

On encodings of phylogenetic networks of bounded level

Philippe Gambette · Katharina T. Huber

Received: 29 June 2009 / Revised: 22 June 2011 / Published online: 14 July 2011
© Springer-Verlag 2011

Abstract Phylogenetic networks have now joined phylogenetic trees in the center of phylogenetics research. Like phylogenetic trees, such networks canonically induce collections of phylogenetic trees, clusters, and triplets, respectively. Thus it is not surprising that many network approaches aim to reconstruct a phylogenetic network from such collections. Related to the well-studied perfect phylogeny problem, the following question is of fundamental importance in this context: When does one of the above collections encode (i.e. uniquely describe) the network that induces it? For the large class of level-1 (phylogenetic) networks we characterize those level-1 networks for which an encoding in terms of one (or equivalently all) of the above collections exists. In addition, we show that three known distance measures for comparing phylogenetic networks are in fact metrics on the resulting subclass and give the diameter for two of them. Finally, we investigate the related concept of indistinguishability and also show that many properties enjoyed by level-1 networks are not satisfied by networks of higher level.

This work was supported by the French ANR projects ANR-08-EMER-011-01 (PhylARIANE) and Moonlight R09127AA. The authors would like to thank the organizers of the MIEP 2008 workshop where this work was initiated and also the two referees for their helpful comments and suggestions. Finally, they would also like to thank the London Mathematical Society for supporting part of this work.

P. Gambette
IML, CNRS, Université Marseille 2, Marseille, France
e-mail: philippe.gambette@gmail.com

K. T. Huber (✉)
School of Computing Sciences, University of East Anglia,
Norwich NR4 7TJ, UK
e-mail: Katharina.Huber@cmp.uea.ac.uk

Keywords Phylogenetic networks · Triplets · Clusters · Supernetwork · Level-1 network · Level- k network · Weak hierarchy · Consistency · Metric · Indistinguishable

Mathematics Subject Classification (2000) 92B10

1 Introduction

An improved understanding of the complex processes that drive evolution has lent support to the idea that reticulate evolutionary events, such as lateral gene transfer or hybridization, are more common than originally thought, rendering a phylogenetic tree (essentially a rooted leaf labelled graph-theoretical tree) too simplistic a model to fully understand the complex processes that drive evolution. Reflecting this, phylogenetic networks have now joined phylogenetic trees in the center of phylogenetics research. Influenced by the diversity of questions posed by evolutionary biologists that can be addressed with a phylogenetic network, various alternative definitions of these types of networks have been developed over the years (see e.g. [Huson et al. 2011](#) for a recent overview). These include split networks ([Bryant and Moulton 2004](#); [Bandelt et al. 1995](#); [Holland et al. 2004](#)) as well as ancestral recombination graphs ([Song and Hein 2005](#)), TOM networks ([Willson 2006](#)), level- k networks¹ with k a non-negative integer that in a some sense captures how complex the network structure is, networks for studying the evolution of polyploid organisms ([Moulton and Huber 2006](#)), tree-child and tree-sibling networks ([Cardona et al. 2008](#)), to name just a few.

Apart from split networks which aim to give an implicit model of evolution and are not the focus of this paper, all other phylogenetic networks mentioned above aim to provide an explicit model of evolution. Although slightly different in detail, they are all based on the concept of a leaf-labelled rooted connected directed acyclic graph (see the next section for a definition). For the convenience of the reader, we depict an example of a phylogenetic network in the form of a level-1 network in Fig. 1a. Concerning these types of phylogenetic networks, it should be noted that they are closely related to *galled trees* ([Wang et al. 2001](#); [Gusfield et al. 2003](#)) and that, in addition to constituting the first layer of the hierarchy of level- k networks, they also give rise to a large subclass of the class of tree-sibling networks ([Arenas et al. 2008](#)).

Due to the rich combinatorial structure of phylogenetic networks, different combinatorial objects have been used to reconstruct them from biological data. For a set X of taxa (e.g. species or organisms), these include *cluster systems* of X , that is, collections of non-empty subsets of X ([Bandelt and Dress 1989](#); [Huson and Rupp 2008](#)), *triplet systems* on X , that is, collections of phylogenetic trees with just three leaves which are generally called (*rooted*) *triplets* ([Jansson and Sung 2006](#); [To and Habib 2009](#)), and *tree systems*, that is, collections of phylogenetic trees which all have leaf set X ([Semple 2007](#)). The underlying rationale being that any phylogenetic network N

¹ Note that these networks were originally introduced in [Choy et al. \(2004\)](#), but the definition commonly used now is slightly different with the main difference being that every vertex of the network with indegree 2 must have outdegree 1 (see e.g. [Jansson et al. 2006](#)).