CAN HYBRIDISATION NETWORKS BE CONSTRUCTED FROM LOCAL INFORMATION?

BHALCHANDRA D. THATTE

Abstract. It is proved that a hybridisation networks cannot in general be constructed uniquely from its proper sub-networks, thus negatively answering a question of Semple.

1. Introduction

Evolutionary relationships between groups of species are typically represented by their *phylogenetic trees*. But it is well known that the evolutionary history of certain groups of species is not suitably represented by a single phylogenetic tree as a result of hybridisation, horizontal gene transfer, and recombination. Several authors have therefore considered *hybridisation networks* (also sometimes called *hybrid phylogenies* or *phylogenetic networks*) for the purpose of depicting reticulation events. Constructing a hybridisation network from a collection of phylogenetic trees that represent tree-like evolution of different parts of the genome is an NP-hard problem. Recently Semple [2] asked the question: can a hybridisation network be constructed from local information such as hybridisation networks on subsets of taxa? The purpose of this short note is to present an infinite family of networks that cannot be constructed from their collections of sub-networks. In particular, we show that there are non-isomorphic networks $\mathcal{P}$ and $\mathcal{Q}$ that have the same collection of hybridisation networks on subsets of their sets of taxa, and that there are infinitely many such pairs. The constructions are based on similar constructions for pedigrees of populations [4].

The following definition is based on [1].

A *hybridisation network* $\mathcal{H}(X)$ on a set X is a rooted acyclic diagraph with root ρ in which

(1) X is the set of vertices of out-degree zero;
(2) the out-degree $d^+(\rho)$ of ρ is at least 2;
(3) for all vertices v with out-degree $d^+(v) = 1$, the in-degree $d^-(v)$ is at least 2.

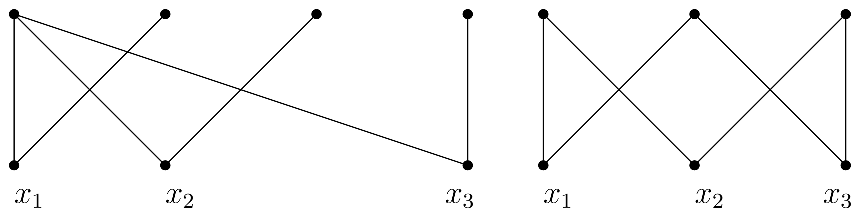
Key words and phrases. pedigrees, phylogenetics, hybridisation networks.
Supported by the Allan Wilson Centre for Molecular Ecology and Evolution.

The set X in the definition represents the set of taxa. Suppose $Y \subseteq X$. A sub-network $\mathcal{H}_Y(X)$ of $\mathcal{H}(X)$ is obtained by deleting vertices that have no descendents in Y, and then suppressing all vertices that have both in-degree and out-degree 1. Note that all vertices in $X \backslash Y$ are deleted (since they are treated as their own descendents), and when a vertex v is deleted, all arcs to and from v are deleted as well.

One formulation of Semple's question may now be stated as follows: can the hybridisation network $\mathcal{H}(X)$ be constructed uniquely (that is, up to isomorphism in the graph theoretic sense,) from the collection $\{\mathcal{H}_Y(X) | Y \subset X, |Y| = k\}$, where k is a fixed positive integer smaller than $|X| = n$?

A *general pedigree* is a directed acyclic graph in which every vertex has zero or two in-coming arcs. The vertices with no in-coming arcs are called the *founder vertices*. The *extant* vertices have no out-going arcs. In [4], I considered the problem of constructing the pedigree of a population from the collection of pedigrees on its proper sub-populations. The question for pedigrees was posed by Steel and Hein [3].

Figure 1 shows non-isomorphic pedigrees $\mathcal{T}$ and $\mathcal{U}$ that have identical collections of sub-pedigrees.



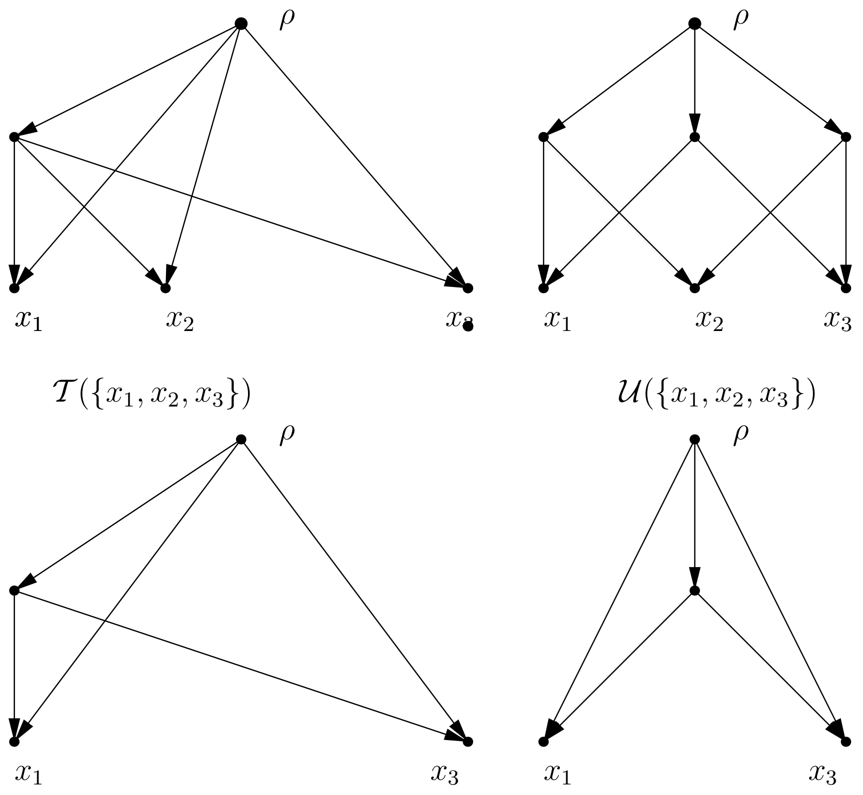
$\mathcal{T}(\{x_1, x_2, x_3\})$ $\mathcal{U}(\{x_1, x_2, x_3\})$

$$\mathcal{T}(\{x_i, x_j\}) \cong \mathcal{U}(\{x_i, x_j\}) \forall i, j$$
$$\mathcal{T} \ncong \mathcal{U}$$

FIGURE 1. Pedigrees that are not reconstructible from their sub-pedigrees.

The above example was used as the base case of an infinite family of non-isomorphic pedigrees having correspondingly isomorphic sub-pedigrees.

Here we modify the above examples as follows. In each of the above diagraphs, we add a root vertex ρ, and make all founder vertices as children of ρ. We then suppress all vertices that have both in-degree and out-degree 1. Figure 2 shows the two hybridisation networks and their proper sub-networks.



$$\mathcal{T}(\{x_i, x_j\}) \cong \mathcal{U}(\{x_i, x_j\}) \forall i, j$$

$$\mathcal{T} \ncong \mathcal{U}$$

FIGURE 2. Hybridisation networks that cannot be constructed from their proper sub-networks.

The infinite family of non-reconstructible pedigrees in [4] can be similarly modified to construct an infinitely family of hybridisation networks that cannot be constructed from their proper sub-networks, details of which are omitted in this note.

Acknowledgements. I would like to thank Charles Semple for useful conversations.

References

[1] Charles Semple. Hybridisation networks. Technical report, Department of Mathematics and Statistics, University of Canterbury, May 2006.

[2] Charles Semple. Hybridisation networks. Talk given at the Annual Phylogenetics Meeting, Doom, New Zealand, February 2007.

[3] M. Steel and J. Hein. Reconstructing pedigrees: a combinatorial perspective. *Journal of Theoretical Biology*, 2006.

[4] Bhalchandra D. Thatte. Combinatorics of pedigrees. *Preprint, arXiv:math.CO/0609264*, 16 pages, 2006.

Biomathematics Research Centre, Mathematics and Computer Science Building, University of Canterbury, Private Bag 4800, Christchurch, New Zealand

E-mail address, Bhalchandra D. Thatte: bdthatte@gmail.com