

A Metric on the Space of Reduced Phylogenetic Networks

Luay Nakhleh

Abstract—Phylogenetic networks are leaf-labeled, rooted, acyclic, and directed graphs that are used to model reticulate evolutionary histories. Several measures for quantifying the topological dissimilarity between two phylogenetic networks have been devised, each of which was proven to be a metric on certain restricted classes of phylogenetic networks. A biologically motivated class of phylogenetic networks, namely, *reduced* phylogenetic networks, was recently introduced. None of the existing measures is a metric on the space of reduced phylogenetic networks. In this paper, we provide a metric on the space of reduced phylogenetic networks that is computable in time polynomial in the size of the networks.

Index Terms—Phylogeny, phylogenetic network, indistinguishability, reduced phylogenetic network, metric.

1 INTRODUCTION

PHYLOGENETIC trees model the vertical transmission of genetic material from ancestors to descendants. When reticulate evolutionary events, such as horizontal gene transfer and hybrid speciation, occur, the evolutionary history of the set of organisms is more appropriately modeled as a special rooted, directed, acyclic graph, called *phylogenetic network*.

Measuring the distance between a pair of phylogenies plays an important role in a variety of tasks, including clustering, error estimation of reconstruction methods, and detection of species/gene tree incongruities.

A distance measure, or metric m on a space S satisfies three properties for all $a, b, c \in S$:

- P1. $m(a, b) = 0$ if and only if $a = b$,
- P2. $m(a, b) = m(b, a)$, and
- P3. $m(a, b) + m(b, c) \geq m(a, c)$.

In addition to these three properties, in phylogenetics, it is desired that the measure provides information about the similarity of the two evolutionary scenarios represented by the two phylogenies being compared. For example, viewing each branch of a phylogenetic tree as indicative of a bipartition of the set of taxa at the leaves, the Robinson-Foulds measure quantifies the average number of bipartitions that the two trees disagree on [1]. As another example, the *subtree prune and regraft* (SPR) measure quantifies an edit distance that is correlated with the number of reticulate evolutionary events required to reconcile the two trees being compared [2].

To illustrate a metric that is of little or no utility in phylogenetic comparison, consider the measure m on the space of all phylogenetic trees, where $m(P_1, P_2) = 0$ if the

two phylogenies P_1 and P_2 are isomorphic (where the bijection also respects the leaf labeling), and $m(P_1, P_2) = 1$ otherwise. While it is clearly a metric on the space of phylogenetic trees, such a “binary” measure tells whether two phylogenetic trees are identical (up to isomorphism) or not, but does not quantify the degree of similarity/dissimilarity between them.

The literature on general labeled graphs contains metrics that can be applied to phylogenetic networks, e.g., [3], [4], but the results they yield do not necessarily provide insight into the similarity or dissimilarity of the evolutionary history scenarios provided by a pair of phylogenetic networks. To address this issue, several measures have been introduced to quantify the dissimilarity between a pair of phylogenetic network topologies, each of which is metric on a restricted class of phylogenetic networks [5], [6], [7], [8], [9], [10], [11], [12].

In 2004, Moret et al. introduced the concepts of phylogenetic network *indistinguishability* and *reduced* phylogenetic networks to account for issues related to reconstructibility of phylogenetic networks [6]. In that paper, a dissimilarity measure was introduced, which was later shown not to satisfy some of the properties P1-P3, even on reduced phylogenetic networks [8]. In this paper, we review the concepts of phylogenetic network reduction and indistinguishability, and provide a novel metric on the space of reduced phylogenetic networks.

2 INDISTINGUISHABILITY AND REDUCED PHYLOGENETIC NETWORKS

In this section, we review the concepts of indistinguishability and reduced phylogenetic networks, which were first presented in [6]. We begin with general phylogenetic networks.

Definition 1. A phylogenetic \mathcal{X} -network, or \mathcal{X} -network for short, N is an ordered pair (G, f) , where¹

1. We use *indeg* and *outdeg* to denote the in-degree (number of parents) and out-degree (number of children) of a node, respectively.

• The author is with the Department of Computer Science, Rice University, 6100 Main Street, MS 132, Houston, TX 77005.
E-mail: nakhleh@cs.rice.edu.

Manuscript received 12 Dec. 2007; revised 10 July 2008; accepted 31 Dec. 2008; published online 6 Jan. 2009.

For information on obtaining reprints of this article, please send e-mail to: tcbb@computer.org, and reference IEEECS Log Number TCBB-2007-12-0174. Digital Object Identifier no. 10.1109/TCBB.2009.2.

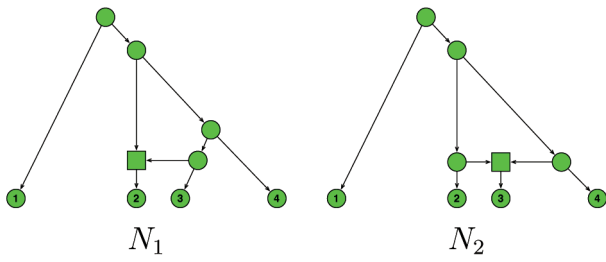


Fig. 1. Two phylogenetic \mathcal{X} -networks, with $\mathcal{X} = \{1, 2, 3, 4\}$.

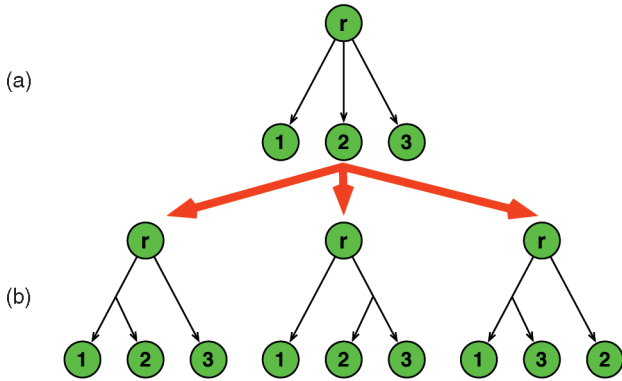


Fig. 2. (a) A soft polytomy and (b) its three possible refinements into binary trees.

- $G = (V, E)$ is a directed, acyclic graph (DAG) with V being the union of four pairwise-disjoint sets $\{r\}$, V_L , V_T , and V_N , where
 - $indeg(r) = 0$ (r is the root of N);
 - $\forall v \in V_L, indeg(v) = 1$ and $outdeg(v) = 0$ (V_L are the leaves of N);
 - $\forall v \in V_T, indeg(v) = 1$ and $outdeg(v) \geq 2$ (V_T are the tree nodes of N); and
 - $\forall v \in V_N, indeg(v) \geq 2$ and $outdeg(v) \geq 1$ (V_N are the network nodes of N),
 and $E \subseteq V \times V$ are the network's edges.

- $f : V_L \rightarrow \mathcal{X}$ is the leaf-labeling function, which is a bijection from V_L to \mathcal{X} .

We also use the notation $L(N)$ for the set of leaves of phylogenetic network N . Notice that the definition implies that the undirected graph underlying a phylogenetic network is connected. Fig. 1 shows two phylogenetic \mathcal{X} -networks, where $\mathcal{X} = \{1, 2, 3, 4\}$. When the context is clear, we may omit the labeling function f and write $N = (V, E)$. Further, when the set of taxa \mathcal{X} labeling the leaves of a phylogenetic \mathcal{X} -network is clear from the context, we may omit the set description and refer to it simply as a phylogenetic network.

In [6], Moret et al. discussed the issue of *indistinguishability* among phylogenetic networks from a reconstruction standpoint, and argued for the need of a measure that separates phylogenetic networks up to indistinguishability, or in other words, a measure that is a metric on the space of all *reduced* phylogenetic networks. In this section, we briefly review the concepts of distinguishability and reduced (phylogenetic) networks, and in Section 3, we present a metric on the space of all reduced networks that is computable in polynomial time in the size of the networks.

Moret et al. introduced the concept of indistinguishability to account for what we term here as “soft in-polytomy.” In the context of (rooted) phylogenetic trees, a (soft) polytomy is represented by a node that has more than two children, and indicates the lack of phylogenetic signal to resolve, or refine, the evolutionary relationship among these children. Fig. 2 shows a soft polytomy at node r , along with the three possible refinements of that polytomy. Even though the true evolutionary history is one of the three possible refinements, due to lack of phylogenetic signal, the only scenario that might be reconstructible is the polytomy.

In the case of phylogenetic networks, a lack of phylogenetic signal may also result in nodes with more than two parents. To illustrate, consider the scenario in Fig. 3. Four genes, or markers, from five taxa 1, 2, 3, 4, and x yield “gene trees” that differ in the placement of x such that in each of the trees, it is a sibling of a different taxon. If this difference is due to, say, hybrid speciation, then x is clearly a hybrid.

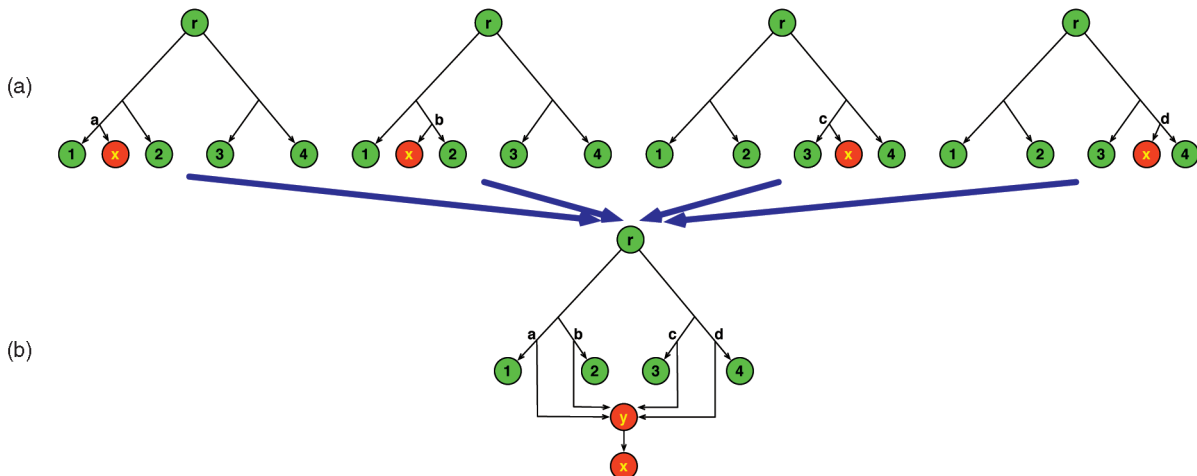


Fig. 3. A phylogenetic network with a soft in-polytomy at node y (b) resulting from the different placements of x in the four different gene trees (a). This network with in-polytomy can be refined into nine different networks in which each node has at most two parents, by adding two nodes x_1 and x_2 , and refining node y in all nine possible ways, as described in Table. 1.

TABLE 1

The Possible Refinements of Node y in the Phylogenetic Network in Fig. 3, Which Result in Networks in Which Each Nodes Has at Most Two Parents

Refinement	Parents of x_1	Parents of x_2	Parents of y
1	a, b	c, d	x_1, x_2
2	a, c	b, d	x_1, x_2
3	a, d	b, c	x_1, x_2
4	a, b	x_1, c	x_2, d
5	a, b	x_1, d	x_2, c
6	a, c	x_1, b	x_2, d
7	a, c	x_1, d	x_2, b
8	a, d	x_1, b	x_2, c
9	a, d	x_1, c	x_2, b

The phylogenetic network resulting from refinement (1) is shown in Fig. 4.

Nonetheless, due to lack of phylogenetic signal (in this case, it is massive extinction events or very “sparse” taxon sampling), the phylogenetic network that reconciles these four gene trees is one in which node y has in-degree 4, as shown in Fig. 3b. This phylogenetic network may be refined into nine different phylogenetic networks in which each node has at most two parents, as described in Table 1. Fig. 4 shows the network resulting from refinement (1) in the table. However, in the absence of any additional information, such as divergence times, selecting one of these refinements over the others is arbitrary. Using the terminology of [6], all these nine networks are indistinguishable from a reconstruction point of view, even though they are not isomorphic. Hence, Moret et al. introduced the concept of reducing these networks so as to eliminate the distinction among arbitrary refinements that are not supported by the data. In this case, the network in Fig. 3 is the reduced version of all these nine networks. In any of the networks resulting from the refinements described in Table 1, the set $\{x_1, x_2\}$ of nodes is referred to in [6] as a *maximal convergent set*. The network reduction procedure basically entails identifying maximal convergent sets in the network, and for each such set, connecting its parent nodes directly to the maximal subtrees, or clades, reachable from it, and eliminating all previously existing paths from the nodes in the set to these clades. We now review the formal definitions, as given in [6], of the three concepts of a maximal convergent set, a reduced network, and network indistinguishability.

Definition 2. Let $N = (V, E)$ be a phylogenetic \mathcal{X} -network. A set $U \subseteq V$ of internal nodes is convergent if 1) $|U| \geq 2$ and 2) every leaf reachable from some node in U is reachable from all nodes in U . If there exists no convergent set $U' \subseteq V$ such that $U \subset U'$, we say that U is a maximal convergent set.

The reduction procedure of [6] proceeds as follows,² when applied to phylogenetic \mathcal{X} -network $N = (V, E)$:

1. For each maximal subtree (or, clade) t (that includes no network nodes) of leaves $X' \subseteq \mathcal{X}$, rooted at node r_t , create a new unique node h_t and a new edge

² The reduction procedure of [6] is, in fact, inaccurate, in that it does not correctly capture all cases of indistinguishability. In this paper, we do not attempt to fix the procedure, but rather review it as it was given in [6].

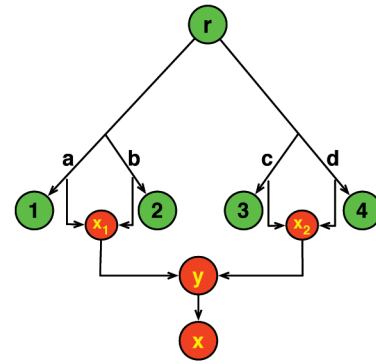


Fig. 4. Refinement (1) in Table 1 of the reduced phylogenetic network in Fig. 3.

(p_t, h_t) , where p_t is the parent of r_t , delete the edge (p_t, r_t) , and remove the subtree t . The node h_t becomes a *symbolic leaf* that represents the clade t . Let the resulting network be N' .

2. Repeat the following two steps on N' until no change occurs:
 - a. For each convergent set U with leaf set L_U , remove all vertices and edges on the paths from a vertex in U to a leaf in L_U , including all vertices in U and excluding vertices in L_U . For any edge (x, v) for which v is in the deleted set, replace it by a set of edges $\{(x, \ell) : \ell \in L_U \text{ is reachable from } v\}$.
 - b. For each node w in the network, with $\text{indeg}(w) = \text{outdeg}(w) = 1$, say with edges (u, w) and (w, v) , replace these two edges with a single edge (u, v) , remove node w , and remove any duplicate edges. Repeat until no such node exists.
3. Reattach to each symbolic leaf h_t the clade t by its root r_t .

Definition 3. Let $N = (V, E, f)$ be a phylogenetic \mathcal{X} -network. Its reduced version, denoted by $R(N)$, is the network obtained from N by application of the reduction procedure.

In Fig. 4, the only clade t found in Step 1 of the reduction procedure is the one that contains one leaf x . Its root is $r_t = x$ and its parent is $p_t = y$. A new node h_t is added, with the new edge (y, h_t) , and the clade (x) , along with the edge from y to it, is removed. Then, nodes x_1, x_2 , and y are removed along with all edges incident with them. Four new edges $(a, h_t), (b, h_t), (c, h_t),$ and (d, h_t) are then added. Finally, the clade (x) is reattached to node h_t , resulting in the network shown in Fig. 3 (with the node h_t labeled y).

We are now in a position to define network indistinguishability.

Definition 4. Two phylogenetic networks N_1 and N_2 are indistinguishable if their reduced versions $R(N_1)$ and $R(N_2)$ are isomorphic.

For example, all nine phylogenetic networks resulting from the node refinements described in Table 1 are (pairwise) indistinguishable. To determine if two phylogenetic networks are indistinguishable, one can first reduce them, and

then, compare their reduced versions. This requires a metric for comparing reduced phylogenetic networks, and we develop and present such a metric in the next section.

3 A METRIC ON THE SPACE OF REDUCED PHYLOGENETIC NETWORKS

The results of Cardona et al. [8] show that the tripartition-based measure introduced in [6] is not a metric on the space of reduced networks. In this section, we introduce a metric on the space of reduced phylogenetic \mathcal{X} -networks that is computable in time polynomial in the size of the networks. We begin with the notion of node equivalence.

Definition 5. Given a phylogenetic \mathcal{X} -network $N = ((V, E), f)$, we say that two nodes $u, v \in V$ are equivalent, denoted by $u \equiv v$, if

- $u, v \in V_L$ and $f(u) = f(v)$, or
- Node u has k children u_1, u_2, \dots, u_k , node v has k children v_1, v_2, \dots, v_k , and $u_i \equiv v_i$ for $1 \leq i \leq k$.

Since in this paper, we are concerned with comparing networks with identical leaf sets, the notion of node equivalence can be extended to nodes from two different networks, as established in the following equivalence mapping:

Definition 6. Let $N_1 = ((V_1, E_1), f_1)$ and $N_2 = ((V_2, E_2), f_2)$ be two phylogenetic \mathcal{X} -networks. We define the equivalence mapping between N_1 and N_2 , $h : V_1 \rightarrow 2^{V_2}$, so that $v \in h(u)$, for $u \in V_1$ and $v \in V_2$, if:

- $u \in L(N_1)$, $v \in L(N_2)$, and $f_1(u) = f_2(v)$, or
- Node u has k children u_1, u_2, \dots, u_k , node v has k children v_1, v_2, \dots, v_k , and $v_i \in h(u_i)$ for $1 \leq i \leq k$.

We have the following theorem:

Theorem 1. The equivalence of nodes as given in Definition 5 is an equivalence relation.

The proof is straightforward and follows from the properties of set equality.

Observation 1. Let $N = ((V, E), f)$ be a phylogenetic \mathcal{X} -network, and let $u, v \in V$ be two nodes where $u \equiv v$. Then, the set $\{u, v\}$ is convergent.

We have the following lemma:

Lemma 1. Let $N = ((V, E), f)$ be a reduced phylogenetic \mathcal{X} -network. If $\{u, v\} \subseteq V$ is a convergent set, then $u \not\equiv v$.

Proof. Based on the phylogenetic network reduction procedure, all convergent sets in a phylogenetic network are eliminated, except for one special type of convergent sets: a set of two nodes $\{u, v\}$, where u is a network node and v is the only child of u . In this case, u and v are not equivalent, since no child of v is equivalent to v . \square

Given two phylogenetic networks $N_1 = (V_1, E_1)$ and $N_2 = (V_2, E_2)$, and a node $v_1 \in V_1$, we call the set $h(v_1)v_1$'s mates in N_2 , where h is the equivalence mapping, as given in Definition 6. Notice that $h(v_1)$ is empty when v_1 has no equivalent nodes in N_2 . Further, while in phylogenetic trees, we always have $|h(v_1)| \leq 1$, it may be the case in general phylogenetic networks that $|h(v_1)| > 1$ for some

nodes. Since all nodes in $h(v_1)$ are pairwise equivalent, we use $h(v_1)$ to denote an arbitrary node in the set, and NIL when the set is empty.

Assume that $V_1 = \{v_1, v_2, \dots, v_p\}$. Then, the unique nodes of N_1 , denoted by $U(N_1)$, are the set $\{v_i : \forall j < i, v_j \not\equiv v_i, 1 \leq i \leq p\}$. We define $U(N_2)$ similarly. Further, for each node $v_i \in V_1$, we define $\kappa_{N_1}(v_i) = |\{v \in V_1 : v \equiv v_i\}|$, and $\kappa_{N_2}(u_i)$ similarly for each node $u_i \in V_2$. We define $\kappa(\text{NIL}) = 0$ for any network N . When the context is clear, we drop the subscript of κ . We are now in position to define the measure on pairs of phylogenetic \mathcal{X} -networks.

Definition 7. Let $N_1 = (V_1, E_1)$ and $N_2 = (V_2, E_2)$ be two phylogenetic \mathcal{X} -networks. Then, $m(N_1, N_2)$ equals

$$\frac{1}{2} \left(\sum_{v \in U(N_1)} \max\{0, \kappa(v) - \kappa(v')\} + \sum_{u \in U(N_2)} \max\{0, \kappa(u) - \kappa(u')\} \right), \quad (1)$$

where v' (u') is a node in N_2 (N_1) that is equivalent to v (u), and if no such equivalent node exists, then v' (u') is NIL.

The rationale behind the measure m is that it roughly quantifies the number of rooted subnetworks that are in one but not both of the networks. In the special case where the two networks are two rooted trees T_1 and T_2 , then $m(T_1, T_2)$ yields half the symmetric difference of their sets of rooted subtrees, where two subtrees from T_1 and T_2 are equal if they are isomorphic with respect to the leaf labels. Further, like the Robinson-Foulds metric [1], this measure is very sensitive to small perturbations in certain cases. For example, even though the only difference between the two networks in Fig. 1 is the orientation of a single edge, which is the edge between the parents of leaves 2 and 3, we have $m(N_1, N_2) = 5$. One of the most commonly used distance metrics for comparing phylogenetic trees, namely, the RF distance, has a similar property. For example, for the two trees $T_1 = (a, (b, (c, (d, (e, f))))))$ and $T_2 = (f, (a, (b, (c, (d, e))))))$, whose edit distance is 1, due to the different placement of the leaf f , have RF distance of 3, when considered unrooted, and 4 when considered rooted (and this effect can be further "dramatized" by considering larger such "caterpillar" trees).

We now establish properties of the measure m .

Lemma 2. If $m(N_1, N_2) = 0$ for two reduced phylogenetic networks $N_1 = (V_1, E_1)$ and $N_2 = (V_2, E_2)$, then:

1. $|V_1| = |V_2|$.
2. $\sum_{v_1 \in U(N_1)} \kappa(v_1) = \sum_{v_2 \in U(N_2)} \kappa(v_2)$.

Proof. Let $h_1 : V_1 \rightarrow V_2$ and $h_2 : V_2 \rightarrow V_1$ be two equivalence mappings, as given by Definition 6. Since $m(N_1, N_2) = 0$, it follows that $\kappa(v_1) = \kappa(h_1(v_1))$ for all $v_1 \in V_1$ and $\kappa(v_2) = \kappa(h_2(v_2))$ for all $v_2 \in V_2$. From this, both results follow. \square

Lemma 3. If $m(N_1, N_2) = 0$ for two reduced phylogenetic \mathcal{X} -networks $N_1 = ((V_1, E_1), f_1)$ and $N_2 = ((V_2, E_2), f_2)$, then the equivalence mappings $h_1 : V_1 \rightarrow V_2$ and $h_2 : V_2 \rightarrow V_1$ are both isomorphisms.

Proof. We show the proof for h_1 ; the proof for h_2 is identical. By Lemma 2, it follows that $|V_1| = |V_2|$, and from Observation 1 and Lemma 1, it follows that $h(v_1)$ is defined and unique for each $v_1 \in V_1$. We now show that if $(u, v) \in E_1$, then $(u', v') \in E_2$, where $u' = h_1(u)$ and $v' = h_1(v)$. Given that $u' = h_1(u)$, it follows that u' and u are equivalent, which, by the definition of equivalence, implies that both u and u' have equivalent children. By Observation 1 and Lemma 1, and from the assumption that $m(N_1, N_2) = 0$, N_2 has exactly one node that is equivalent to v , which must be v' . Therefore, $(u', v') \in E_2$. \square

We now prove that the measure m satisfies property P1 stated in Section 1.

Lemma 4. Let $N_1 = (V_1, E_1)$ and $N_2 = (V_2, E_2)$ be two reduced phylogenetic \mathcal{X} -networks. Then, N_1 and N_2 are isomorphic if and only if $m(N_1, N_2) = 0$.

Proof. For the “only if” direction, let $h_1 : V_1 \rightarrow V_2$ be the equivalence mapping as given in Definition 6. Mapping h_1 is a bijection, since N_1 and N_2 are isomorphic. Then, based on Definition 7, we have $m(N_1, N_2) = 0$.

The “if” direction follows directly from Lemma 3. \square

From the definition of the measure, property P2 follows immediately.

Lemma 5. For any pair of phylogenetic \mathcal{X} -networks N_1 and N_2 , we have $m(N_1, N_2) = m(N_2, N_1)$.

The measure $m(N_1, N_2)$ can be viewed as half the symmetric difference of two multisets on the same set of elements, where the multiplicity of element u in N_1 is $\kappa_{N_1}(u)$, and similarly, for N_2 . Since the symmetric difference defines a metric on multisets [13], we have the following result:

Lemma 6. Let N_1, N_2 , and N_3 be three phylogenetic \mathcal{X} -networks. Then, $m(N_1, N_2) + m(N_2, N_3) \geq m(N_1, N_3)$.

From Lemmas 4, 5, and 6, we have the following main result:

Theorem 2. The measure m is a metric on the space of reduced phylogenetic networks.

We have proved that the measure m , as given by Definition 7, is a metric on the space of all reduced phylogenetic \mathcal{X} -networks. Finally, it is worth noting that the metric is computable in polynomial time in the size of the networks.

4 CONCLUSIONS

In this paper, we reviewed the concepts of phylogenetic network indistinguishability and reduction, and devised a polynomially-computable metric on the space of all reduced phylogenetic networks. This fixes the problem with the measure introduced in [6], which was later shown not to be metric on the space of reduced phylogenetic networks [8]. To determine whether two phylogenetic \mathcal{X} -networks are indistinguishable, the two networks are first reduced, using the reduction procedure of [6] (reviewed in Section 2), and the measure m is applied to the two resulting networks.

ACKNOWLEDGMENTS

The author would like to thank Gabriel Cardona, C. Randal Linder, Bernard M.E. Moret, Francesc Rosselló, Gabriel Valiente, Tandy Warnow, and the three anonymous reviewers for very helpful comments on the manuscript. This work was supported in part by the US Department of Energy (DOE) grant DE-FG02-06ER25734, the US National Science Foundation (NSF) grant CCF-0622037, and grant R01LM009494 from the US National Library of Medicine. The contents are solely the responsibility of the author and do not necessarily represent the official views of the DOE, NSF, US National Library of Medicine, or the US National Institutes of Health.

REFERENCES

- [1] D. Robinson and L. Foulds, “Comparison of Phylogenetic Trees,” *Math. Biosciences*, vol. 53, pp. 131-147, 1981.
- [2] L. Nakhleh, T. Warnow, and C. Linder, “Reconstructing Reticulate Evolution in Species-Theory and Practice,” *Proc. Eighth Ann. Int’l Conf. Computational Molecular Biology*, pp. 337-346, 2004.
- [3] H. Bunke and K. Shearer, “A Graph Distance Metric Based on the Maximal Common Subgraph,” *Pattern Recognition Letters*, vol. 19, pp. 255-259, 1998.
- [4] H. Bunke, X. Jiang, and A. Kandel, “On the Minimum Common Supergraph of Two Graphs,” *Computing*, vol. 65, pp. 13-25, 2000.
- [5] L. Nakhleh, J. Sun, T. Warnow, R. Linder, B. Moret, and A. Tholse, “Towards the Development of Computational Tools for Evaluating Phylogenetic Network Reconstruction Methods,” *Proc. Eighth Pacific Symp. Biocomputing*, pp. 315-326, 2003.
- [6] B. Moret, L. Nakhleh, T. Warnow, C. Linder, A. Tholse, A. Padolina, J. Sun, and R. Timme, “Phylogenetic Networks: Modeling, Reconstructibility, and Accuracy,” *IEEE/ACM Trans. Computational Biology and Bioinformatics*, vol. 1, no. 1, pp. 13-23, Jan.-Mar. 2004.
- [7] M. Baroni, C. Semple, and M. Steel, “A Framework for Representing Reticulate Evolution,” *Annals of Combinatorics*, vol. 8, pp. 391-408, 2004.
- [8] G. Cardona, F. Rosselló, and G. Valiente, “Tripartitions Do Not Always Discriminate Phylogenetic Networks,” *Math. Biosciences*, vol. 211, no. 2, pp. 356-370, 2008.
- [9] G. Cardona, M. Llabrés, F. Rosselló, and G. Valiente, “A Distance Metric for a Class of Tree-Sibling Phylogenetic Networks,” *Bioinformatics*, vol. 24, no. 13, pp. 1481-1488, 2008.
- [10] G. Cardona, M. Llabrés, F. Rosselló, and G. Valiente, “Metrics for Phylogenetic Networks I: Generalizations of the Robinson-Foulds Metric,” *IEEE/ACM Trans. Computational Biology and Bioinformatics*, vol. 6, no. 1, pp. 46-61, Jan.-Mar. 2009.
- [11] G. Cardona, F. Rosselló, and G. Valiente, “Comparison of Tree-Child Phylogenetic Networks,” *IEEE/ACM Trans. Computational Biology and Bioinformatics*, vol. 6, no. 4, pp. 552-569, Oct.-Dec. 2009.
- [12] G. Cardona, M. Llabrés, F. Rosselló, and G. Valiente, “Metrics for Phylogenetic Networks II: Nodal and Triplets Metrics,” *IEEE/ACM Trans. Computational Biology and Bioinformatics*, vol. 6, no. 3, pp. 454-469, July-Sept. 2009.
- [13] F. Restle, “A Metric and an Ordering on Sets,” *Psychometrika*, vol. 24, no. 3, pp. 207-220, 1959.



Luay Nakhleh received the BSc degree in computer science from the Technion—Israel Institute of Technology in 1996, the master’s degree in computer science from Texas A&M University in 1998, and the PhD degree in computer science from The University of Texas at Austin. He is an assistant professor of computer science at Rice University. His research interests fall in the general areas of computational biology and bioinformatics; in particular, he works on computational phylogenetics, comparative genomics, and biological network analysis. He received the Roy E. Campbell Faculty Development Award from Rice University in May 2006 and the US Department of Energy (DOE) Early Career Award in August 2006.