Transforming phylogenetic networks: Moving beyond tree space

Katharina T. Huber, Vincent Moulton, Taoyang Wu *

School of Computing Sciences, University of East Anglia, Norwich NR4 7TJ, UK

HIGHLIGHTS

- Two new operations are introduced for transforming unrooted phylogenetic networks.
- Any network can be transformed into any other network with the same size using NNI operations.
- Any network can be transformed into any other using only two operations.
- New connections between phylogenetic networks and cubic graphs are employed.
- Novel spaces of networks arise which generalize phylogenetic tree space.

ARTICLE INFO

ABSTRACT

Phylogenetic networks are a generalization of phylogenetic trees that are used to represent reticulate evolution. Unrooted phylogenetic networks form a special class of such networks, which naturally generalize unrooted phylogenetic trees. In this paper we define two operations on unrooted phylogenetic networks, one of which is a generalization of the well-known nearest-neighbor interchange (NNI) operation on phylogenetic trees. We show that any unrooted phylogenetic network can be transformed into any other such network using only these operations. This generalizes the well-known fact that any phylogenetic tree can be transformed into any other such tree using only NNI operations. It also allows us to define a generalization of tree space and to define some new metrics on unrooted phylogenetic networks. To prove our main results, we employ some fascinating new connections between phylogenetic networks and cubic graphs that we have recently discovered. Our results should be useful in developing new strategies to search for optimal phylogenetic networks, a topic that has recently generated some interest in the literature, as well as for providing new ways to compare networks.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Phylogenetic networks are a generalization of phylogenetic trees that are gaining growing acceptance by biologists due to their importance in representing reticulate evolution (Bapteste et al., 2013). Certain types of networks, such as neighbornets (see e.g. Bryant and Moulton, 2004; Huson and Bryant, 2006) and median networks (see e.g. Bandelt et al., 1995) are now commonly used in the literature. Moreover, there has recently been much focus on developing ways to construct special classes of networks to explicitly model evolution (see e.g. Huson et al., 2010; Nakhleh, 2011; Gusfield, 2014). Even so, there are still several aspects of phylogenetic network theory that remain to be more fully explored. One such aspect is how to transform one network into another one by using a collection of specified network operations (see e.g. Cardona et al., 2009; Yu et al., 2014; Huber et al., 2015, for some results in this direction) which we consider in this paper.

Transformations of phylogenetic trees have been studied for several years, and have applications to tree search algorithms and comparing trees (cf. e.g. Felsenstein, 2004, Chapters 4 and 30). Probably the best known and simplest way to transform one phylogenetic tree into another is to use a *nearest-neighbor interchange (NNI) operation* which we now recall. For a set X of three or more species or taxa, a *phylogenetic tree (on X)* is a tree in which every vertex has degree 1 or 3 with leaf set X. A pair of distinct trees differ by one NNI operation if one tree can be obtained from the other by swapping two of the four subtrees adjacent with an interior edge (Fig. 1(i)). Note that the NNI operation is *reversible*, i.e. there is a unique NNI operation (or *reverse* operation) that can be applied to get back to the original tree. A well-known result concerning the NNI operation states that given any pair of phylogenetic trees T, T' on X, it is possible to transform T into T' by some sequence of NNI operations (Robinson, 1971). This implies

* Corresponding author.
E-mail addresses: k.huber@uea.ac.uk (K.T. Huber),
v.moulton@uea.ac.uk (V. Moulton), taoyang.wu@gmail.com (T. Wu).
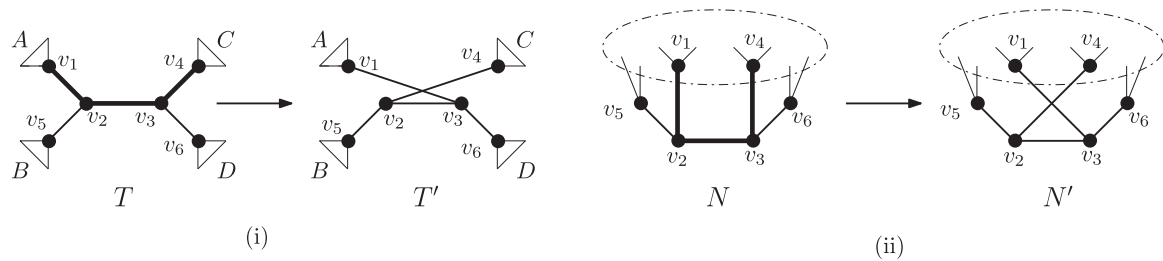
Fig. 1. (i) An NNI operation on a phylogenetic tree T. The tree T' is obtained from T by performing an operation on the path highlighted in bold that results in the subtrees labeled A and C being swapped. (ii) An NNI operation on a network N. The network N' is obtained from N by one NNI operation that is performed on the path highlighted in bold, just like for phylogenetic trees. Note that vertices v_1, v_4, v_5 and v_6 in N could all have degree 1 or 3, and that N contains neither edge $\{v_1, v_3\}$ nor edge $\{v_2, v_4\}$.

that NNI operations can be used to explore all possible phylogenetic trees on a set X, a useful fact that underpins several algorithms for reconstructing phylogenetic trees (see e.g. Felsenstein, 2004).

We now turn to the analogous problem for phylogenetic networks. More specifically, we shall consider the problem of transforming *unrooted phylogenetic binary networks* (on X), or *networks* for short, into one another. Such networks are connected graphs in which every vertex has degree 1 or 3 and whose leaf set is X (Gambette et al., 2012). Here we concentrate on binary networks since these generalize binary trees which are commonly employed within tree-inference methods (see e.g. Felsenstein, 2004, Chapter 4) and, moreover, multifurcations may be problematic for phylogenetic inference (Whelan and Money, 2010). In Fig. 2, we present an example of such a network R that was referred to as the "ring of life" in a study concerning the genome fusion origin of eukaryotes (Rivera and Lake, 2004, Fig. 1). Based on that network, whose construction employed whole genomes spanning the diversity of life, the authors concluded that the eukaryotic nuclear genome has resulted from a fusion of a relative of a protobacterium (P_γ) and a relative of an archeal ecocyte (E). Properties of networks have been studied in Gambette et al. (2012), and they can be generated by software such as T-REX (cf. Makarenkov, 2001) and Splitstree (cf. Huson and Bryant, 2006; Huson et al., 2005). In this paper, we show that it is possible to extend results concerning NNI operations on phylogenetic trees in a natural way to networks, which we expect could lead to applications to network search algorithms and comparison of networks.

Our starting point is to extend NNI operations to networks. This is based on the observation that the tree T' in Fig. 1(i) can be obtained from T by replacing the length 3 path v_1, v_2, v_3, v_4

highlighted in bold with the path v_1, v_3, v_2, v_4 whilst preserving all other edges. This definition immediately extends to give what we shall call an NNI operation on networks. In particular, we start again with a path v_1, v_2, v_3, v_4 in a network N on X for which neither $\{v_1, v_3\}$ nor $\{v_2, v_4\}$ is an edge, and obtain a new network N' on X by replacing this path with the path v_1, v_3, v_2, v_4 (Fig. 1(ii)). Note that N' has the same number of vertices as the original network, and that, just as with phylogenetic trees, the NNI operation is reversible.

In the first of our main results we show that, just as with phylogenetic trees, we can transform any network N on X to any other network N' on X with the same number of vertices as N by just using NNI operations (Theorem 3). We illustrate this in Fig. 2: here network R is transformed into the network R'' – which can be considered as an alternative ring of life hypothesis – by applying a sequence of two NNI operations. Even so, it is not possible to explore all alternative ways to potentially represent the ring of life using only NNI operations. To see this, note that we cannot transform the tree T in Fig. 4 to the network R in Fig. 2 using only NNI operations since, even though these networks are on the same leaf set, they have a different number of vertices and NNI operations must preserve this number.

In our second main result (Theorem 4) we show that we need only one additional operation and its reverse to be able to transform any network on X into any other network on X. We call these operations *triangle operations* or Δ *operations* and picture them in Fig. 3; they involve either inserting or removing a triangle (or, more technically, a length three cycle) from a network. We illustrate this result in Fig. 4. The tree T in this figure is one of the phylogenetic trees presented in Rivera and Lake (2004, Fig. 1). It is one of the five most probable trees computed using the method of
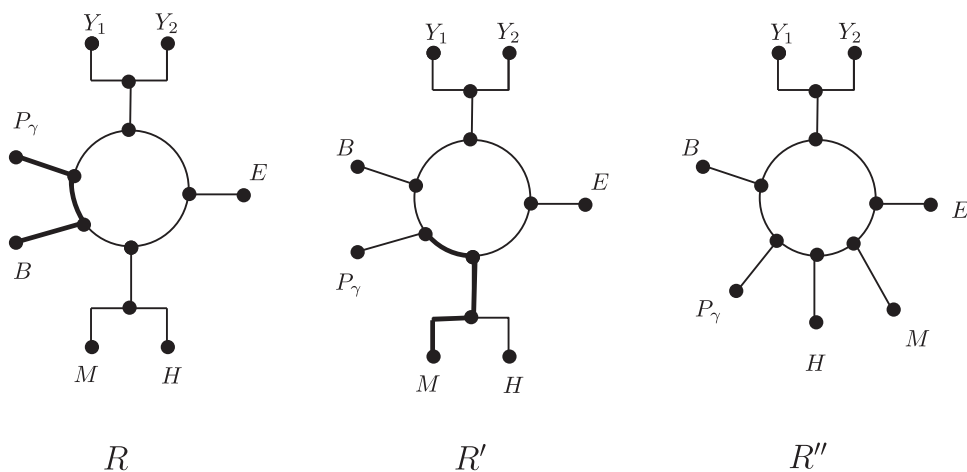
Fig. 2. An example of transforming phylogenetic networks using NNI operations. Network R is from Rivera and Lake (2004, Fig. 1), and is labeled by two yeasts (Y_1, *Schizosaccharomyces pombe* and Y_2, *Saccharomyces cerevisiae*), a γ-proteobacterium (P_γ, *Xylella fastidiosa*), a bacillus (B, *Staphylococcus aureus* MW2), a halobacterium (H, *Halobacterium* sp. NRC-1), an eocyte (E, *Sulfolobus tokodaii*) and a methanococcus (M, *Methanosarcina mazei Goe*1). The networks R' and R'' are obtained from R and R', respectively, by applying an NNI operations to the path highlighted in bold.