)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_2z_0z_{01}z_{13}z_{23}\right)+2\gamma(1-\gamma)\left(1-\frac{2}{3}z_2z_0\right)+\gamma^2\left(1-\frac{2}{3}z_2z_0z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_2z_0z_{01}z_{13}z_{23}+2\gamma(1-\gamma)\frac{1}{3}z_2z_0+\gamma^2\frac{1}{3}z_2z_0z_{02}$ $(1-\gamma)^2\frac{1}{3}z_2z_0z_{01}z_{13}z_{23}+2\gamma(1-\gamma)\frac{1}{3}z_2z_0+\gamma^2\frac{1}{3}z_2z_0z_{02}$
$(0,1,2,1)$	$(2,0,1,1)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{13}z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{23}\right)+\gamma^2\left(1-\frac{2}{3}z_0z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_0z_{13}z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}\right)+\gamma^2\frac{1}{3}z_0z_{02}$ $(1-\gamma)^2\frac{1}{3}z_0z_{13}z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}\right)+\gamma^2\frac{1}{3}z_0z_{02}$
$(0,1,1,2)$	$(1,0,2,1)$	$(1-\gamma)\left(1-\frac{2}{3}z_{23}z_2\right)+\gamma\left(1-\frac{2}{3}z_2\right)$ $(1-\gamma)\frac{1}{3}z_{23}z_2+\gamma\frac{1}{3}z_2$ $(1-\gamma)\frac{1}{3}z_{23}z_2+\gamma\frac{1}{3}z_2$
$(0,2,2,0)$	$(2,0,0,2)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_3z_0z_{13}z_{01}\right)+2\gamma(1-\gamma)\left(1-\frac{2}{3}z_3z_0\right)+\gamma^2\left(1-\frac{2}{3}z_3z_0z_{23}z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_3z_0z_{13}z_{01}+2\gamma(1-\gamma)\frac{1}{3}z_3z_0+\gamma^2\frac{1}{3}z_3z_0z_{23}z_{02}$ $(1-\gamma)^2\frac{1}{3}z_3z_0z_{13}z_{01}+2\gamma(1-\gamma)\frac{1}{3}z_3z_0+\gamma^2\frac{1}{3}z_3z_0z_{23}z_{02}$
$(0,2,1,1)$	$(1,0,1,2)$	$(1-\gamma)\left(1-\frac{2}{3}z_3\right)+\gamma\left(1-\frac{2}{3}z_{23}z_3\right)$ $(1-\gamma)\frac{1}{3}z_3+\gamma\frac{1}{3}z_{23}z_3$ $(1-\gamma)\frac{1}{3}z_3+\gamma\frac{1}{3}z_{23}z_3$
$(0,2,0,2)$	$(0,0,2,2)$	$1-\frac{2}{3}z_2z_{23}z_3$ $\frac{1}{3}z_2z_{23}z_3$ $\frac{1}{3}z_2z_{23}z_3$
$(1,0,2,1)$	$(2,1,1,0)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{23}z_{13}\right)+\gamma^2\left(1-\frac{2}{3}z_0z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}$
$(1,0,1,2)$	$(1,1,2,0)$	$(1-\gamma)\left(1-\frac{2}{3}z_{13}z_{23}z_2\right)+\gamma\left(1-\frac{2}{3}z_2\right)$ $(1-\gamma)\frac{1}{3}z_{13}z_{23}z_2+\gamma\frac{1}{3}z_2$ $(1-\gamma)\frac{1}{3}z_{13}z_{23}z_2+\gamma\frac{1}{3}z_2$
$(1,1,2,0)$	$(2,1,0,1)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{13}\right)+\gamma^2\left(1-\frac{2}{3}z_0z_{02}z_{23}\right)$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}z_{23}$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}z_{23}$

$(1,1,0,2)$	$(0,1,2,1)$	$1-\frac{2}{3}z_{23}z_2$ $\frac{1}{3}z_{23}z_2$ $\frac{1}{3}z_{23}z_2$
$(1,2,1,0)$	$(1,1,0,2)$	$(1-\gamma)\left(1-\frac{2}{3}z_{13}z_3\right)+\gamma\left(1-\frac{2}{3}z_3\right)$ $(1-\gamma)\frac{1}{3}z_{13}z_3+\gamma\frac{1}{3}z_3$ $(1-\gamma)\frac{1}{3}z_{13}z_3+\gamma\frac{1}{3}z_3$
$(1,2,0,1)$	$(0,1,1,2)$	$1-\frac{2}{3}z_3$ $\frac{1}{3}z_3$ $\frac{1}{3}z_3$
$(2,0,2,0)$	$(2,2,0,0)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_1z_0z_{01}\right)+2\gamma(1-\gamma)\left(1-\frac{2}{3}z_1z_0\right)+\gamma^2\left(1-\frac{2}{3}z_1z_0z_{02}z_{23}z_{13}\right)$ $(1-\gamma)^2\frac{1}{3}z_1z_0z_{01}+2\gamma(1-\gamma)\frac{1}{3}z_1z_0+\gamma^2\frac{1}{3}z_1z_0z_{02}z_{23}z_{13}$ $(1-\gamma)^2\frac{1}{3}z_1z_0z_{01}+2\gamma(1-\gamma)\frac{1}{3}z_1z_0+\gamma^2\frac{1}{3}z_1z_0z_{02}z_{23}z_{13}$
$(2,0,1,1)$	$(1,2,1,0)$	$(1-\gamma)\left(1-\frac{2}{3}z_1\right)+\gamma\left(1-\frac{2}{3}z_{23}z_{13}z_1\right)$ $(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{23}z_{13}z_1$ $(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{23}z_{13}z_1$
$(2,0,0,2)$	$(0,2,2,0)$	$1-\frac{2}{3}z_2z_{23}z_{13}z_1$ $\frac{1}{3}z_2z_{23}z_{13}z_1$ $\frac{1}{3}z_2z_{23}z_{13}z_1$
$(2,1,1,0)$	$(1,2,0,1)$	$(1-\gamma)\left(1-\frac{2}{3}z_1\right)+\gamma\left(1-\frac{2}{3}z_{13}z_1\right)$ $(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{13}z_1$ $(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{13}z_1$
$(2,1,0,1)$	$(0,2,1,1)$	$1-\frac{2}{3}z_{13}z_1$ $\frac{1}{3}z_{13}z_1$ $\frac{1}{3}z_{13}z_1$
$(2,2,0,0)$	$(0,2,0,2)$	$1-\frac{2}{3}z_3z_{13}z_1$ $\frac{1}{3}z_3z_{13}z_1$ $\frac{1}{3}z_3z_{13}z_1$
$(1,1,1,1)$		$(1-\gamma)\frac{1}{3}z_{01}+\gamma(1-\frac{2}{3}z_{13})$ $(1-\gamma)(1-\frac{2}{3}z_{01})+\gamma\frac{1}{3}z_{13}$ $(1-\gamma)\frac{1}{3}z_{01}+\gamma\frac{1}{3}z_{13}$

Table 2: CF equations for $\mathcal{N}_{right}$: $CF(\mathcal{N}_{right}, \boldsymbol{z}, \boldsymbol{\gamma})$

A.4 CF equations for $\mathcal{N}_{left}$: $CF(\mathcal{N}_{left}, \boldsymbol{z}, \boldsymbol{\gamma})$

$n \in \mathcal{N}_{left}$	$n \in \mathcal{N}_{down}$	CF Formula
$(0,0,2,2)$	$(0,2,0,2)$	$1-\frac{2}{3}z_3z_{13}z_1$ $\frac{1}{3}z_3z_{13}z_1$ $\frac{1}{3}z_3z_{13}z_1$
$(0,1,2,1)$	$(1,1,0,2)$	$(1-\gamma)\left(1-\frac{2}{3}z_{13}z_3\right)+\gamma\left(1-\frac{2}{3}z_3\right)$ $(1-\gamma)\frac{1}{3}z_{13}z_3+\gamma\frac{1}{3}z_3$ $(1-\gamma)\frac{1}{3}z_{13}z_3+\gamma\frac{1}{3}z_3$
$(0,1,1,2)$	$(1,2,0,1)$	$(1-\gamma)\left(1-\frac{2}{3}z_1\right)+\gamma\left(1-\frac{2}{3}z_{13}z_1\right)$ $(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{13}z_1$ $(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{13}z_1$
$(0,2,2,0)$	$(2,0,0,2)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_3z_0z_{13}z_{01}\right)+2\gamma(1-\gamma)\left(1-\frac{2}{3}z_3z_0\right)+\gamma^2\left(1-\frac{2}{3}z_3z_0z_{23}z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_3z_0z_{13}z_{01}+2\gamma(1-\gamma)\frac{1}{3}z_3z_0+\gamma^2\frac{1}{3}z_3z_0z_{23}z_{02}$ $(1-\gamma)^2\frac{1}{3}z_3z_0z_{13}z_{01}+2\gamma(1-\gamma)\frac{1}{3}z_3z_0+\gamma^2\frac{1}{3}z_3z_0z_{23}z_{02}$
$(0,2,1,1)$	$(2,1,0,1)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{13}\right)+\gamma^2\left(1-\frac{2}{3}z_0z_{02}z_{23}\right)$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}z_{23}$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}z_{23}$
$(0,2,0,2)$	$(2,2,0,0)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_1z_0z_{01}\right)+2\gamma(1-\gamma)\left(1-\frac{2}{3}z_1z_0\right)+\gamma^2\left(1-\frac{2}{3}z_1z_0z_{02}z_{23}z_{13}\right)$ $(1-\gamma)^2\frac{1}{3}z_1z_0z_{01}+2\gamma(1-\gamma)\frac{1}{3}z_1z_0+\gamma^2\frac{1}{3}z_1z_0z_{02}z_{23}z_{13}$ $(1-\gamma)^2\frac{1}{3}z_1z_0z_{01}+2\gamma(1-\gamma)\frac{1}{3}z_1z_0+\gamma^2\frac{1}{3}z_1z_0z_{02}z_{23}z_{13}$
$(1,0,2,1)$	$(0,1,1,2)$	$1-\frac{2}{3}z_3$ $\frac{1}{3}z_3$ $\frac{1}{3}z_3$
$(1,0,1,2)$	$(0,2,1,1)$	$1-\frac{2}{3}z_{13}z_1$ $\frac{1}{3}z_{13}z_1$ $\frac{1}{3}z_{13}z_1$
$(1,1,2,0)$	$(1,0,1,2)$	$(1-\gamma)\left(1-\frac{2}{3}z_3\right)+\gamma\left(1-\frac{2}{3}z_{23}z_3\right)$ $(1-\gamma)\frac{1}{3}z_3+\gamma\frac{1}{3}z_{23}z_3$ $(1-\gamma)\frac{1}{3}z_3+\gamma\frac{1}{3}z_{23}z_3$
$(1,1,0,2)$	$(1,2,1,0)$	$(1-\gamma)\left(1-\frac{2}{3}z_1\right)+\gamma\left(1-\frac{2}{3}z_{23}z_{13}z_1\right)$

		$(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{23}z_{13}z_1$ $(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{23}z_{13}z_1$
$(1,2,1,0)$	$(2,0,1,1)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{13}z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{23}\right)+\gamma^2\left(1-\frac{2}{3}z_0z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_0z_{13}z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}\right)+\gamma^2\frac{1}{3}z_0z_{02}$ $(1-\gamma)^2\frac{1}{3}z_0z_{13}z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}\right)+\gamma^2\frac{1}{3}z_0z_{02}$
$(1,2,0,1)$	$(2,1,1,0)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{23}z_{13}\right)+\gamma^2\left(1-\frac{2}{3}z_0z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}$
$(2,0,2,0)$	$(0,0,2,2)$	$1-\frac{2}{3}z_2z_{23}z_3$ $\frac{1}{3}z_2z_{23}z_3$ $\frac{1}{3}z_2z_{23}z_3$
$(2,0,1,1)$	$(0,1,2,1)$	$1-\frac{2}{3}z_{23}z_2$ $\frac{1}{3}z_{23}z_2$ $\frac{1}{3}z_{23}z_2$
$(2,0,0,2)$	$(0,2,2,0)$	$1-\frac{2}{3}z_2z_{23}z_{13}z_1$ $\frac{1}{3}z_2z_{23}z_{13}z_1$ $\frac{1}{3}z_2z_{23}z_{13}z_1$
$(2,1,1,0)$	$(1,0,2,1)$	$(1-\gamma)\left(1-\frac{2}{3}z_{23}z_2\right)+\gamma\left(1-\frac{2}{3}z_2\right)$ $(1-\gamma)\frac{1}{3}z_{23}z_2+\gamma\frac{1}{3}z_2$ $(1-\gamma)\frac{1}{3}z_{23}z_2+\gamma\frac{1}{3}z_2$
$(2,1,0,1)$	$(1,1,2,0)$	$(1-\gamma)\left(1-\frac{2}{3}z_{13}z_{23}z_2\right)+\gamma\left(1-\frac{2}{3}z_2\right)$ $(1-\gamma)\frac{1}{3}z_{13}z_{23}z_2+\gamma\frac{1}{3}z_2$ $(1-\gamma)\frac{1}{3}z_{13}z_{23}z_2+\gamma\frac{1}{3}z_2$
$(2,2,0,0)$	$(2,0,2,0)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_2z_0z_{01}z_{13}z_{23}\right)+2\gamma(1-\gamma)\left(1-\frac{2}{3}z_2z_0\right)+\gamma^2\left(1-\frac{2}{3}z_2z_0z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_2z_0z_{01}z_{13}z_{23}+2\gamma(1-\gamma)\frac{1}{3}z_2z_0+\gamma^2\frac{1}{3}z_2z_0z_{02}$ $(1-\gamma)^2\frac{1}{3}z_2z_0z_{01}z_{13}z_{23}+2\gamma(1-\gamma)\frac{1}{3}z_2z_0+\gamma^2\frac{1}{3}z_2z_0z_{02}$
$(1,1,1,1)$		$(1-\gamma)\frac{1}{3}z_{23}+\gamma(1-\frac{2}{3}z_{02})$ $(1-\gamma)(1-\frac{2}{3}z_{23})+\gamma\frac{1}{3}z_{02}$ $(1-\gamma)\frac{1}{3}z_{23}+\gamma\frac{1}{3}z_{02}$

Table 3: CF equations for $\mathcal{N}_{left}$: $CF(\mathcal{N}_{left}, \boldsymbol{z}, \boldsymbol{\gamma})$

A.5 CF equations for $\mathcal{N}_{up}$: $CF(\mathcal{N}_{up}, \boldsymbol{z}, \boldsymbol{\gamma})$

$n \in \mathcal{N}_{up}$	$n \in \mathcal{N}_{down}$	CF Formula
$(0,0,2,2)$	$(2,2,0,0)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_1z_0z_{01}\right)+2\gamma(1-\gamma)\left(1-\frac{2}{3}z_1z_0\right)+\gamma^2\left(1-\frac{2}{3}z_1z_0z_{02}z_{23}z_{13}\right)$ $(1-\gamma)^2\frac{1}{3}z_1z_0z_{01}+2\gamma(1-\gamma)\frac{1}{3}z_1z_0+\gamma^2\frac{1}{3}z_1z_0z_{02}z_{23}z_{13}$ $(1-\gamma)^2\frac{1}{3}z_1z_0z_{01}+2\gamma(1-\gamma)\frac{1}{3}z_1z_0+\gamma^2\frac{1}{3}z_1z_0z_{02}z_{23}z_{13}$
$(0,1,2,1)$	$(1,2,1,0)$	$(1-\gamma)\left(1-\frac{2}{3}z_1\right)+\gamma\left(1-\frac{2}{3}z_{23}z_{13}z_1\right)$ $(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{23}z_{13}z_1$ $(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{23}z_{13}z_1$
$(0,1,1,2)$	$(2,1,1,0)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{23}z_{13}\right)+\gamma^2\left(1-\frac{2}{3}z_0z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}$
$(0,2,2,0)$	$(0,2,2,0)$	$1-\frac{2}{3}z_2z_{23}z_{13}z_1$ $\frac{1}{3}z_2z_{23}z_{13}z_1$ $\frac{1}{3}z_2z_{23}z_{13}z_1$
$(0,2,1,1)$	$(1,1,2,0)$	$(1-\gamma)\left(1-\frac{2}{3}z_{13}z_{23}z_2\right)+\gamma\left(1-\frac{2}{3}z_2\right)$ $(1-\gamma)\frac{1}{3}z_{13}z_{23}z_2+\gamma\frac{1}{3}z_2$ $(1-\gamma)\frac{1}{3}z_{13}z_{23}z_2+\gamma\frac{1}{3}z_2$
$(0,2,0,2)$	$(2,0,2,0)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_2z_0z_{01}z_{13}z_{23}\right)+2\gamma(1-\gamma)\left(1-\frac{2}{3}z_2z_0\right)+\gamma^2\left(1-\frac{2}{3}z_2z_0z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_2z_0z_{01}z_{13}z_{23}+2\gamma(1-\gamma)\frac{1}{3}z_2z_0+\gamma^2\frac{1}{3}z_2z_0z_{02}$ $(1-\gamma)^2\frac{1}{3}z_2z_0z_{01}z_{13}z_{23}+2\gamma(1-\gamma)\frac{1}{3}z_2z_0+\gamma^2\frac{1}{3}z_2z_0z_{02}$
$(1,0,2,1)$	$(1,2,0,1)$	$(1-\gamma)\left(1-\frac{2}{3}z_1\right)+\gamma\left(1-\frac{2}{3}z_{13}z_1\right)$ $(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{13}z_1$ $(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{13}z_1$
$(1,0,1,2)$	$(2,1,0,1)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{13}\right)+\gamma^2\left(1-\frac{2}{3}z_0z_{02}z_{23}\right)$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}z_{23}$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}z_{23}$
$(1,1,2,0)$	$(0,2,1,1)$	$1-\frac{2}{3}z_{13}z_1$ $\frac{1}{3}z_{13}z_1$ $\frac{1}{3}z_{13}z_1$

$(1,1,0,2)$	$(2,0,1,1)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{13}z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{23}\right)+\gamma^2\left(1-\frac{2}{3}z_0z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_0z_{13}z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}\right)+\gamma^2\frac{1}{3}z_0z_{02}$ $(1-\gamma)^2\frac{1}{3}z_0z_{13}z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}\right)+\gamma^2\frac{1}{3}z_0z_{02}$
$(1,2,1,0)$	$(0,1,2,1)$	$1-\frac{2}{3}z_{23}z_2$ $\frac{1}{3}z_{23}z_2$ $\frac{1}{3}z_{23}z_2$
$(1,2,0,1)$	$(1,0,2,1)$	$(1-\gamma)\left(1-\frac{2}{3}z_{23}z_2\right)+\gamma\left(1-\frac{2}{3}z_2\right)$ $(1-\gamma)\frac{1}{3}z_{23}z_2+\gamma\frac{1}{3}z_2$ $(1-\gamma)\frac{1}{3}z_{23}z_2+\gamma\frac{1}{3}z_2$
$(2,0,2,0)$	$(0,2,0,2)$	$1-\frac{2}{3}z_3z_{13}z_1$ $\frac{1}{3}z_3z_{13}z_1$ $\frac{1}{3}z_3z_{13}z_1$
$(2,0,1,1)$	$(1,1,0,2)$	$(1-\gamma)\left(1-\frac{2}{3}z_{13}z_3\right)+\gamma\left(1-\frac{2}{3}z_3\right)$ $(1-\gamma)\frac{1}{3}z_{13}z_3+\gamma\frac{1}{3}z_3$ $(1-\gamma)\frac{1}{3}z_{13}z_3+\gamma\frac{1}{3}z_3$
$(2,0,0,2)$	$(2,0,0,2)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_3z_0z_{13}z_{01}\right)+2\gamma(1-\gamma)\left(1-\frac{2}{3}z_3z_0\right)+\gamma^2\left(1-\frac{2}{3}z_3z_0z_{23}z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_3z_0z_{13}z_{01}+2\gamma(1-\gamma)\frac{1}{3}z_3z_0+\gamma^2\frac{1}{3}z_3z_0z_{23}z_{02}$ $(1-\gamma)^2\frac{1}{3}z_3z_0z_{13}z_{01}+2\gamma(1-\gamma)\frac{1}{3}z_3z_0+\gamma^2\frac{1}{3}z_3z_0z_{23}z_{02}$
$(2,1,1,0)$	$(0,1,1,2)$	$1-\frac{2}{3}z_3$ $\frac{1}{3}z_3$ $\frac{1}{3}z_3$
$(2,1,0,1)$	$(1,0,1,2)$	$(1-\gamma)\left(1-\frac{2}{3}z_3\right)+\gamma\left(1-\frac{2}{3}z_{23}z_3\right)$ $(1-\gamma)\frac{1}{3}z_3+\gamma\frac{1}{3}z_{23}z_3$ $(1-\gamma)\frac{1}{3}z_3+\gamma\frac{1}{3}z_{23}z_3$
$(2,2,0,0)$	$(0,0,2,2)$	$1-\frac{2}{3}z_2z_{23}z_3$ $\frac{1}{3}z_2z_{23}z_3$ $\frac{1}{3}z_2z_{23}z_3$
$(1,1,1,1)$		$(1-\gamma)(1-\frac{2}{3}z_{02})+\gamma\frac{1}{3}z_{01}$ $(1-\gamma)\frac{1}{3}z_{02}+\gamma(1-\frac{2}{3}z_{01})$ $(1-\gamma)\frac{1}{3}z_{02}+\gamma\frac{1}{3}z_{01}$

Table 4: CF equations for $\mathcal{N}_{up}$: $CF(\mathcal{N}_{up}, \boldsymbol{z}, \boldsymbol{\gamma})$

B Phylogenetic networks as algebraic varieties: $n = 6$ taxa

B.1 CF equations for 4-cycle network $\mathcal{N}_{down}$ for $N = 1212$ (Diamond 6)

n	CF Formula
$(0,1,1,2)$	$1-\frac{2}{3}z_3$ $\frac{1}{3}z_3$ $\frac{1}{3}z_3$
$(0,2,1,1)$	$1-\frac{2}{3}z_{13}z_1$ $\frac{1}{3}z_{13}z_1$ $\frac{1}{3}z_{13}z_1$
$(0,2,0,2)$	$1-\frac{2}{3}z_3z_{13}z_1$ $\frac{1}{3}z_3z_{13}z_1$ $\frac{1}{3}z_3z_{13}z_1$
$(1,0,1,2)$	$(1-\gamma)\left(1-\frac{2}{3}z_3\right)+\gamma\left(1-\frac{2}{3}z_{23}z_3\right)$ $(1-\gamma)\frac{1}{3}z_3+\gamma\frac{1}{3}z_{23}z_3$ $(1-\gamma)\frac{1}{3}z_3+\gamma\frac{1}{3}z_{23}z_3$
$(1,1,0,2)$	$(1-\gamma)\left(1-\frac{2}{3}z_{13}z_3\right)+\gamma\left(1-\frac{2}{3}z_3\right)$ $(1-\gamma)\frac{1}{3}z_{13}z_3+\gamma\frac{1}{3}z_3$ $(1-\gamma)\frac{1}{3}z_{13}z_3+\gamma\frac{1}{3}z_3$
$(1,2,1,0)$	$(1-\gamma)\left(1-\frac{2}{3}z_1\right)+\gamma\left(1-\frac{2}{3}z_{23}z_{13}z_1\right)$ $(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{23}z_{13}z_1$ $(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{23}z_{13}z_1$
$(1,2,0,1)$	$(1-\gamma)\left(1-\frac{2}{3}z_1\right)+\gamma\left(1-\frac{2}{3}z_{13}z_1\right)$ $(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{13}z_1$ $(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{13}z_1$
$(1,1,1,1)$	$(1-\gamma)\left(1-\frac{2}{3}z_{13}\right)+\gamma\frac{1}{3}z_{23}$ $(1-\gamma)\frac{1}{3}z_{13}+\gamma\left(1-\frac{2}{3}z_{23}\right)$ $(1-\gamma)\frac{1}{3}z_{13}+\gamma\frac{1}{3}z_{23}$

Table 5: CF equations for 4-cycle network $\mathcal{N}_{down}$ for $N = 1212$ (Diamond 6)

B.2 CF equations for 4-cycle network $\mathcal{N}_{right}$ for $N = 1122$ (Diamond 9)

$n \in \mathcal{N}_{right}$	$n \in \mathcal{N}_{down}$	CF Formula
$(0,1,1,2)$	$(1,0,2,1)$	$(1-\gamma)\left(1-\frac{2}{3}z_{23}z_2\right)+\gamma\left(1-\frac{2}{3}z_2\right)$ $(1-\gamma)\frac{1}{3}z_{23}z_2+\gamma\frac{1}{3}z_2$ $(1-\gamma)\frac{1}{3}z_{23}z_2+\gamma\frac{1}{3}z_2$
$(0,2,1,1)$	$(1,0,1,2)$	$(1-\gamma)\left(1-\frac{2}{3}z_3\right)+\gamma\left(1-\frac{2}{3}z_{23}z_3\right)$ $(1-\gamma)\frac{1}{3}z_3+\gamma\frac{1}{3}z_{23}z_3$ $(1-\gamma)\frac{1}{3}z_3+\gamma\frac{1}{3}z_{23}z_3$
$(0,2,0,2)$	$(0,0,2,2)$	$1-\frac{2}{3}z_2z_{23}z_3$ $\frac{1}{3}z_2z_{23}z_3$ $\frac{1}{3}z_2z_{23}z_3$
$(1,0,1,2)$	$(1,1,2,0)$	$(1-\gamma)\left(1-\frac{2}{3}z_{13}z_{23}z_2\right)+\gamma\left(1-\frac{2}{3}z_2\right)$ $(1-\gamma)\frac{1}{3}z_{13}z_{23}z_2+\gamma\frac{1}{3}z_2$ $(1-\gamma)\frac{1}{3}z_{13}z_{23}z_2+\gamma\frac{1}{3}z_2$
$(1,1,0,2)$	$(0,1,2,1)$	$1-\frac{2}{3}z_{23}z_2$ $\frac{1}{3}z_{23}z_2$ $\frac{1}{3}z_{23}z_2$
$(1,2,1,0)$	$(1,1,0,2)$	$(1-\gamma)\left(1-\frac{2}{3}z_{13}z_3\right)+\gamma\left(1-\frac{2}{3}z_3\right)$ $(1-\gamma)\frac{1}{3}z_{13}z_3+\gamma\frac{1}{3}z_3$ $(1-\gamma)\frac{1}{3}z_{13}z_3+\gamma\frac{1}{3}z_3$
$(1,2,0,1)$	$(0,1,1,2)$	$1-\frac{2}{3}z_3$ $\frac{1}{3}z_3$ $\frac{1}{3}z_3$
$(1,1,1,1)$		$(1-\gamma)\frac{1}{3}z_{01}+\gamma(1-\frac{2}{3}z_{13})$ $(1-\gamma)(1-\frac{2}{3}z_{01})+\gamma\frac{1}{3}z_{13}$ $(1-\gamma)\frac{1}{3}z_{01}+\gamma\frac{1}{3}z_{13}$

Table 6: CF equations for 4-cycle network $\mathcal{N}_{right}$ for $N = 1122$ (Diamond 9)

B.3 CF equations for 4-cycle network $\mathcal{N}_{left}$ for $N = 2211$ (Diamond 7)

$n \in \mathcal{N}_{left}$	$n \in \mathcal{N}_{down}$	CF Formula
$(0,1,1,2)$	$(1,2,0,1)$	$(1-\gamma)\left(1-\frac{2}{3}z_1\right)+\gamma\left(1-\frac{2}{3}z_{13}z_1\right)$ $(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{13}z_1$ $(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{13}z_1$
$(0,2,1,1)$	$(2,1,0,1)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{13}\right)+\gamma^2\left(1-\frac{2}{3}z_0z_{02}z_{23}\right)$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}z_{23}$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}z_{23}$
$(0,2,0,2)$	$(2,2,0,0)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_1z_0z_{01}\right)+2\gamma(1-\gamma)\left(1-\frac{2}{3}z_1z_0\right)+\gamma^2\left(1-\frac{2}{3}z_1z_0z_{02}z_{23}z_{13}\right)$ $(1-\gamma)^2\frac{1}{3}z_1z_0z_{01}+2\gamma(1-\gamma)\frac{1}{3}z_1z_0+\gamma^2\frac{1}{3}z_1z_0z_{02}z_{23}z_{13}$ $(1-\gamma)^2\frac{1}{3}z_1z_0z_{01}+2\gamma(1-\gamma)\frac{1}{3}z_1z_0+\gamma^2\frac{1}{3}z_1z_0z_{02}z_{23}z_{13}$
$(1,0,1,2)$	$(0,2,1,1)$	$1-\frac{2}{3}z_{13}z_1$ $\frac{1}{3}z_{13}z_1$ $\frac{1}{3}z_{13}z_1$
$(1,1,0,2)$	$(1,2,1,0)$	$(1-\gamma)\left(1-\frac{2}{3}z_1\right)+\gamma\left(1-\frac{2}{3}z_{23}z_{13}z_1\right)$ $(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{23}z_{13}z_1$ $(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{23}z_{13}z_1$
$(1,2,1,0)$	$(2,0,1,1)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{13}z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{23}\right)+\gamma^2\left(1-\frac{2}{3}z_0z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_0z_{13}z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}\right)+\gamma^2\frac{1}{3}z_0z_{02}$ $(1-\gamma)^2\frac{1}{3}z_0z_{13}z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}\right)+\gamma^2\frac{1}{3}z_0z_{02}$
$(1,2,0,1)$	$(2,1,1,0)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{23}z_{13}\right)+\gamma^2\left(1-\frac{2}{3}z_0z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}$
$(1,1,1,1)$		$(1-\gamma)\frac{1}{3}z_{23}+\gamma(1-\frac{2}{3}z_{02})$ $(1-\gamma)(1-\frac{2}{3}z_{23})+\gamma\frac{1}{3}z_{02}$ $(1-\gamma)\frac{1}{3}z_{23}+\gamma\frac{1}{3}z_{02}$

Table 7: CF equations for 4-cycle network $\mathcal{N}_{left}$ for $N = 2211$ (Diamond 7)

B.4 CF equations for 4-cycle network $\mathcal{N}_{up}$ for $N = 2121$ (Diamond 8)

$n \in \mathcal{N}_{up}$	$n \in \mathcal{N}_{down}$	CF Formula
$(0,1,1,2)$	$(2,1,1,0)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{23}z_{13}\right)+\gamma^2\left(1-\frac{2}{3}z_0z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}$
$(0,2,1,1)$	$(1,1,2,0)$	$(1-\gamma)\left(1-\frac{2}{3}z_{13}z_{23}z_2\right)+\gamma\left(1-\frac{2}{3}z_2\right)$ $(1-\gamma)\frac{1}{3}z_{13}z_{23}z_2+\gamma\frac{1}{3}z_2$ $(1-\gamma)\frac{1}{3}z_{13}z_{23}z_2+\gamma\frac{1}{3}z_2$
$(0,2,0,2)$	$(2,0,2,0)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_2z_0z_{01}z_{13}z_{23}\right)+2\gamma(1-\gamma)\left(1-\frac{2}{3}z_2z_0\right)+\gamma^2\left(1-\frac{2}{3}z_2z_0z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_2z_0z_{01}z_{13}z_{23}+2\gamma(1-\gamma)\frac{1}{3}z_2z_0+\gamma^2\frac{1}{3}z_2z_0z_{02}$ $(1-\gamma)^2\frac{1}{3}z_2z_0z_{01}z_{13}z_{23}+2\gamma(1-\gamma)\frac{1}{3}z_2z_0+\gamma^2\frac{1}{3}z_2z_0z_{02}$
$(1,0,1,2)$	$(2,1,0,1)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{13}\right)+\gamma^2\left(1-\frac{2}{3}z_0z_{02}z_{23}\right)$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}z_{23}$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}z_{23}$
$(1,1,0,2)$	$(2,0,1,1)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{13}z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{23}\right)+\gamma^2\left(1-\frac{2}{3}z_0z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_0z_{13}z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}\right)+\gamma^2\frac{1}{3}z_0z_{02}$ $(1-\gamma)^2\frac{1}{3}z_0z_{13}z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}\right)+\gamma^2\frac{1}{3}z_0z_{02}$
$(1,2,1,0)$	$(0,1,2,1)$	$1-\frac{2}{3}z_{23}z_2$ $\frac{1}{3}z_{23}z_2$ $\frac{1}{3}z_{23}z_2$
$(1,2,0,1)$	$(1,0,2,1)$	$(1-\gamma)\left(1-\frac{2}{3}z_{23}z_2\right)+\gamma\left(1-\frac{2}{3}z_2\right)$ $(1-\gamma)\frac{1}{3}z_{23}z_2+\gamma\frac{1}{3}z_2$ $(1-\gamma)\frac{1}{3}z_{23}z_2+\gamma\frac{1}{3}z_2$
$(1,1,1,1)$		$(1-\gamma)(1-\frac{2}{3}z_{02})+\gamma\frac{1}{3}z_{01}$ $(1-\gamma)\frac{1}{3}z_{02}+\gamma(1-\frac{2}{3}z_{01})$ $(1-\gamma)\frac{1}{3}z_{02}+\gamma\frac{1}{3}z_{01}$

Table 8: CF equations for 4-cycle network $\mathcal{N}_{up}$ for $N = 2121$ (Diamond 8)

C Phylogenetic networks as algebraic varieties: $n = 7$ taxa

C.1 CF equations for $\mathcal{N}_{down}$ for $N = 1222$ (Diamond 2)

n	CF Formula
$(0,0,2,2)$	$1-\frac{2}{3}z_2z_{23}z_3$ $\frac{1}{3}z_2z_{23}z_3$ $\frac{1}{3}z_2z_{23}z_3$
$(0,1,2,1)$	$1-\frac{2}{3}z_{23}z_2$ $\frac{1}{3}z_{23}z_2$ $\frac{1}{3}z_{23}z_2$
$(0,1,1,2)$	$1-\frac{2}{3}z_3$ $\frac{1}{3}z_3$ $\frac{1}{3}z_3$
$(0,2,2,0)$	$1-\frac{2}{3}z_2z_{23}z_{13}z_1$ $\frac{1}{3}z_2z_{23}z_{13}z_1$ $\frac{1}{3}z_2z_{23}z_{13}z_1$
$(0,2,1,1)$	$1-\frac{2}{3}z_{13}z_1$ $\frac{1}{3}z_{13}z_1$ $\frac{1}{3}z_{13}z_1$
$(0,2,0,2)$	$1-\frac{2}{3}z_3z_{13}z_1$ $\frac{1}{3}z_3z_{13}z_1$ $\frac{1}{3}z_3z_{13}z_1$
$(1,0,2,1)$	$(1-\gamma)\left(1-\frac{2}{3}z_{23}z_2\right)+\gamma\left(1-\frac{2}{3}z_2\right)$ $(1-\gamma)\frac{1}{3}z_{23}z_2+\gamma\frac{1}{3}z_2$ $(1-\gamma)\frac{1}{3}z_{23}z_2+\gamma\frac{1}{3}z_2$
$(1,0,1,2)$	$(1-\gamma)\left(1-\frac{2}{3}z_3\right)+\gamma\left(1-\frac{2}{3}z_{23}z_3\right)$ $(1-\gamma)\frac{1}{3}z_3+\gamma\frac{1}{3}z_{23}z_3$ $(1-\gamma)\frac{1}{3}z_3+\gamma\frac{1}{3}z_{23}z_3$
$(1,1,2,0)$	$(1-\gamma)\left(1-\frac{2}{3}z_{13}z_{23}z_2\right)+\gamma\left(1-\frac{2}{3}z_2\right)$ $(1-\gamma)\frac{1}{3}z_{13}z_{23}z_2+\gamma\frac{1}{3}z_2$ $(1-\gamma)\frac{1}{3}z_{13}z_{23}z_2+\gamma\frac{1}{3}z_2$

$(1,1,0,2)$	$(1-\gamma)\left(1-\frac{2}{3}z_{13}z_3\right)+\gamma\left(1-\frac{2}{3}z_3\right)$ $(1-\gamma)\frac{1}{3}z_{13}z_3+\gamma\frac{1}{3}z_3$ $(1-\gamma)\frac{1}{3}z_{13}z_3+\gamma\frac{1}{3}z_3$
$(1,2,1,0)$	$(1-\gamma)\left(1-\frac{2}{3}z_1\right)+\gamma\left(1-\frac{2}{3}z_{23}z_{13}z_1\right)$ $(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{23}z_{13}z_1$ $(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{23}z_{13}z_1$
$(1,2,0,1)$	$(1-\gamma)\left(1-\frac{2}{3}z_1\right)+\gamma\left(1-\frac{2}{3}z_{13}z_1\right)$ $(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{13}z_1$ $(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{13}z_1$
$(1,1,1,1)$	$(1-\gamma)\left(1-\frac{2}{3}z_{13}\right)+\gamma\frac{1}{3}z_{23}$ $(1-\gamma)\frac{1}{3}z_{13}+\gamma\left(1-\frac{2}{3}z_{23}\right)$ $(1-\gamma)\frac{1}{3}z_{13}+\gamma\frac{1}{3}z_{23}$

Table 9: CF equations for $\mathcal{N}_{down}$ for $N = 1222$ (Diamond 2)

C.2 CF equations for $\mathcal{N}_{right}$ for $N = 2122$ (Diamond 5)

$n \in \mathcal{N}_{right}$	$n \in \mathcal{N}_{down}$	CF Formula
$(0,0,2,2)$	$(2,0,2,0)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_2z_0z_{01}z_{13}z_{23}\right)+2\gamma(1-\gamma)\left(1-\frac{2}{3}z_2z_0\right)+\gamma^2\left(1-\frac{2}{3}z_2z_0z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_2z_0z_{01}z_{13}z_{23}+2\gamma(1-\gamma)\frac{1}{3}z_2z_0+\gamma^2\frac{1}{3}z_2z_0z_{02}$ $(1-\gamma)^2\frac{1}{3}z_2z_0z_{01}z_{13}z_{23}+2\gamma(1-\gamma)\frac{1}{3}z_2z_0+\gamma^2\frac{1}{3}z_2z_0z_{02}$
$(0,1,2,1)$	$(2,0,1,1)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{13}z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{23}\right)+\gamma^2\left(1-\frac{2}{3}z_0z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_0z_{13}z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}\right)+\gamma^2\frac{1}{3}z_0z_{02}$ $(1-\gamma)^2\frac{1}{3}z_0z_{13}z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}\right)+\gamma^2\frac{1}{3}z_0z_{02}$
$(0,1,1,2)$	$(1,0,2,1)$	$(1-\gamma)\left(1-\frac{2}{3}z_{23}z_2\right)+\gamma\left(1-\frac{2}{3}z_2\right)$ $(1-\gamma)\frac{1}{3}z_{23}z_2+\gamma\frac{1}{3}z_2$ $(1-\gamma)\frac{1}{3}z_{23}z_2+\gamma\frac{1}{3}z_2$
$(0,2,2,0)$	$(2,0,0,2)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_3z_0z_{13}z_{01}\right)+2\gamma(1-\gamma)\left(1-\frac{2}{3}z_3z_0\right)+\gamma^2\left(1-\frac{2}{3}z_3z_0z_{23}z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_3z_0z_{13}z_{01}+2\gamma(1-\gamma)\frac{1}{3}z_3z_0+\gamma^2\frac{1}{3}z_3z_0z_{23}z_{02}$ $(1-\gamma)^2\frac{1}{3}z_3z_0z_{13}z_{01}+2\gamma(1-\gamma)\frac{1}{3}z_3z_0+\gamma^2\frac{1}{3}z_3z_0z_{23}z_{02}$

$(0,2,1,1)$	$(1,0,1,2)$	$(1-\gamma)\left(1-\frac{2}{3}z_3\right)+\gamma\left(1-\frac{2}{3}z_{23}z_3\right)$ $(1-\gamma)\frac{1}{3}z_3+\gamma\frac{1}{3}z_{23}z_3$ $(1-\gamma)\frac{1}{3}z_3+\gamma\frac{1}{3}z_{23}z_3$
$(0,2,0,2)$	$(0,0,2,2)$	$1-\frac{2}{3}z_2z_{23}z_3$ $\frac{1}{3}z_2z_{23}z_3$ $\frac{1}{3}z_2z_{23}z_3$
$(1,0,2,1)$	$(2,1,1,0)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{23}z_{13}\right)+\gamma^2\left(1-\frac{2}{3}z_0z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}$
$(1,0,1,2)$	$(1,1,2,0)$	$(1-\gamma)\left(1-\frac{2}{3}z_{13}z_{23}z_2\right)+\gamma\left(1-\frac{2}{3}z_2\right)$ $(1-\gamma)\frac{1}{3}z_{13}z_{23}z_2+\gamma\frac{1}{3}z_2$ $(1-\gamma)\frac{1}{3}z_{13}z_{23}z_2+\gamma\frac{1}{3}z_2$
$(1,1,2,0)$	$(2,1,0,1)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{13}\right)+\gamma^2\left(1-\frac{2}{3}z_0z_{02}z_{23}\right)$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}z_{23}$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}z_{23}$
$(1,1,0,2)$	$(0,1,2,1)$	$1-\frac{2}{3}z_{23}z_2$ $\frac{1}{3}z_{23}z_2$ $\frac{1}{3}z_{23}z_2$
$(1,2,1,0)$	$(1,1,0,2)$	$(1-\gamma)\left(1-\frac{2}{3}z_{13}z_3\right)+\gamma\left(1-\frac{2}{3}z_3\right)$ $(1-\gamma)\frac{1}{3}z_{13}z_3+\gamma\frac{1}{3}z_3$ $(1-\gamma)\frac{1}{3}z_{13}z_3+\gamma\frac{1}{3}z_3$
$(1,2,0,1)$	$(0,1,1,2)$	$1-\frac{2}{3}z_3$ $\frac{1}{3}z_3$ $\frac{1}{3}z_3$
$(1,1,1,1)$		$(1-\gamma)\frac{1}{3}z_{01}+\gamma(1-\frac{2}{3}z_{13})$ $(1-\gamma)(1-\frac{2}{3}z_{01})+\gamma\frac{1}{3}z_{13}$ $(1-\gamma)\frac{1}{3}z_{01}+\gamma\frac{1}{3}z_{13}$

Table 10: CF equations for $\mathcal{N}_{right}$ for $N = 2122$ (Diamond 5)

C.3 CF equations for $\mathcal{N}_{left}$ for $N = 2212$ (Diamond 3)

$n \in \mathcal{N}_{left}$	$n \in \mathcal{N}_{down}$	CF Formula
$(0,0,2,2)$	$(0,2,0,2)$	$1-\frac{2}{3}z_3z_{13}z_1$ $\frac{1}{3}z_3z_{13}z_1$ $\frac{1}{3}z_3z_{13}z_1$
$(0,1,2,1)$	$(1,1,0,2)$	$(1-\gamma)\left(1-\frac{2}{3}z_{13}z_3\right)+\gamma\left(1-\frac{2}{3}z_3\right)$ $(1-\gamma)\frac{1}{3}z_{13}z_3+\gamma\frac{1}{3}z_3$ $(1-\gamma)\frac{1}{3}z_{13}z_3+\gamma\frac{1}{3}z_3$
$(0,1,1,2)$	$(1,2,0,1)$	$(1-\gamma)\left(1-\frac{2}{3}z_1\right)+\gamma\left(1-\frac{2}{3}z_{13}z_1\right)$ $(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{13}z_1$ $(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{13}z_1$
$(0,2,2,0)$	$(2,0,0,2)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_3z_0z_{13}z_{01}\right)+2\gamma(1-\gamma)\left(1-\frac{2}{3}z_3z_0\right)+\gamma^2\left(1-\frac{2}{3}z_3z_0z_{23}z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_3z_0z_{13}z_{01}+2\gamma(1-\gamma)\frac{1}{3}z_3z_0+\gamma^2\frac{1}{3}z_3z_0z_{23}z_{02}$ $(1-\gamma)^2\frac{1}{3}z_3z_0z_{13}z_{01}+2\gamma(1-\gamma)\frac{1}{3}z_3z_0+\gamma^2\frac{1}{3}z_3z_0z_{23}z_{02}$
$(0,2,1,1)$	$(2,1,0,1)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{13}\right)+\gamma^2\left(1-\frac{2}{3}z_0z_{02}z_{23}\right)$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}z_{23}$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}z_{23}$
$(0,2,0,2)$	$(2,2,0,0)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_1z_0z_{01}\right)+2\gamma(1-\gamma)\left(1-\frac{2}{3}z_1z_0\right)+\gamma^2\left(1-\frac{2}{3}z_1z_0z_{02}z_{23}z_{13}\right)$ $(1-\gamma)^2\frac{1}{3}z_1z_0z_{01}+2\gamma(1-\gamma)\frac{1}{3}z_1z_0+\gamma^2\frac{1}{3}z_1z_0z_{02}z_{23}z_{13}$ $(1-\gamma)^2\frac{1}{3}z_1z_0z_{01}+2\gamma(1-\gamma)\frac{1}{3}z_1z_0+\gamma^2\frac{1}{3}z_1z_0z_{02}z_{23}z_{13}$
$(1,0,2,1)$	$(0,1,1,2)$	$1-\frac{2}{3}z_3$ $\frac{1}{3}z_3$ $\frac{1}{3}z_3$
$(1,0,1,2)$	$(0,2,1,1)$	$1-\frac{2}{3}z_{13}z_1$ $\frac{1}{3}z_{13}z_1$ $\frac{1}{3}z_{13}z_1$
$(1,1,2,0)$	$(1,0,1,2)$	$(1-\gamma)\left(1-\frac{2}{3}z_3\right)+\gamma\left(1-\frac{2}{3}z_{23}z_3\right)$ $(1-\gamma)\frac{1}{3}z_3+\gamma\frac{1}{3}z_{23}z_3$ $(1-\gamma)\frac{1}{3}z_3+\gamma\frac{1}{3}z_{23}z_3$
$(1,1,0,2)$	$(1,2,1,0)$	$(1-\gamma)\left(1-\frac{2}{3}z_1\right)+\gamma\left(1-\frac{2}{3}z_{23}z_{13}z_1\right)$ $(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{23}z_{13}z_1$

		$(1-\gamma)\frac{1}{3}z_1 + \gamma\frac{1}{3}z_{23}z_{13}z_1$
$(1,2,1,0)$	$(2,0,1,1)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{13}z_{01}\right) + 2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{23}\right) + \gamma^2\left(1-\frac{2}{3}z_0z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_0z_{13}z_{01} + \gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}\right) + \gamma^2\frac{1}{3}z_0z_{02}$ $(1-\gamma)^2\frac{1}{3}z_0z_{13}z_{01} + \gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}\right) + \gamma^2\frac{1}{3}z_0z_{02}$
$(1,2,0,1)$	$(2,1,1,0)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{01}\right) + 2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{23}z_{13}\right) + \gamma^2\left(1-\frac{2}{3}z_0z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_0z_{01} + \gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}z_{13}\right) + \gamma^2\frac{1}{3}z_0z_{02}$ $(1-\gamma)^2\frac{1}{3}z_0z_{01} + \gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}z_{13}\right) + \gamma^2\frac{1}{3}z_0z_{02}$
$(1,1,1,1)$		$(1-\gamma)\frac{1}{3}z_{23} + \gamma(1-\frac{2}{3}z_{02})$ $(1-\gamma)(1-\frac{2}{3}z_{23}) + \gamma\frac{1}{3}z_{02}$ $(1-\gamma)\frac{1}{3}z_{23} + \gamma\frac{1}{3}z_{02}$

Table 11: CF equations for $\mathcal{N}_{left}$ for $N = 2212$ (Diamond 3)

C.4 CF equations for $\mathcal{N}_{up}$ for $N = 2221$ (Diamond 4)

$n \in \mathcal{N}_{up}$	$n \in \mathcal{N}_{down}$	CF Formula
$(0,0,2,2)$	$(2,2,0,0)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_1z_0z_{01}\right) + 2\gamma(1-\gamma)\left(1-\frac{2}{3}z_1z_0\right) + \gamma^2\left(1-\frac{2}{3}z_1z_0z_{02}z_{23}z_{13}\right)$ $(1-\gamma)^2\frac{1}{3}z_1z_0z_{01} + 2\gamma(1-\gamma)\frac{1}{3}z_1z_0 + \gamma^2\frac{1}{3}z_1z_0z_{02}z_{23}z_{13}$ $(1-\gamma)^2\frac{1}{3}z_1z_0z_{01} + 2\gamma(1-\gamma)\frac{1}{3}z_1z_0 + \gamma^2\frac{1}{3}z_1z_0z_{02}z_{23}z_{13}$
$(0,1,2,1)$	$(1,2,1,0)$	$(1-\gamma)\left(1-\frac{2}{3}z_1\right) + \gamma\left(1-\frac{2}{3}z_{23}z_{13}z_1\right)$ $(1-\gamma)\frac{1}{3}z_1 + \gamma\frac{1}{3}z_{23}z_{13}z_1$ $(1-\gamma)\frac{1}{3}z_1 + \gamma\frac{1}{3}z_{23}z_{13}z_1$
$(0,1,1,2)$	$(2,1,1,0)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{01}\right) + 2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{23}z_{13}\right) + \gamma^2\left(1-\frac{2}{3}z_0z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_0z_{01} + \gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}z_{13}\right) + \gamma^2\frac{1}{3}z_0z_{02}$ $(1-\gamma)^2\frac{1}{3}z_0z_{01} + \gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}z_{13}\right) + \gamma^2\frac{1}{3}z_0z_{02}$
$(0,2,2,0)$	$(0,2,2,0)$	$1-\frac{2}{3}z_2z_{23}z_{13}z_1$ $\frac{1}{3}z_2z_{23}z_{13}z_1$ $\frac{1}{3}z_2z_{23}z_{13}z_1$
$(0,2,1,1)$	$(1,1,2,0)$	$(1-\gamma)\left(1-\frac{2}{3}z_{13}z_{23}z_2\right) + \gamma\left(1-\frac{2}{3}z_2\right)$ $(1-\gamma)\frac{1}{3}z_{13}z_{23}z_2 + \gamma\frac{1}{3}z_2$ $(1-\gamma)\frac{1}{3}z_{13}z_{23}z_2 + \gamma\frac{1}{3}z_2$

$(0,2,0,2)$	$(2,0,2,0)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_2z_0z_{01}z_{13}z_{23}\right)+2\gamma(1-\gamma)\left(1-\frac{2}{3}z_2z_0\right)+\gamma^2\left(1-\frac{2}{3}z_2z_0z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_2z_0z_{01}z_{13}z_{23}+2\gamma(1-\gamma)\frac{1}{3}z_2z_0+\gamma^2\frac{1}{3}z_2z_0z_{02}$ $(1-\gamma)^2\frac{1}{3}z_2z_0z_{01}z_{13}z_{23}+2\gamma(1-\gamma)\frac{1}{3}z_2z_0+\gamma^2\frac{1}{3}z_2z_0z_{02}$
$(1,0,2,1)$	$(1,2,0,1)$	$(1-\gamma)\left(1-\frac{2}{3}z_1\right)+\gamma\left(1-\frac{2}{3}z_{13}z_1\right)$ $(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{13}z_1$ $(1-\gamma)\frac{1}{3}z_1+\gamma\frac{1}{3}z_{13}z_1$
$(1,0,1,2)$	$(2,1,0,1)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{13}\right)+\gamma^2\left(1-\frac{2}{3}z_0z_{02}z_{23}\right)$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}z_{23}$ $(1-\gamma)^2\frac{1}{3}z_0z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{13}\right)+\gamma^2\frac{1}{3}z_0z_{02}z_{23}$
$(1,1,2,0)$	$(0,2,1,1)$	$1-\frac{2}{3}z_{13}z_1$ $\frac{1}{3}z_{13}z_1$ $\frac{1}{3}z_{13}z_1$
$(1,1,0,2)$	$(2,0,1,1)$	$(1-\gamma)^2\left(1-\frac{2}{3}z_0z_{13}z_{01}\right)+2\gamma(1-\gamma)\left(1-z_0+\frac{1}{3}z_0z_{23}\right)+\gamma^2\left(1-\frac{2}{3}z_0z_{02}\right)$ $(1-\gamma)^2\frac{1}{3}z_0z_{13}z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}\right)+\gamma^2\frac{1}{3}z_0z_{02}$ $(1-\gamma)^2\frac{1}{3}z_0z_{13}z_{01}+\gamma(1-\gamma)z_0\left(1-\frac{1}{3}z_{23}\right)+\gamma^2\frac{1}{3}z_0z_{02}$
$(1,2,1,0)$	$(0,1,2,1)$	$1-\frac{2}{3}z_{23}z_2$ $\frac{1}{3}z_{23}z_2$ $\frac{1}{3}z_{23}z_2$
$(1,2,0,1)$	$(1,0,2,1)$	$(1-\gamma)\left(1-\frac{2}{3}z_{23}z_2\right)+\gamma\left(1-\frac{2}{3}z_2\right)$ $(1-\gamma)\frac{1}{3}z_{23}z_2+\gamma\frac{1}{3}z_2$ $(1-\gamma)\frac{1}{3}z_{23}z_2+\gamma\frac{1}{3}z_2$
$(1,1,1,1)$		$(1-\gamma)(1-\frac{2}{3}z_{02})+\gamma\frac{1}{3}z_{01}$ $(1-\gamma)\frac{1}{3}z_{02}+\gamma(1-\frac{2}{3}z_{01})$ $(1-\gamma)\frac{1}{3}z_{02}+\gamma\frac{1}{3}z_{01}$

Table 12: CF equations for $\mathcal{N}_{up}$ for $N = 2221$ (Diamond 4)

D Subset of CF equations used in Macaulay2

We present below the table with the major CF for all 4-taxon subsets, except (1,1,1,1) for which we present all equations. All equations are multiplied by 3 for conciseness. All Macaulay2 scripts can be found in `https://github.com/gtiley/diamond-identifiability/tree/main/scripts/macaulay2`.

(n_0, n_1, n_2, n_3)	$\mathcal{N}_{down}$	$\mathcal{N}_{right}$
$(0,0,2,2)$	$3 - 2z_2z_{23}z_3$	$(1-\gamma_u)^2(3-2u_2u_0u_{01}u_{13}u_{23})$ $+2\gamma_u(1-\gamma_u)(3-2u_2u_0)$ $+\gamma_u^2(3-2u_2u_0u_{02})$
$(0,1,2,1)$	$3 - 2z_{23}z_2$	$(1-\gamma_u)^2(3-2u_0u_{13}u_{01})$ $+2\gamma_u(1-\gamma_u)(3-3u_0+u_0u_{23})$ $+\gamma_u^2(3-2u_0u_{02})$
$(0,1,1,2)$	$3 - 2z_3$	$(1-\gamma_u)(3-2u_{23}u_2)+\gamma_u(3-2u_2)$
$(0,2,2,0)$	$3 - 2z_2z_{23}z_{13}z_1$	$(1-\gamma_u)^2(3-2u_3u_0u_{13}u_{01})$ $+2\gamma_u(1-\gamma_u)(3-2u_3u_0)$ $+\gamma_u^2(3-2u_3u_0u_{23}u_{02})$
$(0,2,1,1)$	$3 - 2z_{13}z_1$	$(1-\gamma_u)(3-2z_3)+\gamma_u(3-2z_{23}z_3)$
$(0,2,0,2)$	$3 - 2z_3z_{13}z_1$	$3 - 2z_2z_{23}z_3$
$(1,0,2,1)$	$(1-\gamma)(3-2z_{23}z_2)+\gamma(3-2z_2)$	$(1-\gamma_u)^2(3-2u_0u_{01})$ $+2\gamma_u(1-\gamma_u)(3-3u_0+u_0u_{23}u_{13})$ $+\gamma_u^2(3-2u_0u_{02})$
$(1,0,1,2)$	$(1-\gamma)(3-2z_3)+\gamma(3-2z_{23}z_3)$	$(1-\gamma_u)(3-2u_{13}u_{23}u_2)+\gamma_u(3-2u_2)$
$(1,1,2,0)$	$(1-\gamma)(3-2z_{13}z_{23}z_2)+\gamma(3-2z_2)$	$(1-\gamma_u)^2(3-2u_0u_{01})$ $+2\gamma_u(1-\gamma_u)(3-3u_0+u_0u_{13})$ $+\gamma_u^2(3-2u_0u_{02}u_{23})$
$(1,1,0,2)$	$(1-\gamma)(3-2z_{13}z_3)+\gamma(3-2z_3)$	$3 - 2u_{23}u_2$
$(1,2,1,0)$	$(1-\gamma)(3-2z_1)+\gamma(3-2z_{23}z_{13}z_1)$	$(1-\gamma_u)(3-2u_{13}u_3)+\gamma_u(3-2u_3)$
$(1,2,0,1)$	$(1-\gamma)(3-2z_1)+\gamma(3-2z_{13}z_1)$	$3 - 2u_3$
$(2,0,2,0)$	$(1-\gamma)^2(3-2z_2z_0z_{01}z_{13}z_{23})$ $+2\gamma(1-\gamma)(3-2z_2z_0)$ $+\gamma^2(3-2z_2z_0z_{02})$	$(1-\gamma_u)^2(3-2u_1u_0u_{01})$ $+2\gamma_u(1-\gamma_u)(3-2u_1u_0)$ $+\gamma_u^2(3-2u_1u_0u_{02}u_{23}u_{13})$
$(2,0,1,1)$	$(1-\gamma)^2(3-2z_0z_{13}z_{01})$ $+2\gamma(1-\gamma)(3-3z_0+z_0z_{23})$ $+\gamma^2(3-2z_0z_{02})$	$(1-\gamma_u)(3-2u_1)+\gamma_u(3-2u_{23}u_{13}u_1)$
$(2,0,0,2)$	$(1-\gamma)^2(3-2z_3z_0z_{13}z_{01})$	$3 - 2u_2u_{23}u_{13}u_1$

	$+2\gamma(1-\gamma)\,(3-2z_3z_0)$ $+\gamma^2\,(3-2z_3z_0z_{23}z_{02})$	
$(2,1,1,0)$	$(1-\gamma)^2\,(3-2z_0z_{01})$ $+2\gamma(1-\gamma)\,(3-3z_0+z_0z_{23}z_{13})$ $+\gamma^2\,(3-2z_0z_{02})$	$(1-\gamma_u)\,(3-2u_1)+\gamma_u\,(3-2u_{13}u_1)$
$(2,1,0,1)$	$(1-\gamma)^2\,(3-2z_0z_{01})$ $+2\gamma(1-\gamma)\,(3-3z_0+z_0z_{13})$ $+\gamma^2\,(3-2z_0z_{02}z_{23})$	$3-2u_{13}u_1$
$(2,2,0,0)$	$(1-\gamma)^2\,(3-2z_1z_0z_{01})$ $+2\gamma(1-\gamma)\,(3-2z_1z_0)$ $+\gamma^2\,(3-2z_1z_0z_{02}z_{23}z_{13})$	$3-2u_3u_{13}u_1$
$(1,1,1,1)$	$(1-\gamma)\,(3-2z_{13})+\gamma z_{23}$ $(1-\gamma)z_{13}+\gamma\,(3-2z_{23})$ $(1-\gamma)z_{13}+\gamma z_{23}$	$(1-\gamma_u)u_{01}+\gamma_u(3-2u_{13})$ $(1-\gamma_u)(3-2u_{01})+\gamma_u u_{13}$ $(1-\gamma_u)u_{01}+\gamma_u u_{13}$

Table 13: Subset of CF equations used in Macaulay2 (Part 1)

(n_0,n_1,n_2,n_3)	$\mathcal{N}_{left}$	$\mathcal{N}_{up}$
$(0,0,2,2)$	$3-2v_3v_{13}v_1$	$(1-\gamma_w)^2\,(3-2w_1w_0w_{01})$ $+2\gamma_w(1-\gamma_w)\,(3-2w_1w_0)$ $+\gamma_w^2\,(3-2w_1w_0w_{02}w_{23}w_{13})$
$(0,1,2,1)$	$(1-\gamma_v)\,(3-2v_{13}v_3)+\gamma_v\,(3-2v_3)$	$(1-\gamma_w)\,(3-2w_1)+\gamma_w\,(3-2w_{23}w_{13}w_1)$
$(0,1,1,2)$	$(1-\gamma_v)\,(3-2v_1)+\gamma_v\,(3-2v_{13}v_1)$	$(1-\gamma_w)^2\,(3-2w_0w_{01})$ $+2\gamma_w(1-\gamma_w)\,(3-3w_0+w_0w_{23}w_{13})$ $+\gamma_w^2\,(3-2w_0w_{02})$
$(0,2,2,0)$	$(1-\gamma_v)^2\,(3-2v_3v_0v_{13}v_{01})$ $+2\gamma_v(1-\gamma_v)\,(3-2v_3v_0)$ $+\gamma_v^2\,(3-2v_3v_0v_{23}v_{02})$	$3-2w_2w_{23}w_{13}w_1$
$(0,2,1,1)$	$(1-\gamma_v)^2\,(3-2v_0v_{01})$ $+2\gamma_v(1-\gamma_v)\,(3-3v_0+v_0v_{13})$ $+\gamma_v^2\,(3-2v_0v_{02}v_{23})$	$(1-\gamma_w)\,(3-2w_{13}w_{23}w_2)+\gamma_w\,(3-2w_2)$
$(0,2,0,2)$	$(1-\gamma_v)^2\,(3-2v_1v_0v_{01})$ $+2\gamma_v(1-\gamma_v)\,(3-2v_1v_0)$ $+\gamma_v^2\,(3-2v_1v_0v_{02}v_{23}v_{13})$	$(1-\gamma_w)^2\,(3-2w_2w_0w_{01}w_{13}w_{23})$ $+2\gamma_w(1-\gamma_w)\,(3-2w_2w_0)$ $+\gamma_w^2\,(3-2w_2w_0w_{02})$
$(1,0,2,1)$	$3-2v_3$	$(1-\gamma_w)\,(3-2w_1)+\gamma_w\,(3-2w_{13}w_1)$
$(1,0,1,2)$	$3-2v_{13}v_1$	$(1-\gamma_w)^2\,(3-2w_0w_{01})$

		$+2\gamma_w(1-\gamma_w)\,(3-3w_0+w_0w_{13})$ $+\gamma_w^2\,(3-2w_0w_{02}w_{23})$
$(1,1,2,0)$	$(1-\gamma_v)\,(3-2v_3)+\gamma_v\,(3-2v_{23}v_3)$	$3-2w_{13}w_1$
$(1,1,0,2)$	$(1-\gamma_v)\,(3-2v_1)+\gamma_v\,(3-2v_{23}v_{13}v_1)$	$(1-\gamma_w)^2\,(3-2w_0w_{13}w_{01})$ $+2\gamma_w(1-\gamma_w)\,(3-3w_0+w_0w_{23})$ $+\gamma_w^2\,(3-2w_0w_{02})$
$(1,2,1,0)$	$(1-\gamma_v)^2\,(3-2v_0v_{13}v_{01})$ $+2\gamma_v(1-\gamma_v)\,(3-3v_0+v_0v_{23})$ $+\gamma_v^2\,(3-2v_0v_{02})$	$3-2w_{23}w_2$
$(1,2,0,1)$	$(1-\gamma_v)^2\,(3-2v_0v_{01})$ $+2\gamma_v(1-\gamma_v)\,(3-3v_0+v_0v_{23}v_{13})$ $+\gamma_v^2\,(3-2v_0v_{02})$	$(1-\gamma_w)\,(3-2w_{23}w_2)+\gamma_w\,(3-2w_2)$
$(2,0,2,0)$	$3-2v_2v_{23}v_3$	$3-2w_3w_{13}w_1$
$(2,0,1,1)$	$3-2v_{23}v_2$	$(1-\gamma_w)\,(3-2w_{13}w_3)+\gamma_w\,(3-2w_3)$
$(2,0,0,2)$	$3-2v_2v_{23}v_{13}v_1$	$(1-\gamma_w)^2\,(3-2w_3w_0w_{13}w_{01})$ $+2\gamma_w(1-\gamma_w)\,(3-2w_3w_0)$ $+\gamma_w^2\,(3-2w_3w_0w_{23}w_{02})$
$(2,1,1,0)$	$(1-\gamma_v)\,(3-2v_{23}v_2)+\gamma_v\,(3-2v_2)$	$3-2w_3$
$(2,1,0,1)$	$(1-\gamma_v)\,(3-2v_{13}v_{23}v_2)+\gamma_v\,(3-2v_2)$	$(1-\gamma_w)\,(3-2w_3)+\gamma_w\,(3-2w_{23}w_3)$
$(2,2,0,0)$	$(1-\gamma_v)^2\,(3-2v_2v_0v_{01}v_{13}v_{23})$ $+2\gamma_v(1-\gamma_v)\,(3-2v_2v_0)$ $+\gamma_v^2\,(3-2v_2v_0v_{02})$	$3-2w_2w_{23}w_3$
$(1,1,1,1)$	$(1-\gamma_v)v_{23}+\gamma_v(3-2v_{02})$ $(1-\gamma_v)(3-2v_{23})+\gamma_v v_{02}$ $(1-\gamma_v)v_{23}+\gamma_v v_{02}$	$(1-\gamma_w)(3-2w_{02})+\gamma_w w_{01}$ $(1-\gamma_w)w_{02}+\gamma_w(3-2w_{01})$ $(1-\gamma_w)w_{02}+\gamma_w w_{01}$

Table 14: Subset of CF equations used in Macaulay2 (Part 2)

E Simulation figures and tables

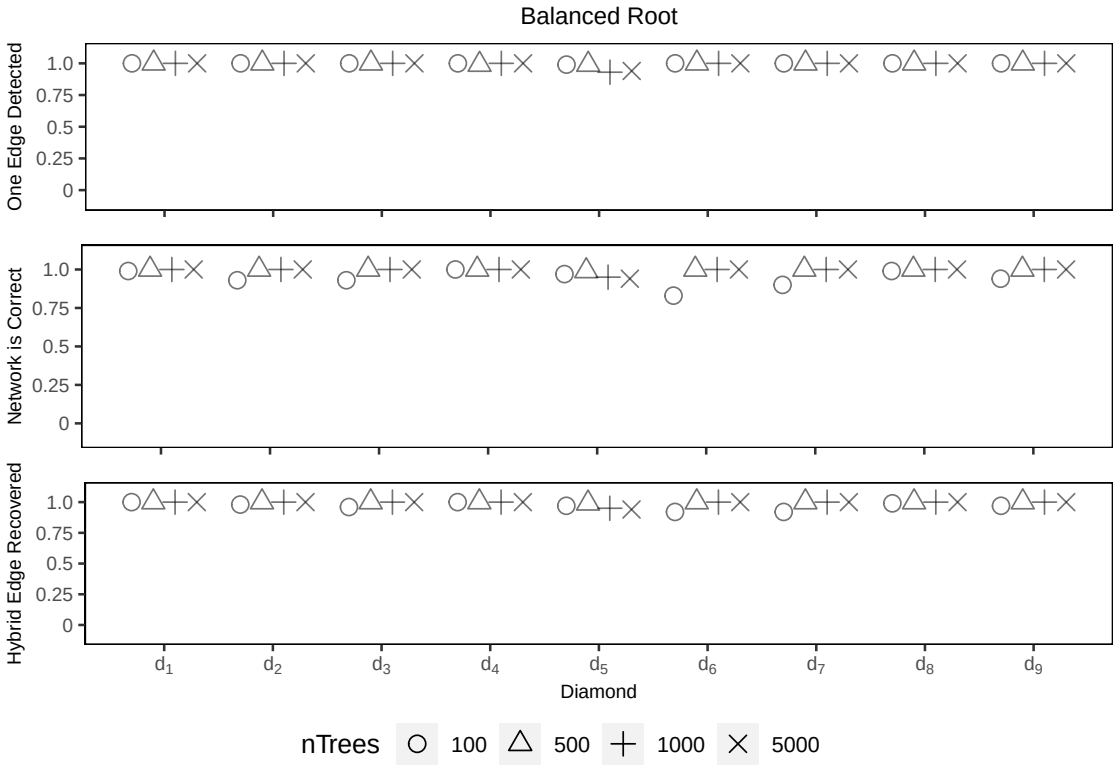

Figure 9: SNaQ results for root 1. *One Edge Detected* is the proportion out of 100 replicates that 1 reticulate edge was correctly inferred by the pseudolikehood scores. *Network is Correct* is the proportion of 100 replicates where the estimated topology when allowing only one reticulation event is identical to the true topology. *Hybrid Edge Recovered* is the proportion of 100 replicates where the correct reticulation edge was inferred, regardless if other parts of the estimated network were incorrect, when allowing only one reticulation event. Diamond d1 through d9 correspond to Fig. 1e.

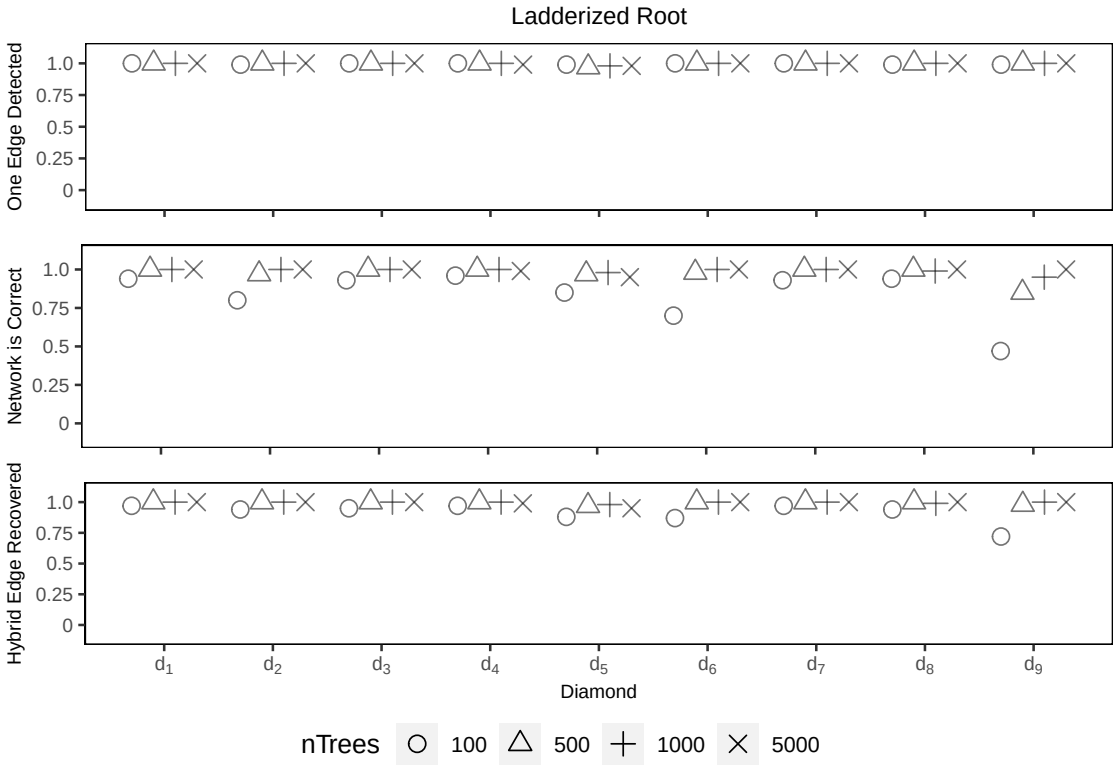

Figure 10: SNaQ results for root 2. *One Edge Detected* is the proportion out of 100 replicates that 1 reticulate edge was correctly inferred by the pseudolikehood scores. *Network is Correct* is the proportion of 100 replicates where the estimated topology when allowing only one reticulation event is identical to the true topology. *Hybrid Edge Recovered* is the proportion of 100 replicates where the correct reticulation edge was inferred, regardless if other parts of the estimated network were incorrect, when allowing only one reticulation event. Diamond d1 through d9 correspond to Fig. 1e.

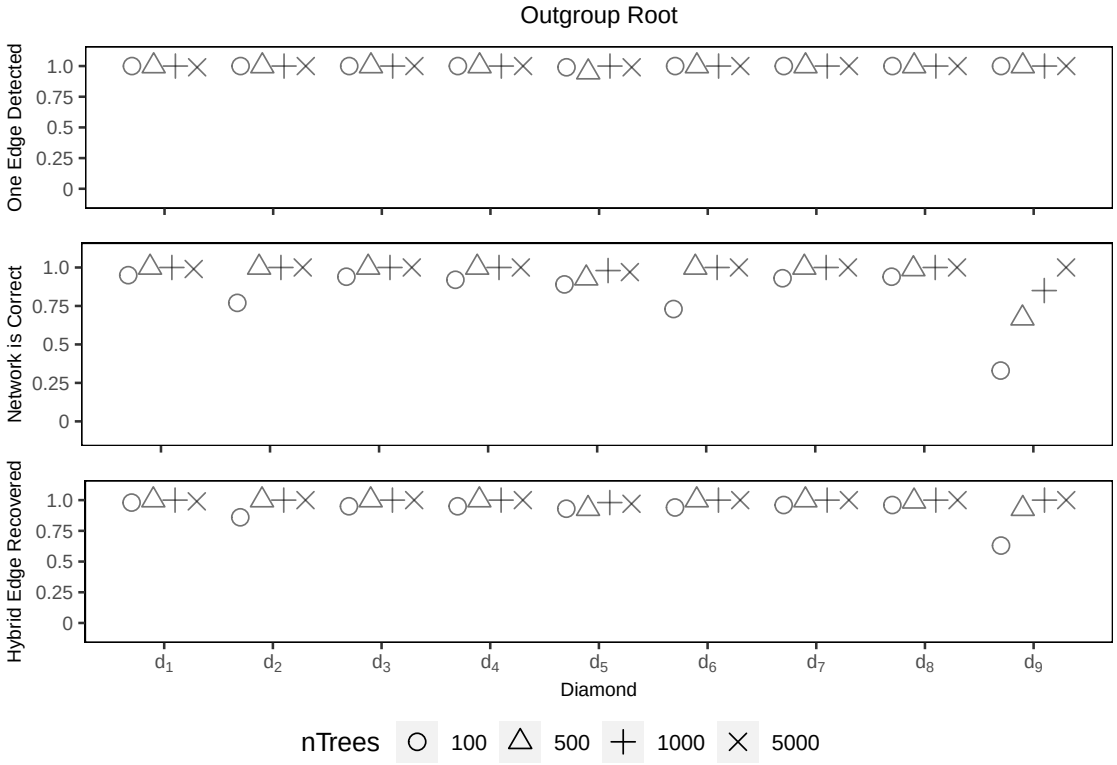

Figure 11: SNaQ results for root 3. *One Edge Detected* is the proportion out of 100 replicates that 1 reticulate edge was correctly inferred by the pseudolikehood scores. *Network is Correct* is the proportion of 100 replicates where the estimated topology when allowing only one reticulation event is identical to the true topology. *Hybrid Edge Recovered* is the proportion of 100 replicates where the correct reticulation edge was inferred, regardless if other parts of the estimated network were incorrect, when allowing only one reticulation event. Diamond d1 through d9 correspond to Fig. 1e.

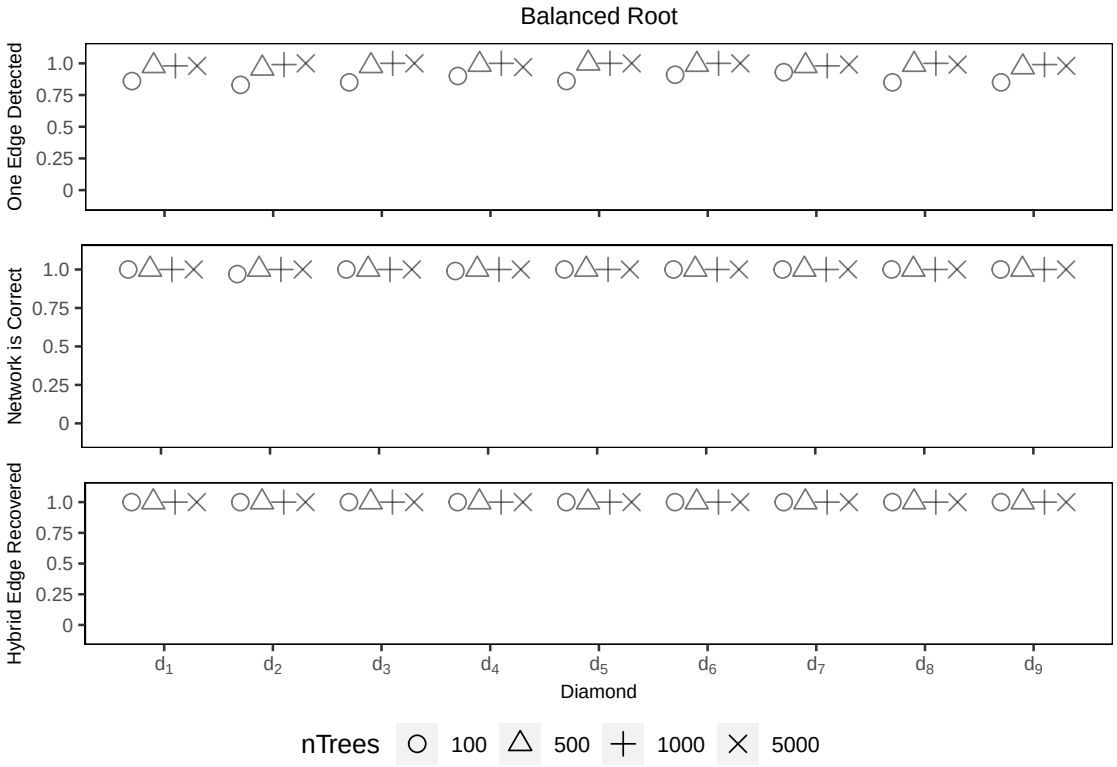

Figure 12: PhyloNet results for root 1. *One Edge Detected* is the proportion out of 100 replicates that 1 reticulate edge was correctly inferred by the pseudolikehood scores. *Network is Correct* is the proportion of 100 replicates where the estimated topology when allowing only one reticulation event is identical to the true topology. *Hybrid Edge Recovered* is the proportion of 100 replicates where the correct reticulation edge was inferred, regardless if other parts of the estimated network were incorrect, when allowing only one reticulation event. Diamond d1 through d9 correspond to Fig. 1e.

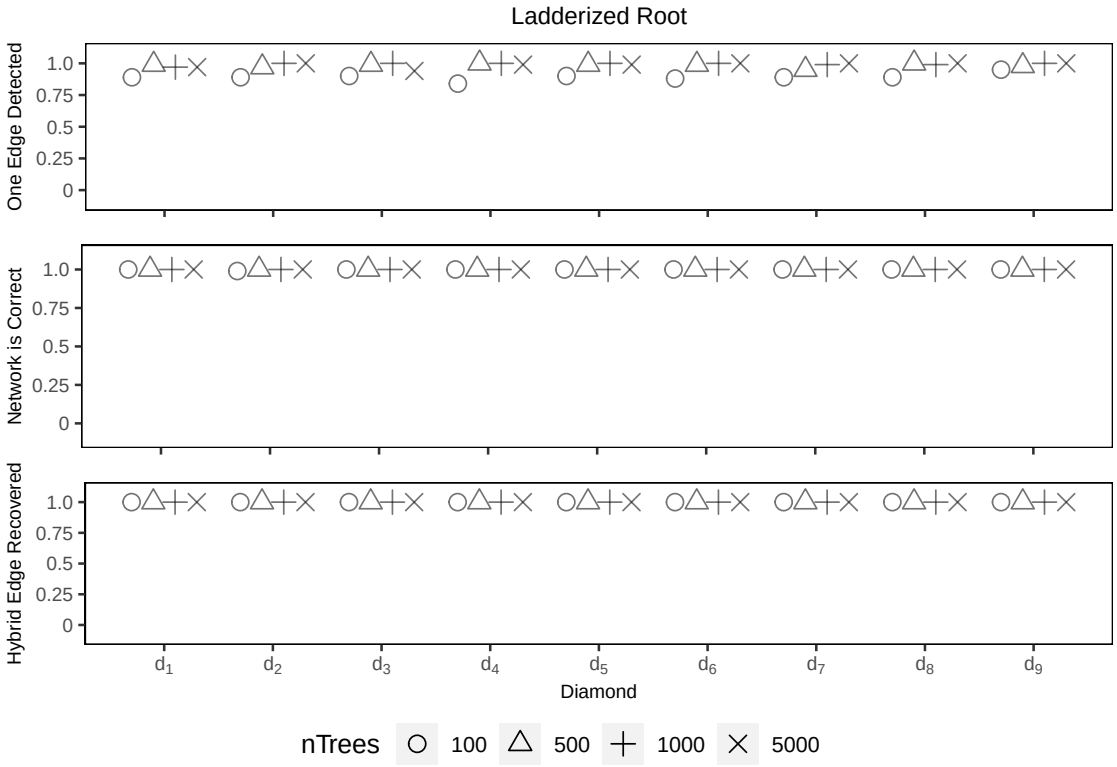

Figure 13: PhyloNet results for root 2. *One Edge Detected* is the proportion out of 100 replicates that 1 reticulate edge was correctly inferred by the pseudolikehood scores. *Network is Correct* is the proportion of 100 replicates where the estimated topology when allowing only one reticulation event is identical to the true topology. *Hybrid Edge Recovered* is the proportion of 100 replicates where the correct reticulation edge was inferred, regardless if other parts of the estimated network were incorrect, when allowing only one reticulation event. Diamond d1 through d9 correspond to Fig. 1e.

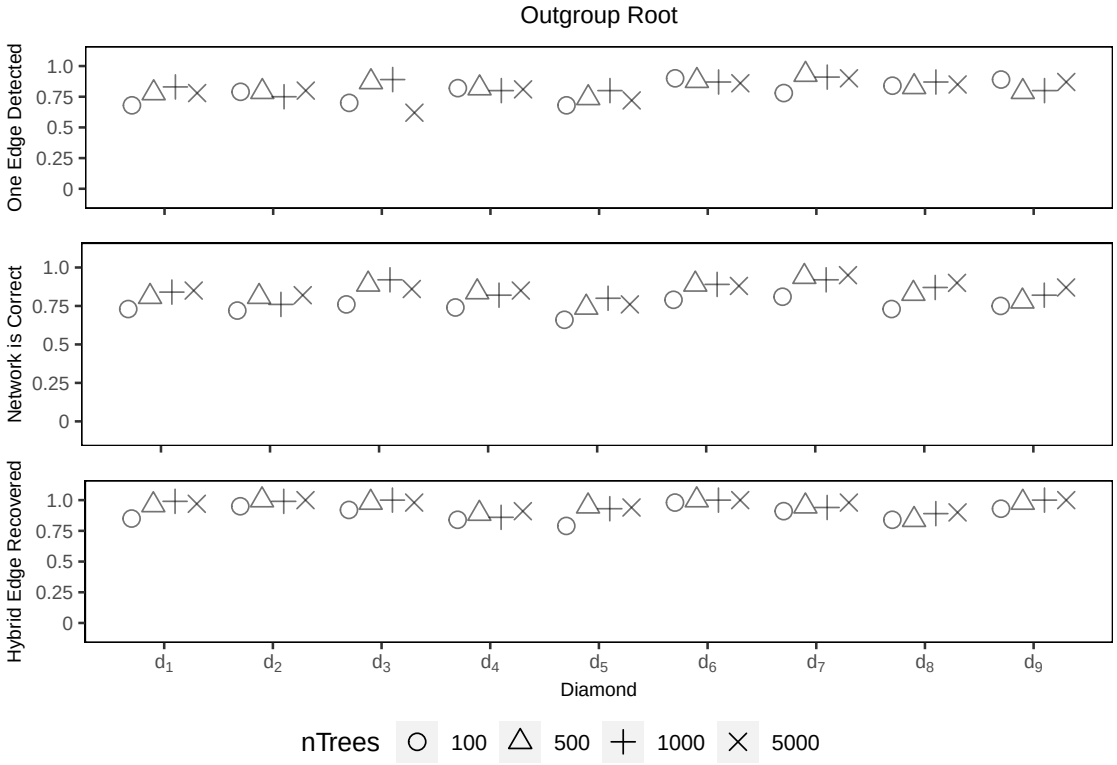

Figure 14: PhyloNet results for root 3. *One Edge Detected* is the proportion out of 100 replicates that 1 reticulate edge was correctly inferred by the pseudolikehood scores. *Network is Correct* is the proportion of 100 replicates where the estimated topology when allowing only one reticulation event is identical to the true topology. *Hybrid Edge Recovered* is the proportion of 100 replicates where the correct reticulation edge was inferred, regardless if other parts of the estimated network were incorrect, when allowing only one reticulation event. Diamond d1 through d9 correspond to Fig. 1e.

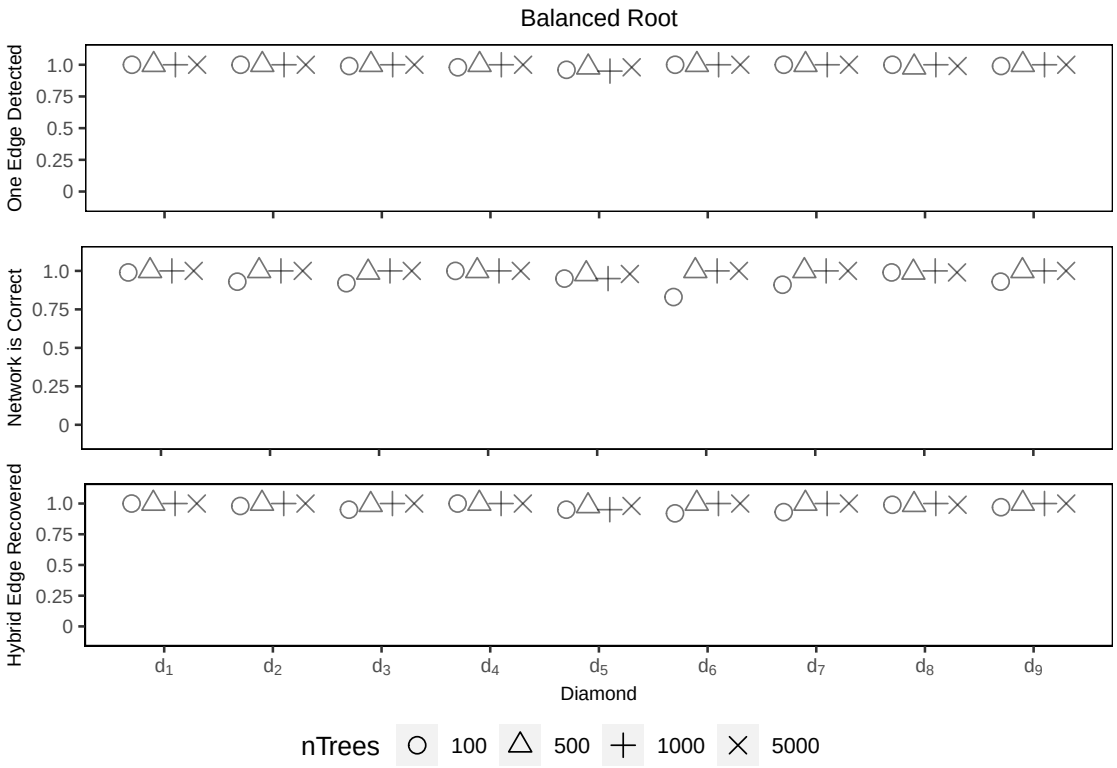

Figure 15: SNaQ results for root 1 with random outgroups for 10% of gene trees.

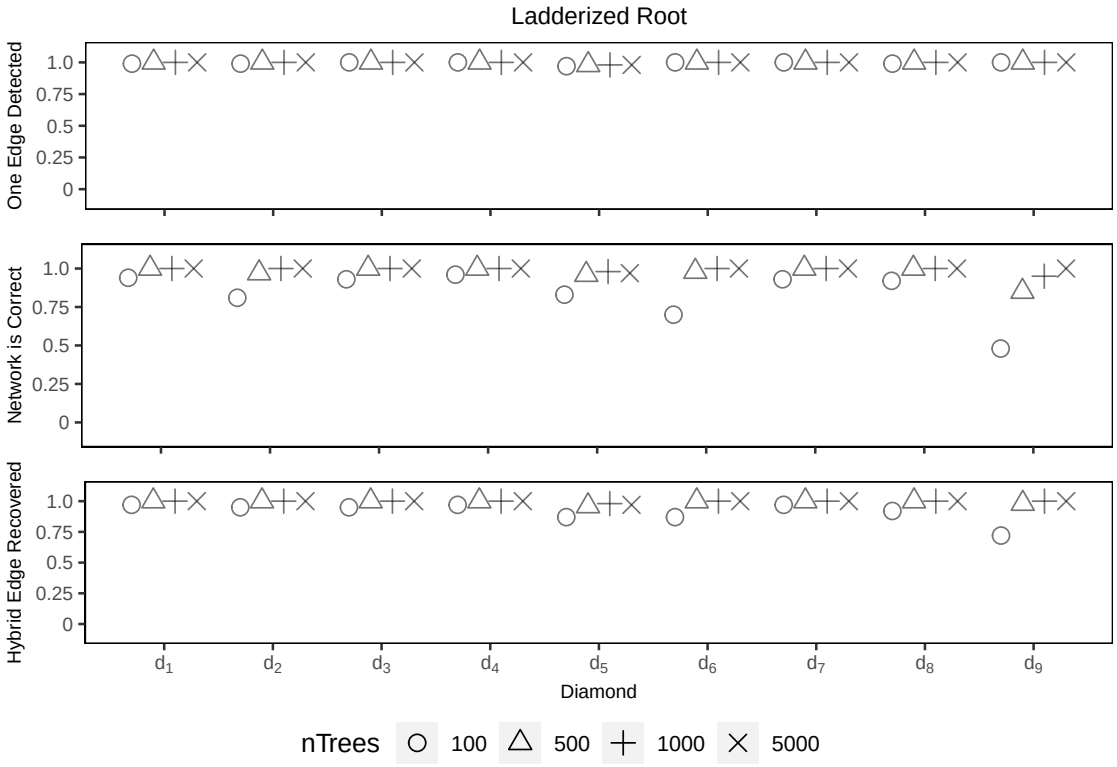

Figure 16: SNaQ results for root 2 with random outgroups for 10% of gene trees.

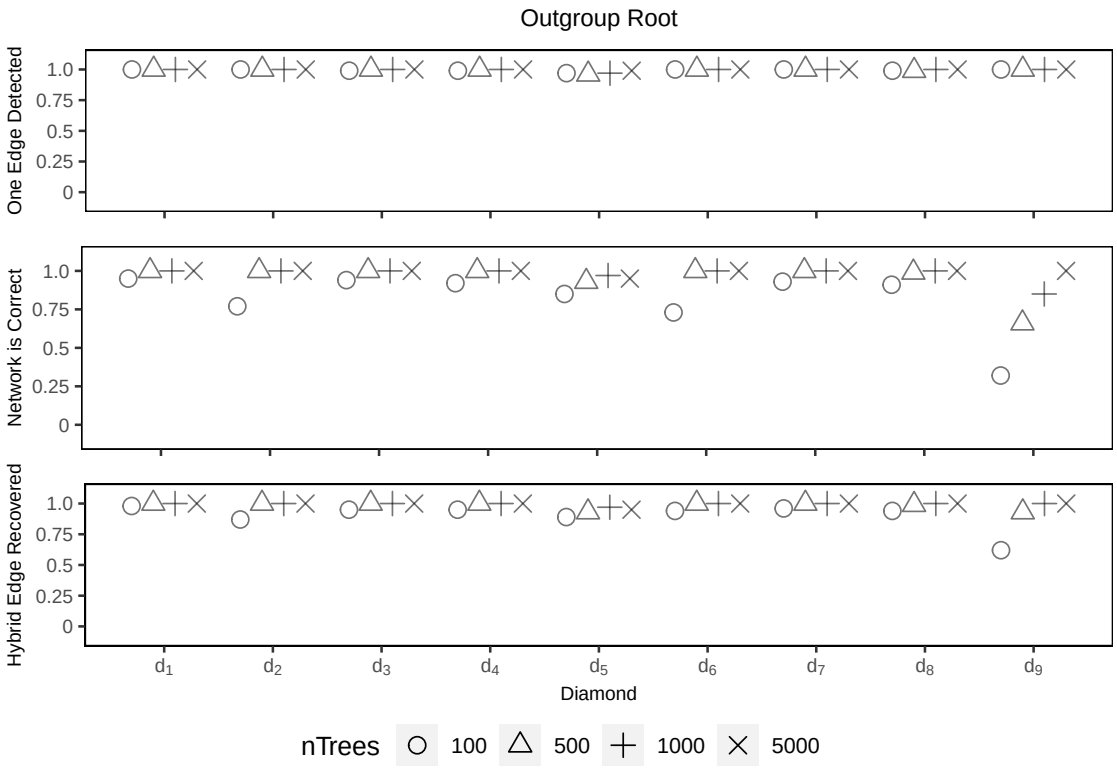

Figure 17: SNaQ results for root 3 with random outgroups for 10% of gene trees.

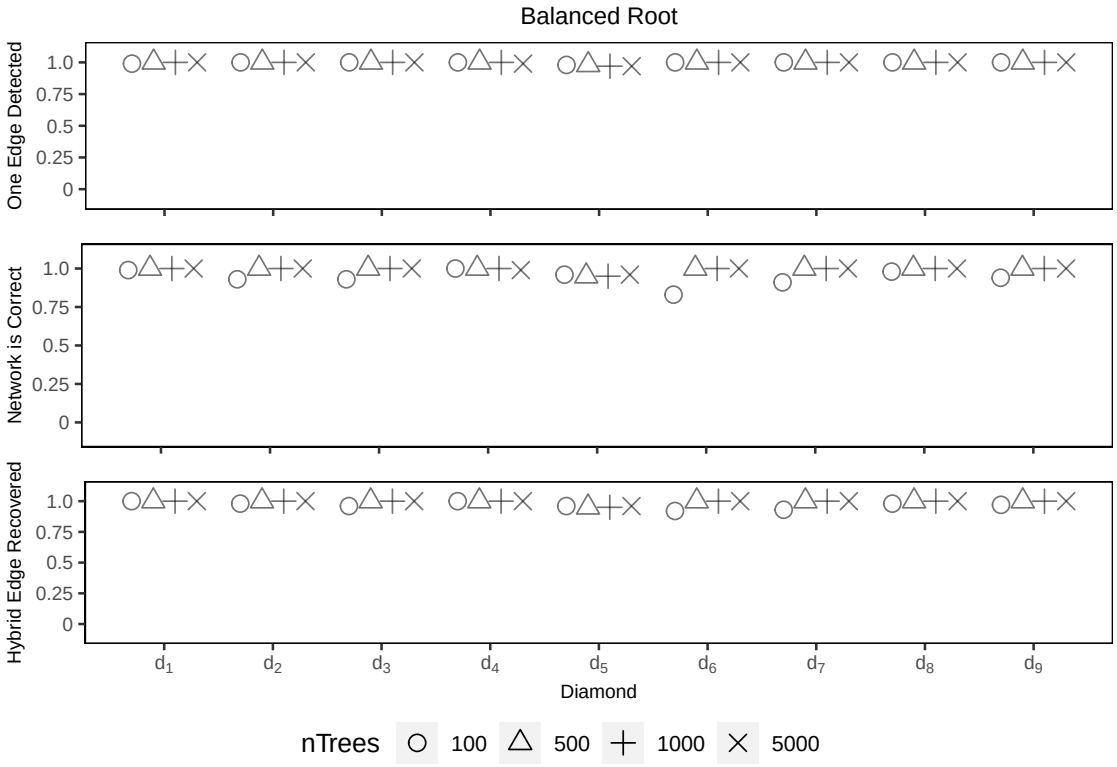

Figure 18: SNaQ results for root 1 with random outgroups for 30% of gene trees.

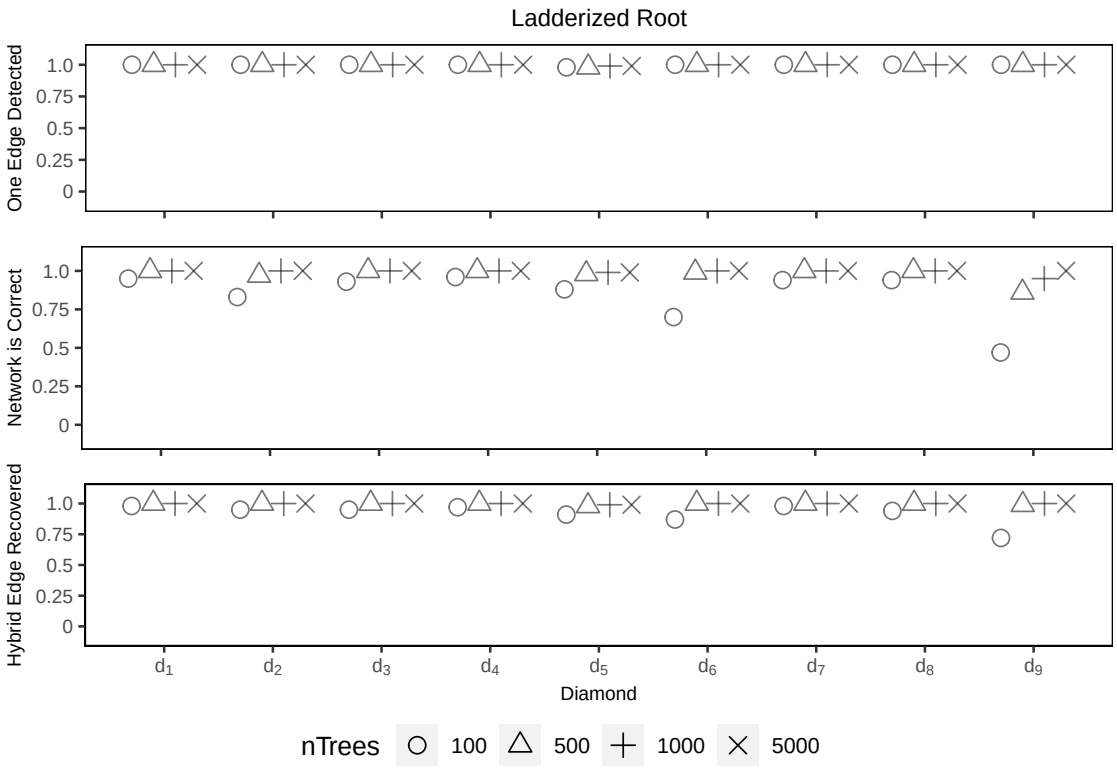

Figure 19: SNaQ results for root 2 with random outgroups for 30% of gene trees.

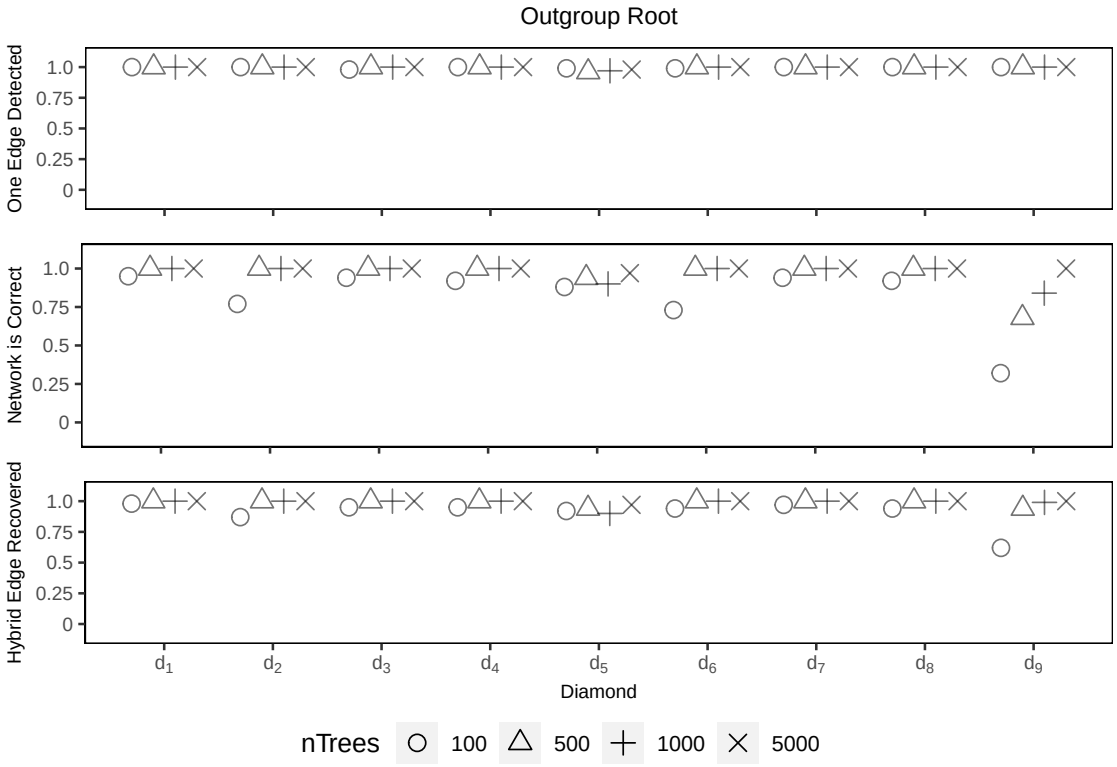

Figure 20: SNaQ results for root 3 with random outgroups for 30% of gene trees.

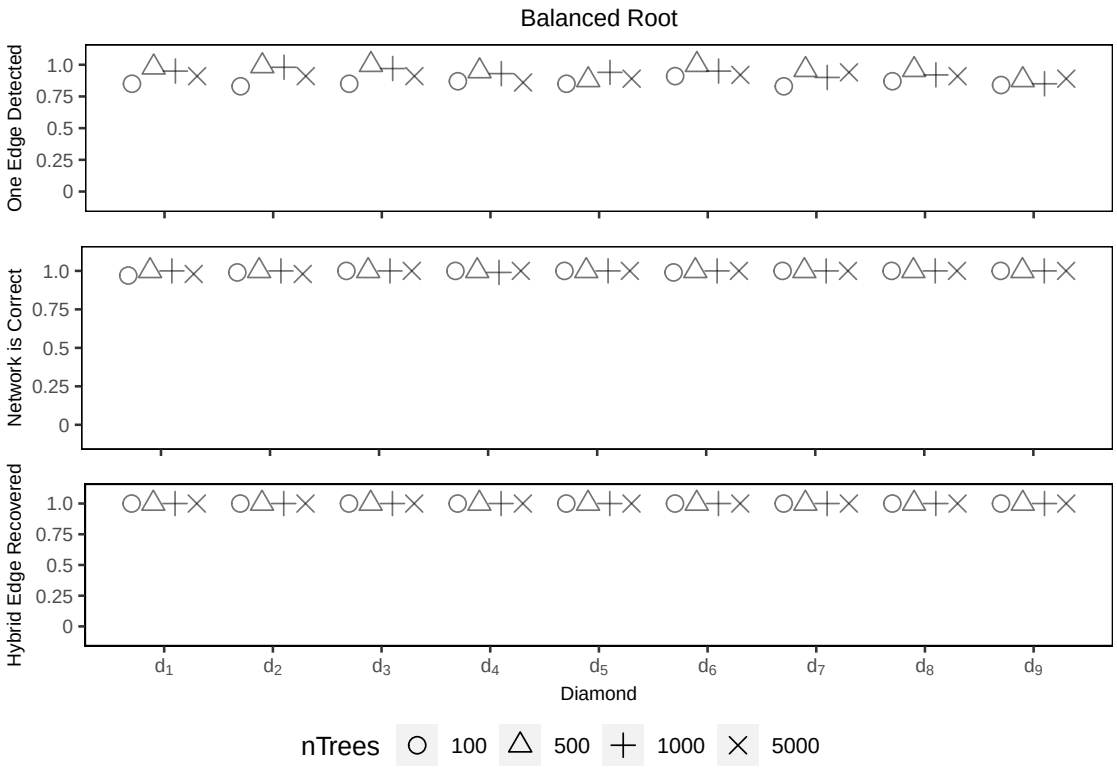

Figure 21: PhyloNet results for root 1 with random outgroups for 10% of gene trees.

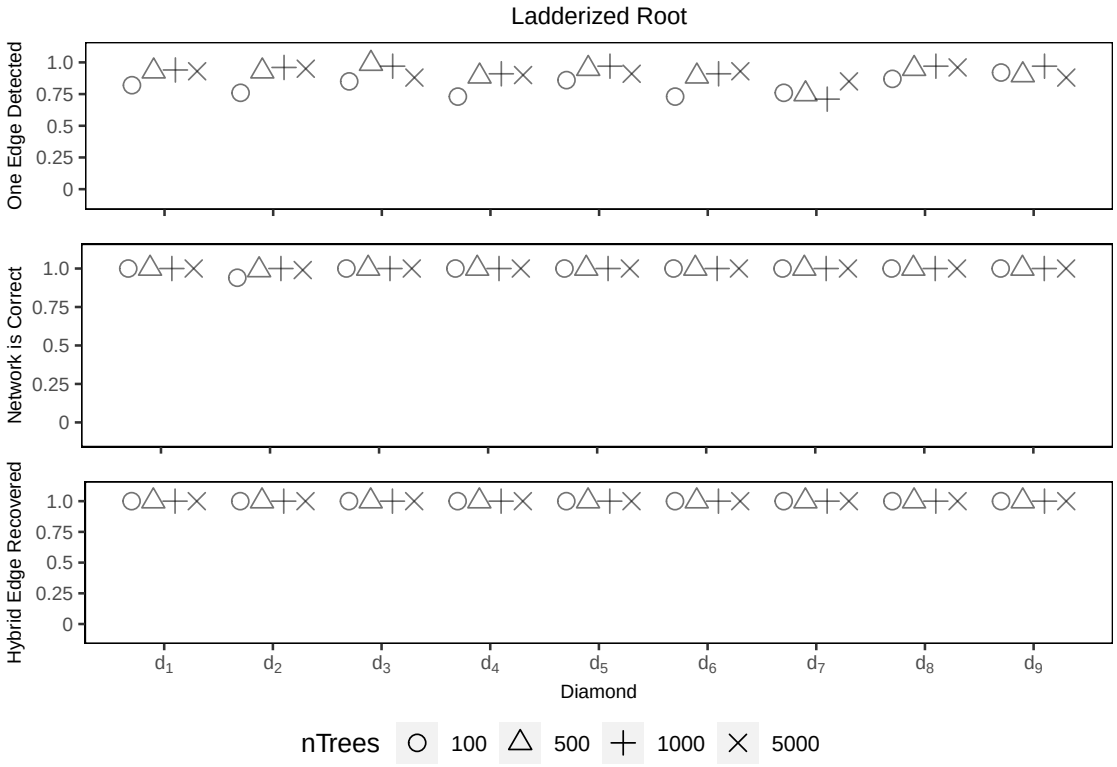

Figure 22: PhyloNet results for root 2 with random outgroups for 10% of gene trees.

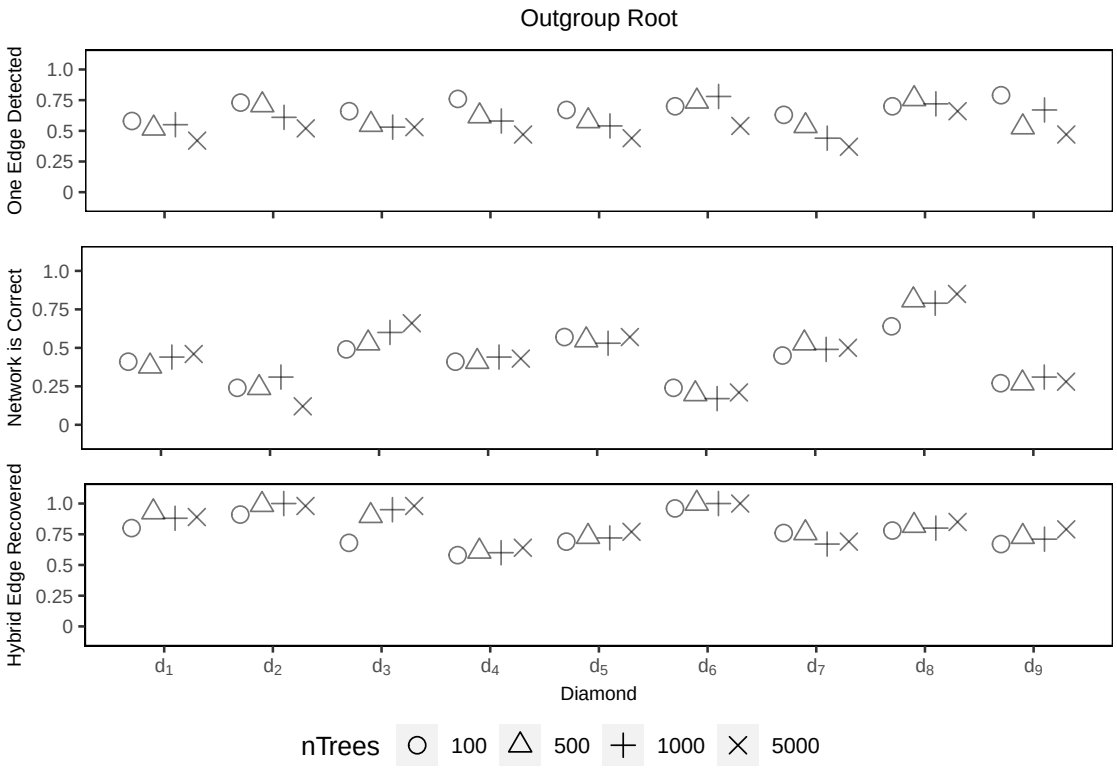

Figure 23: PhyloNet results for root 3 with random outgroups for 10% of gene trees.

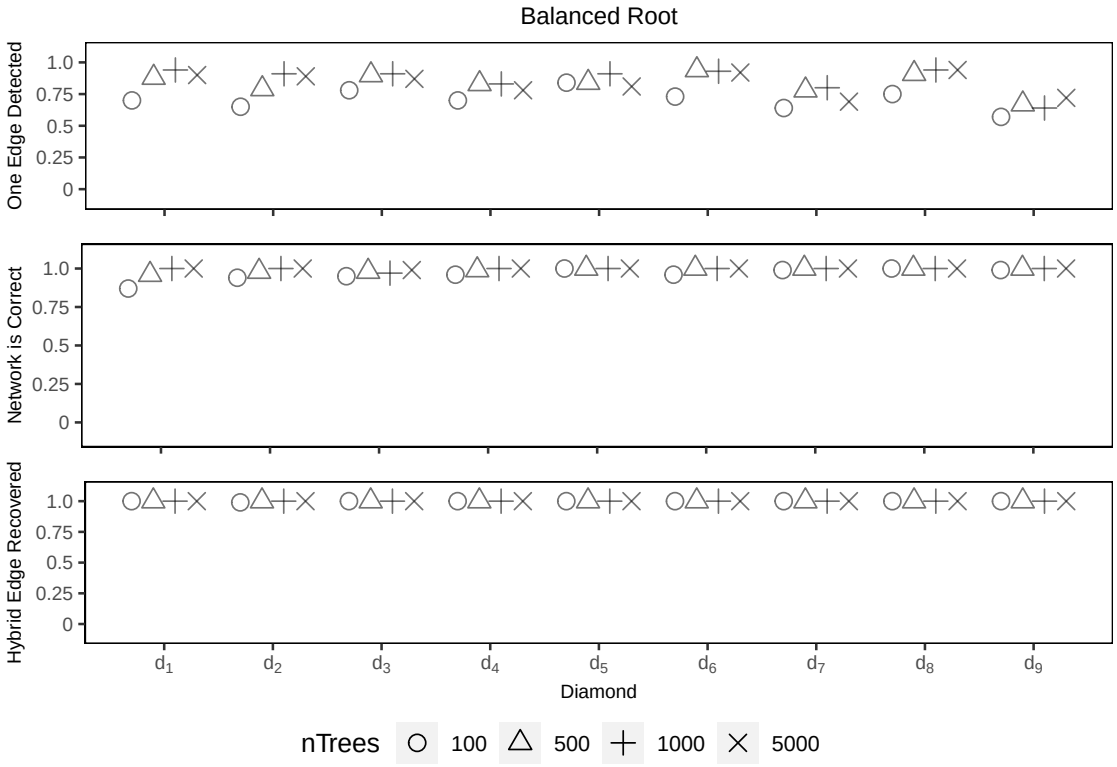

Figure 24: PhyloNet results for root 1 with random outgroups for 30% of gene trees.

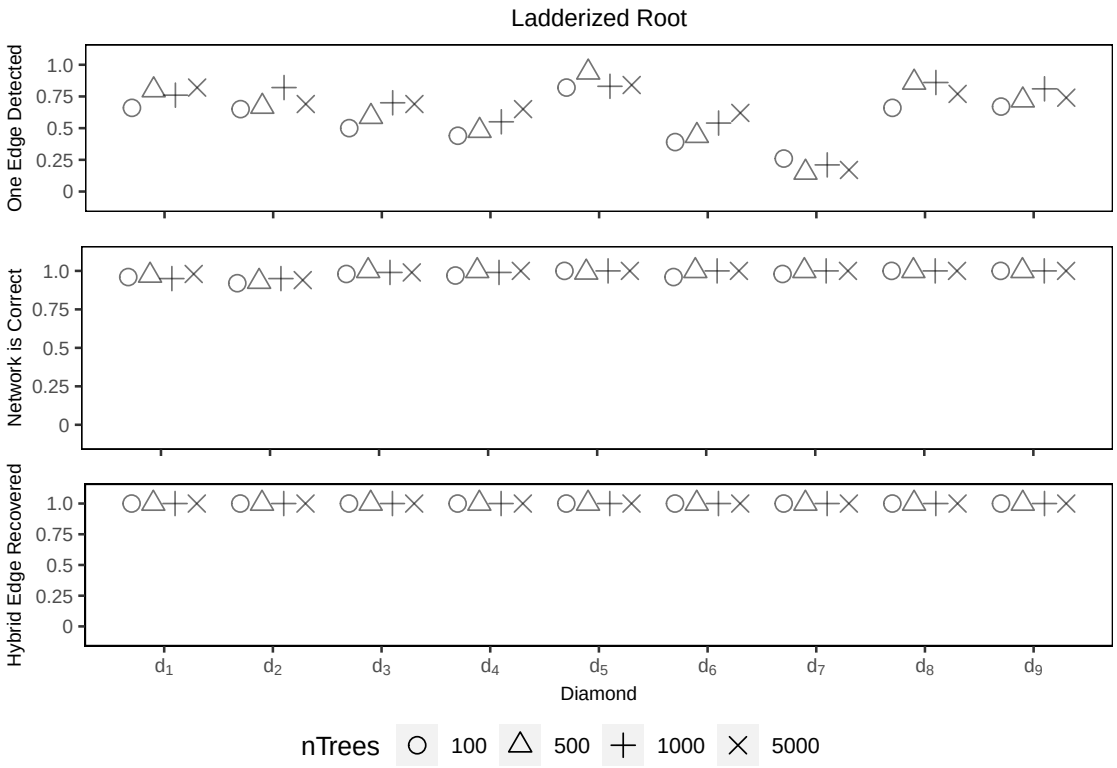

Figure 25: PhyloNet results for root 2 with random outgroups for 30% of gene trees.

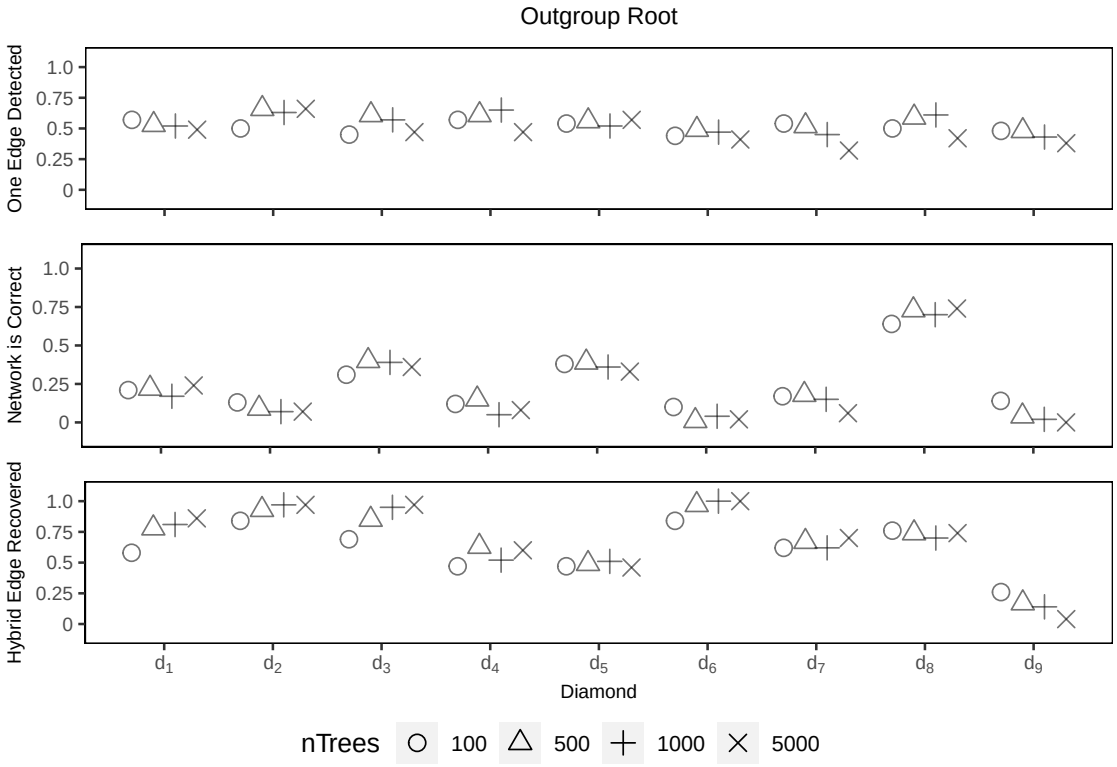

Figure 26: PhyloNet results for root 3 with random outgroups for 30% of gene trees.

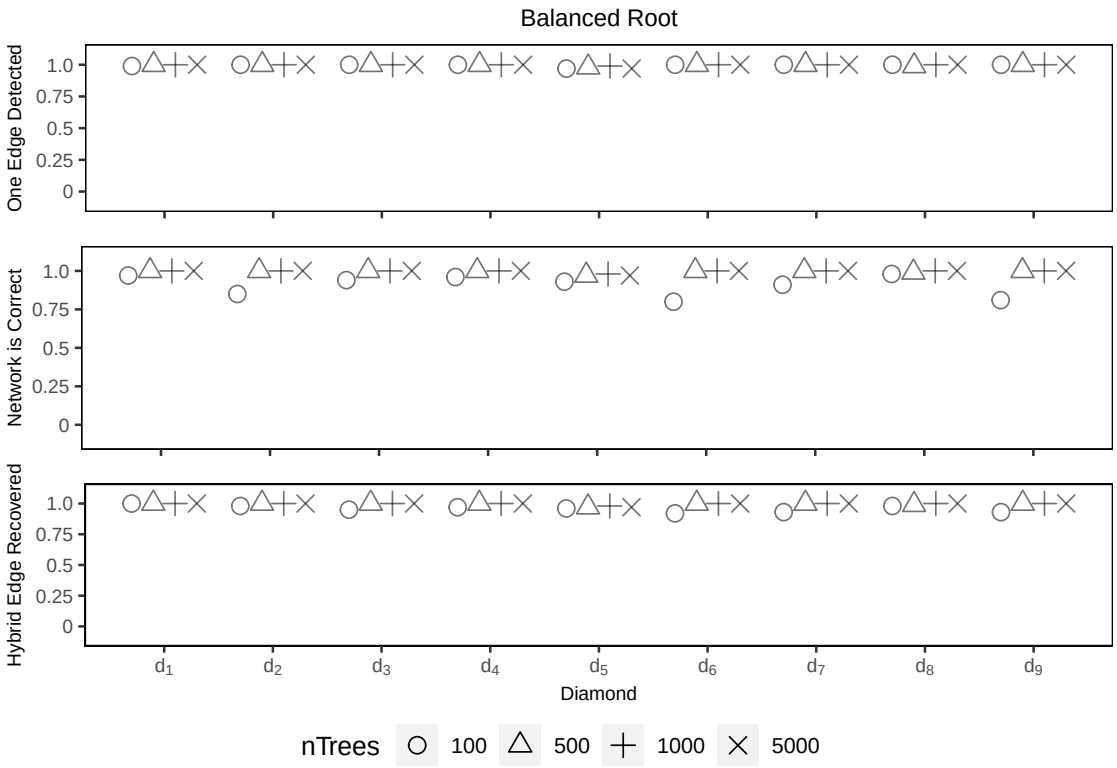

Figure 27: SNaQ results for root 1 with random outgroups and a random NNI move for 10% of gene trees.

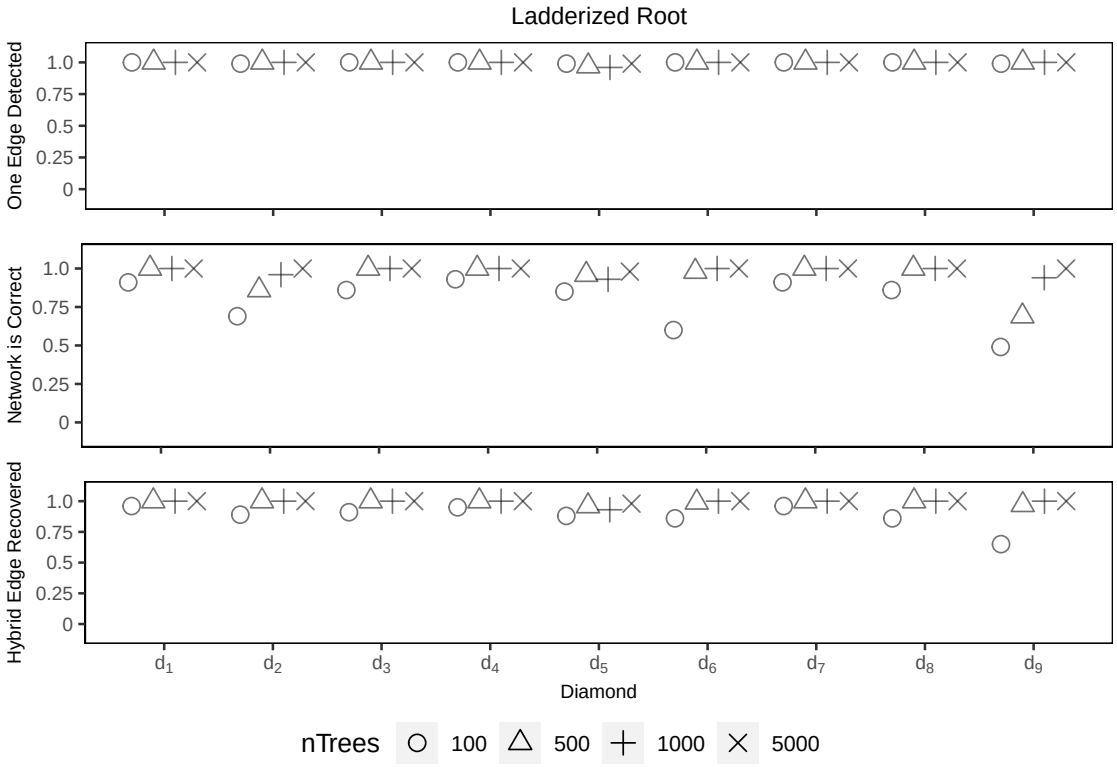

Figure 28: SNaQ results for root 2 with random outgroups and a random NNI move for 10% of gene trees.

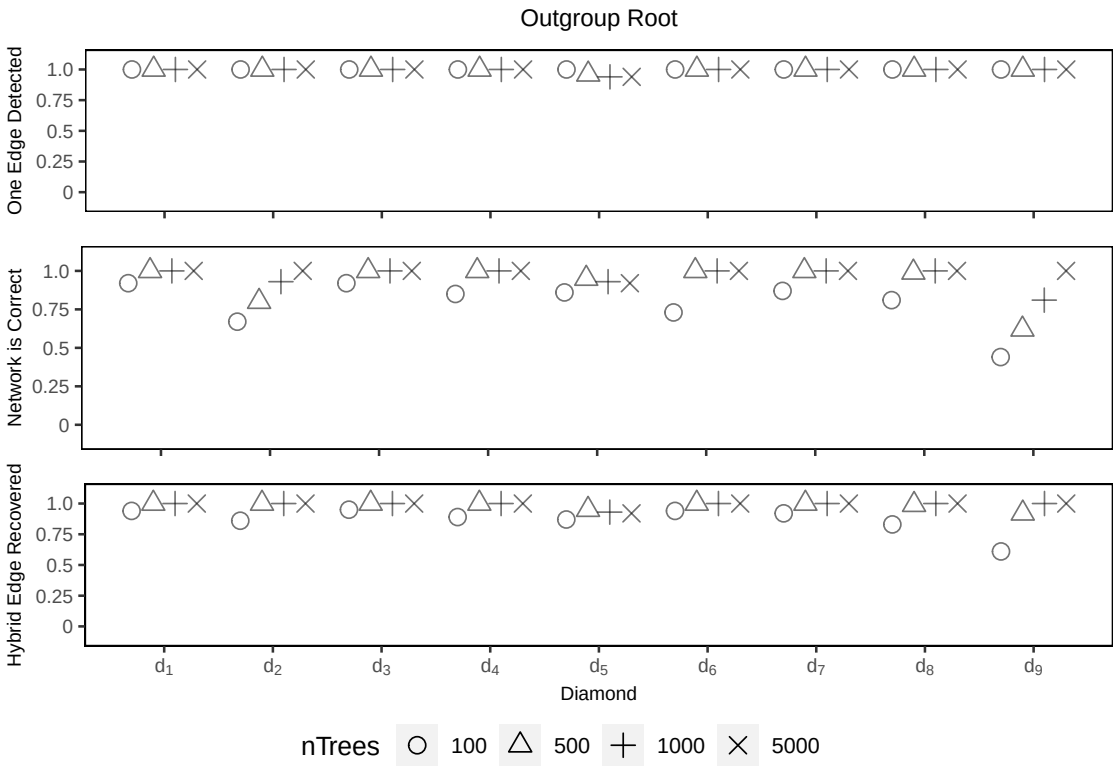

Figure 29: SNaQ results for root 3 with random outgroups and a random NNI move for 10% of gene trees.

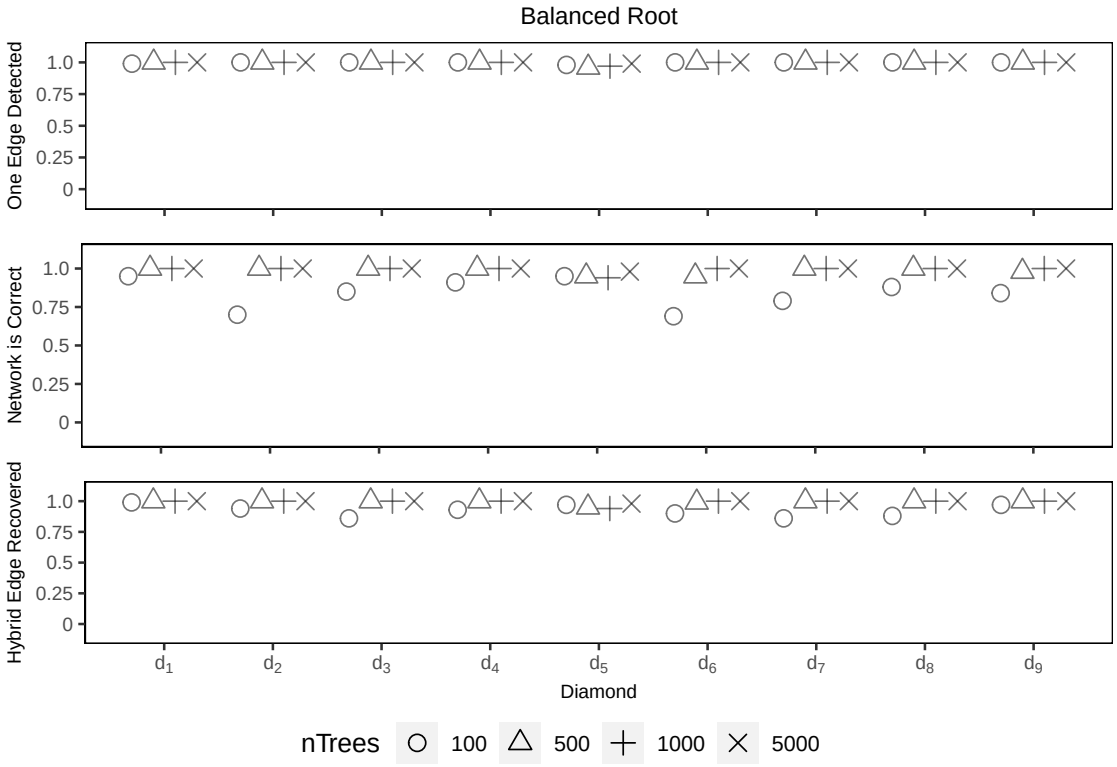

Figure 30: SNaQ results for root 1 with random outgroups and a random NNI move for 30% of gene trees.

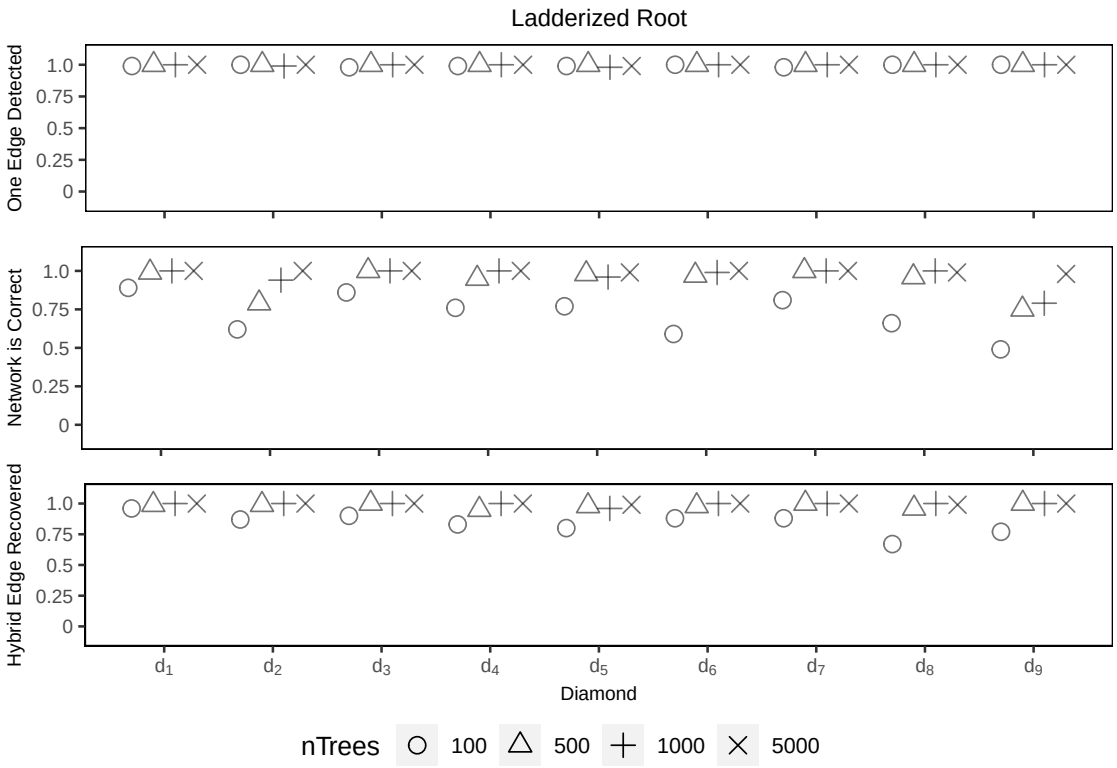

Figure 31: SNaQ results for root 2 with random outgroups and a random NNI move for 30% of gene trees.

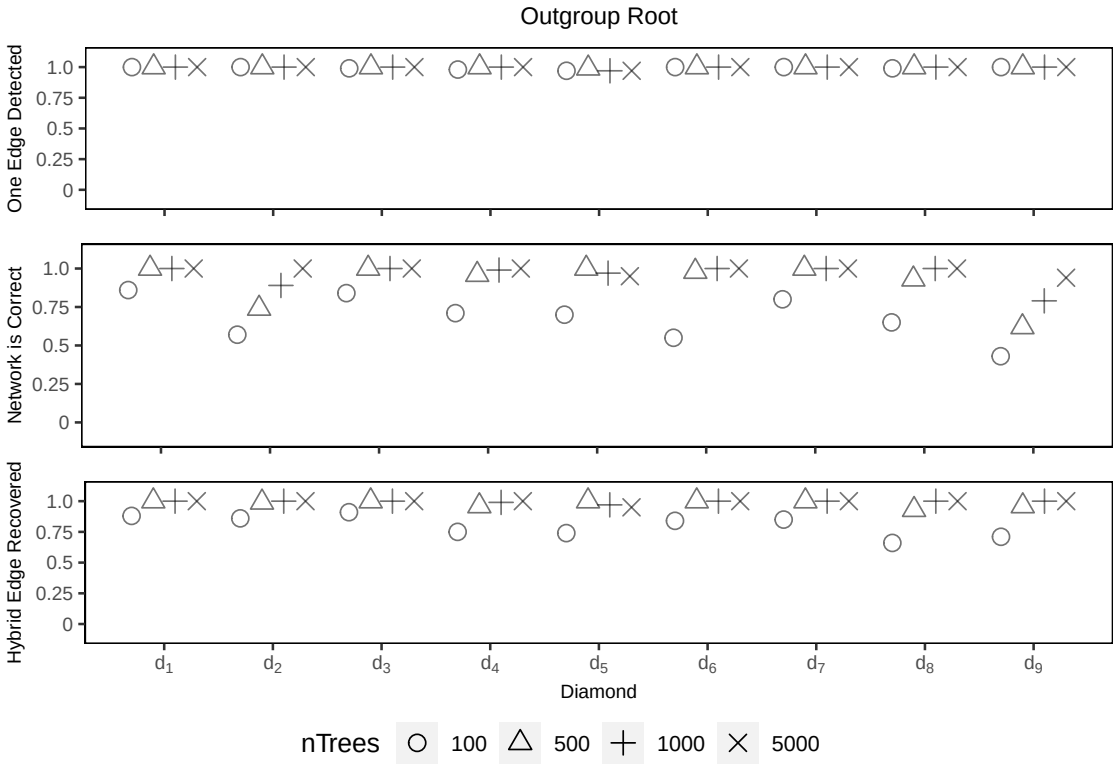

Figure 32: SNaQ results for root 3 with random outgroups and a random NNI move for 30% of gene trees.

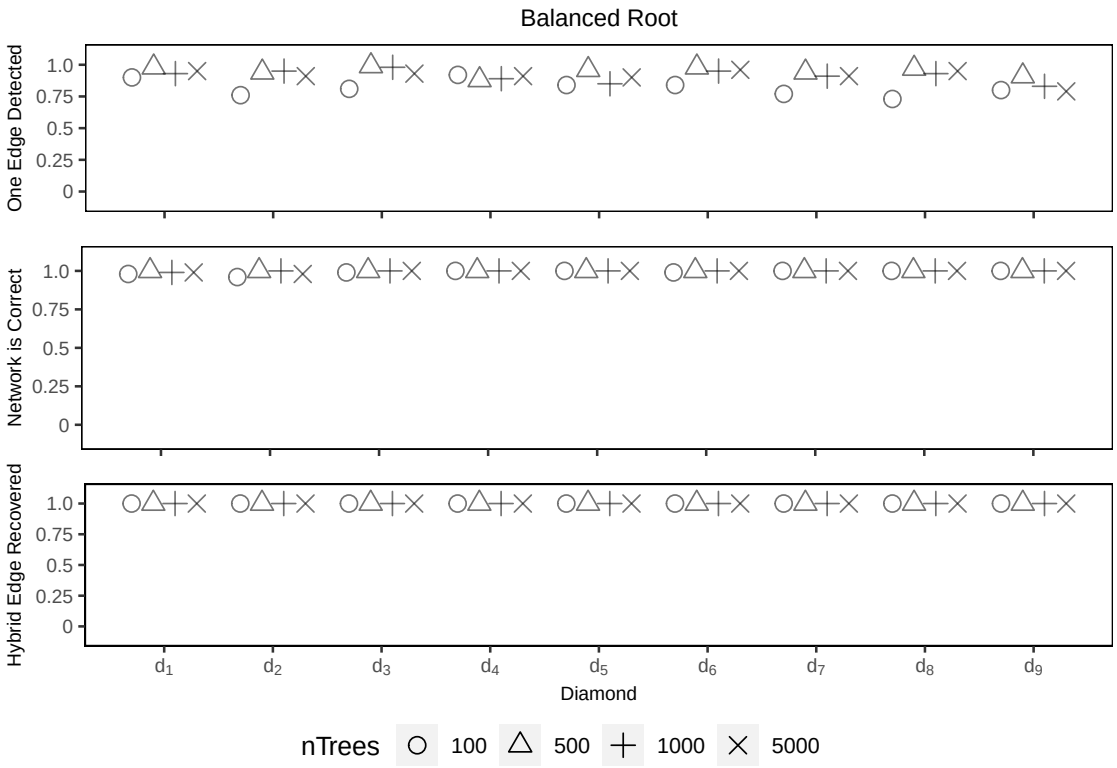

Figure 33: PhyloNet results for root 1 with random outgroups and a random NNI move for 10% of gene trees.

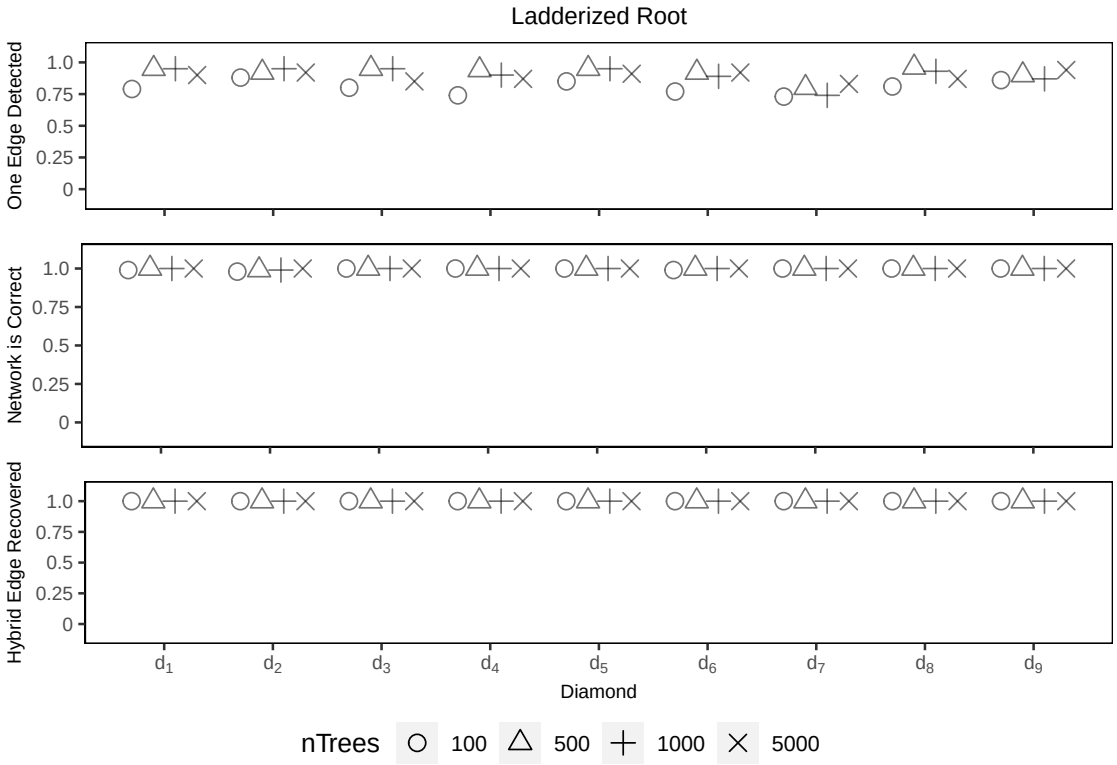

Figure 34: PhyloNet results for root 2 with random outgroups and a random NNI move for 10% of gene trees.

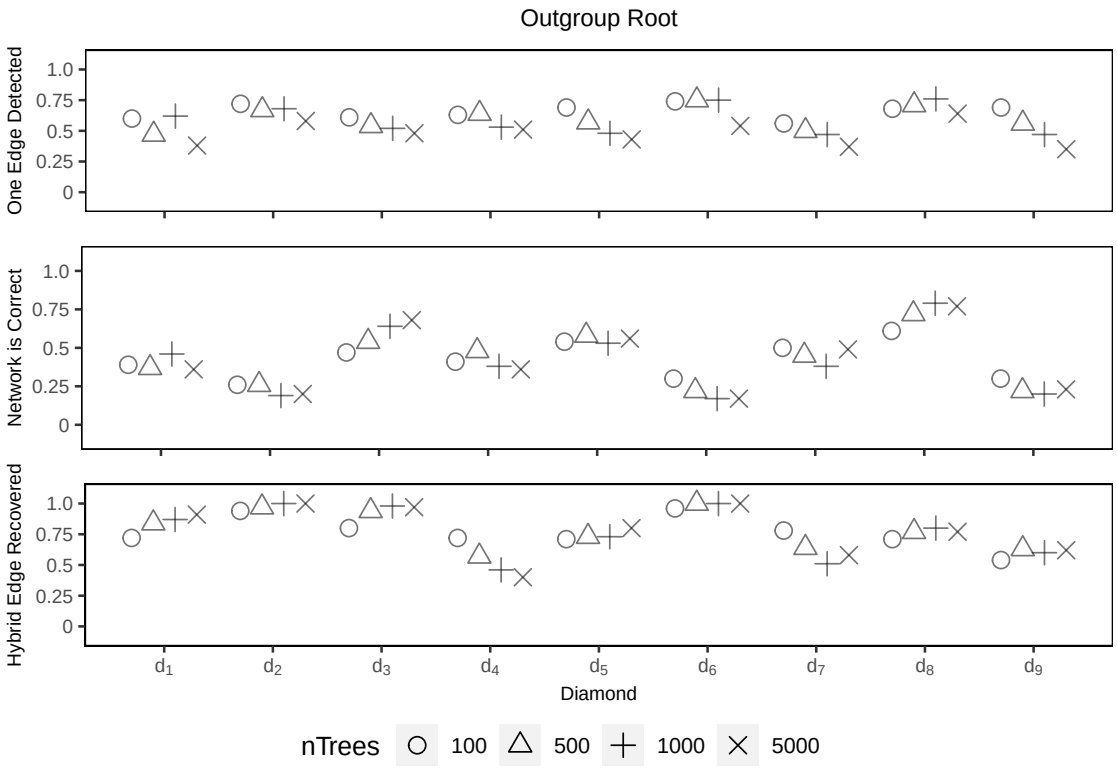

Figure 35: PhyloNet results for root 3 with random outgroups and a random NNI move for 10% of gene trees.

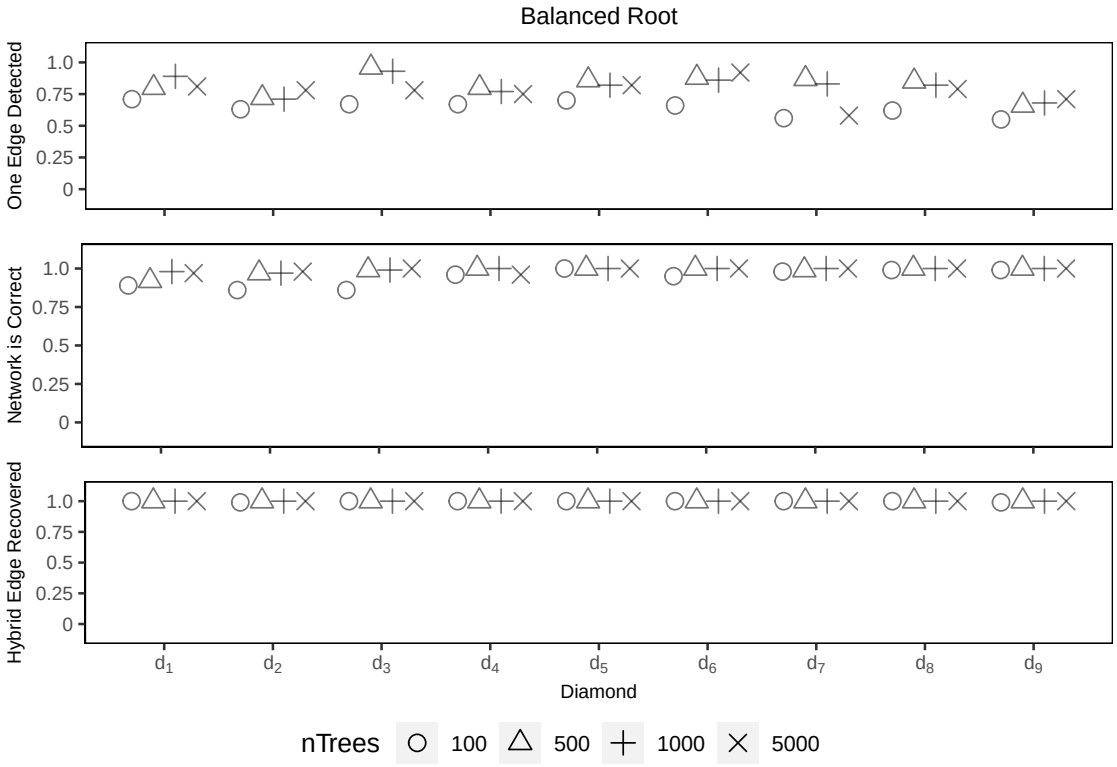

Figure 36: PhyloNet results for root 1 with random outgroups and a random NNI move for 30% of gene trees.

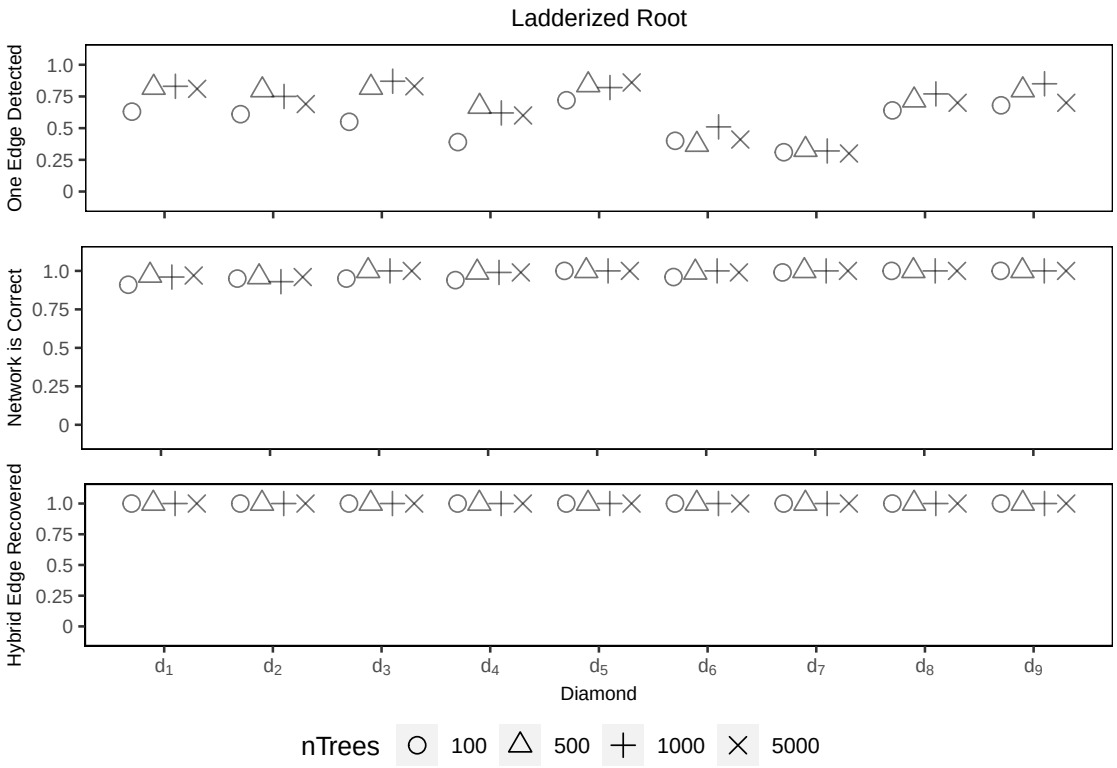

Figure 37: PhyloNet results for root 2 with random outgroups and a random NNI move for 30% of gene trees.

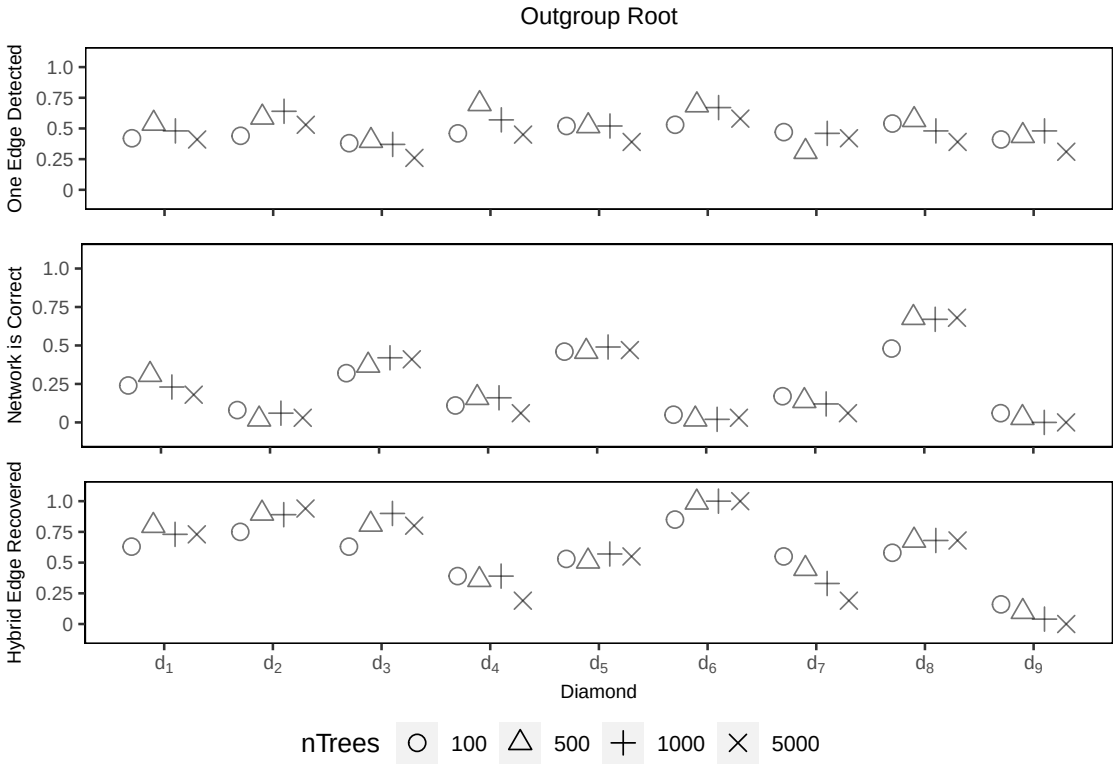

Figure 38: PhyloNet results for root 3 with random outgroups and a random NNI move for 30% of gene trees.

Method	Root	Root Error	NNI Error	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9
SNaQ	r_1	0	0	1	1	1	1	0.94	1	1	1	1
SNaQ	r_1	0.1	0	1	1	1	1	0.98	1	1	0.99	1
SNaQ	r_1	0.1	0.1	1	1	1	1	0.97	1	1	1	1
SNaQ	r_1	0.3	0	1	1	1	0.99	0.96	1	1	1	1
SNaQ	r_1	0.3	0.3	1	1	1	1	0.98	1	1	1	1
SNaQ	r_2	0	0	1	1	1	0.99	0.95	1	1	1	1
SNaQ	r_2	0.1	0	1	1	1	1	0.97	1	1	1	1
SNaQ	r_2	0.1	0.1	1	1	1	1	0.98	1	1	1	1
SNaQ	r_2	0.3	0	1	1	1	1	0.99	1	1	1	1
SNaQ	r_2	0.3	0.3	1	1	1	1	0.99	1	1	0.99	0.98
SNaQ	r_3	0	0	0.99	1	1	1	0.97	1	1	1	1
SNaQ	r_3	0.1	0	1	1	1	1	0.95	1	1	1	1
SNaQ	r_3	0.1	0.1	1	1	1	1	0.92	1	1	1	1
SNaQ	r_3	0.3	0	1	1	1	1	0.97	1	1	1	1
SNaQ	r_3	0.3	0.3	1	1	1	1	0.95	1	1	1	0.94
InferNetwork_MPL	r_1	0	0	1	1	1	1	1	1	1	1	1
InferNetwork_MPL	r_1	0.1	0	0.98	0.98	1	1	1	1	1	1	1
InferNetwork_MPL	r_1	0.1	0.1	0.99	0.98	1	1	1	1	1	1	1
InferNetwork_MPL	r_1	0.3	0	1	1	0.99	1	1	1	1	1	1
InferNetwork_MPL	r_1	0.3	0.3	0.97	0.98	1	0.96	1	1	1	1	1
InferNetwork_MPL	r_2	0	0	1	1	1	1	1	1	1	1	1
InferNetwork_MPL	r_2	0.1	0	1	0.99	1	1	1	1	1	1	1
InferNetwork_MPL	r_2	0.1	0.1	1	1	1	1	1	1	1	1	1
InferNetwork_MPL	r_2	0.3	0	0.98	0.94	0.99	1	1	1	1	1	1
InferNetwork_MPL	r_2	0.3	0.3	0.97	0.96	1	0.99	1	0.99	1	1	1
InferNetwork_MPL	r_3	0	0	0.85	0.82	0.86	0.85	0.76	0.88	0.95	0.9	0.87
InferNetwork_MPL	r_3	0.1	0	0.46	0.12	0.66	0.43	0.57	0.21	0.5	0.85	0.28
InferNetwork_MPL	r_3	0.1	0.1	0.36	0.2	0.68	0.36	0.56	0.17	0.49	0.77	0.23
InferNetwork_MPL	r_3	0.3	0	0.24	0.07	0.36	0.08	0.33	0.02	0.06	0.74	0
InferNetwork_MPL	r_3	0.3	0.3	0.18	0.03	0.41	0.06	0.47	0.03	0.06	0.68	0

Table 15: Proportion of correct networks recovered when using 5,000 gene trees. The Root Error and NNI Error are the proportion of gene trees affected by a random re-rooting or NNI move, respectively.