The Fair Proportion is a Shapley Value on phylogenetic networks too

Tomás M. Coronado, Gabriel Riera, and Francesc Rosselló

Dept. of Mathematics and Computer Science, University of the Balearic Islands, E-07122 Palma, Spain, and Balearic Islands Health Research Institute (IdISBa), E-07010 Palma, Spain {t.martinez,gabriel.riera,cesc.rossello}@uib.edu

Abstract. The Fair Proportion of a species in a phylogenetic tree is a very simple measure that has been used to assess its value relative to the overall phylogenetic diversity represented by the tree. It has recently been proved by Fuchs and Jin to be equal to the Shapley Value of the coallitional game that sends each subset of species to its rooted Phylogenetic Diversity in the tree. We prove in this paper that this result extends to the natural translations of the Fair Proportion and the rooted Phylogenetic Diversity to rooted phylogenetic networks. We also generalize to rooted phylogenetic networks the expression for the Shapley Value of the unrooted Phylogenetic Diversity game on a phylogenetic tree established by Haake, Kashiwada and Su.

1 Introduction

An important problem in ecology is to assess the genetic value of individual species, with the aim of ranking them for conservation prioritization purposes [2]. One of the simplest measures proposed in this connection is the Fair Proportion of a species in a phylogenetic tree, introduced by Redding and Mooers in [11]. This index apportions the overall diversity of a phylogenetic tree among its leaves by equally dividing the weight of each arc among its descendant leaves. Although this index is very easy to define, it is not obvious at first sight that it defines a sound and meaningful ranking of the species' genetic value. On the other hand, the Shapley Value of a species in a phylogenetic tree, introduced by Haake, Kashiwada and Su in [8], which is based on a well-known solution from cooperative game theory to the problem of dividing the global value of a game among its players, lies at the other end of the individual biodiversity measures spectrum, in the sense that it provides a meaningful distribution of the global diversity of a phylogenetic tree among its leaves at the cost of being defined through quite a complex formula, involving a sum of an exponential number of terms. But, in what Steel dubs as an "interesting and not immediately obvious" result [14, p. 141], Fuchs and Jin proved in [5] that

Fair Proportions and Shapley Values are exactly the same on phylogenetic trees, thus yielding an individual biodiversity index which is easy to define and compute and which ranks species in a very clear and meaningful way.

In this note we extend Fuchs and Jin's result from rooted phylogenetic trees to rooted phylogenetic networks [9], graphical models of evolutionary histories that allow the inclusion of reticulate processes like hybridizations, recombinations or lateral gene transfers. More specifically, we show that if we define the Fair Proportion of a leaf in a rooted phylogenetic network exactly as if we were a phylogenetic tree —we split the weight of each arc equally among all its descendant leaves, and then we add up the leaf's share of the weights of all its ancestor arcs— then it is equal to the subnet Shapley Value of the leaf in the network as defined by Wicke and Fischer in [15]. We also extend to rooted phylogenetic networks the simple expression for the unrooted phylogenetic Shapley Value established by Haake, Kashiwada and Su in [8], thus showing in particular that it can be computed efficiently also on rooted phylogenetic networks.

2   Preliminaries

Let Σ be a finite set of labels. A Σ-*rDAG* is a rooted directed acyclic graph with its *leaves* (its nodes of out-degree 0) bijectively labeled in Σ. We shall denote the sets of nodes and arcs of a Σ-rDAG N by $V(N)$ and $E(N)$, respectively, and we shall always identify its leaves with their corresponding labels. A *weighted* Σ-rDAG is a Σ-rDAG endowed with a mapping $\omega : E(N) \to [0, \infty)$ that assigns a weight $\omega(e) \geqslant 0$ to every arc e.

Given two nodes u, v in a Σ-rDAG N, we say that v is a *child* of u, and also that u is a *parent* of v, when $(u, v) \in E(N)$, and that v is a *descendant* of u, and also that u is an *ancestor* of v, when there exists a directed path from u to v in N. The *cluster* $C(e)$ of $e \in E(N)$ is the set of descendant leaves of its end, and we shall denote by $\kappa(e)$ the cardinal of $C(e)$. If $a \in C(e)$, we shall also say that e is an *ancestor arc* of a.

A *phylogenetic network* on Σ is a Σ-rDAG without *elementary* nodes (that is, without nodes of in-degree $\leqslant 1$ and out-degree 1). A node in a phylogenetic network is of *tree type* when its in-degree is 0 (the *root*) or 1, and a *reticulation* when its in-degree is at least 2. An arc is of *tree type* (respectively, of *reticulate type*) when its end is a tree node (resp., in a reticulation). Given a phylogenetic network N on Σ and a subset $X \subseteq \Sigma$, we shall denote by $N(X)$ the subgraph of N induced by the set of all the ancestors of the leaves in X: it is a X-rDAG, with the same root as N.

A *phylogenetic tree* is a phylogenetic network without reticulations. Let us emphasize, hence, that all our phylogenetic trees are rooted, unless otherwise explicitly stated. Given a weighted phylogenetic tree T on Σ, for every $a \in \Sigma$ and for every $X \subseteq \Sigma$:

- The *Fair Proportion* of a in T [11] is

$$FP_T(a) = \sum_{e:\, a \in C(e)} \frac{\omega(e)}{\kappa(e)}.$$

- The *rooted Phylogenetic Diversity* $rPD_T(X)$ of X in T [4] is the *total weight* of $T(X)$, that is, the sum of the weights of its arcs:

$$rPD_T(X) = \sum_{e:\, X \cap C(e) \neq \emptyset} \omega(e).$$

- The *unrooted Phylogenetic Diversity* $uPD_T(X)$ of X in T [4] is the total weight of the smallest unrooted subtree of T containing the leaves in X, or equivalently, the total weight of the subtree of $T(X)$ rooted at the lowest common ancestor $LCA_T(X)$ of X.

A *coalitional game* on a set Σ is simply a set function $W : \mathcal{P}(\Sigma) \to \mathbb{R}$. For every $a \in \Sigma$, the *Shapley Value* on a of a coalitional game W on Σ [12] is a weighted average of the marginal contribution of a to the value, under W, of each coalition $X \subseteq \Sigma$:

$$SV_a(W) = \sum_{a \in X \subseteq \Sigma} \frac{(|X| - 1)!(|\Sigma| - |X|)!}{|\Sigma|!} \big(W(X) - W(X \setminus \{a\})\big)$$

The *Shapley value* of the game W is then the vector $(SV_a(W))_{a \in \Sigma}$.

3 The Fair Proportion is a Shapley Value

Let N be a weighted phylogenetic network on Σ. We define the *Fair Proportion* of $a \in \Sigma$ in N, $FP_N(a)$, exactly as if N were a phylogenetic tree: we split the weight of each arc equally among all its descendant leaves, and then we add up a's share of the weights of all its ancestor arcs:

$$FP_N(a) = \sum_{e:\, a \in C(e)} \frac{\omega(e)}{\kappa(e)}.$$

In particular, if N is a phylogenetic tree, this Fair Proportion is equal to the one defined on phylogenetic trees by Redding and Mooers and recalled in the previous section. Our goal in this section is to show that, as it

already happens on phylogenetic trees [5], this Fair Proportion is the Shapley Value of a certain "phylogenetic diversity" coallitional game: namely, of $rPSD_N : \mathcal{P}(\Sigma) \to \mathbb{R}$ on Σ, where, for every $X \subseteq \Sigma$, $rPSD_N(X)$ is the *rooted Phylogenetic Subnet Diversity* of X in N in the sense of [15, Def. 6], which is defined as the total weight of $N(X)$:

$$rPSD_N(X) = \sum_{e:\, X \cap C(e) \neq \emptyset} \omega(e);$$

in particular, $rPSD_N(\emptyset) = 0$. Notice that if T is a phylogenetic tree, then $rPSD_T$ is equal to Faith's rooted Phylogenetic Diversity rPD_T recalled in the previous section.

For every $a \in \Sigma$, let its *rooted subnet Shapley Value* in N be the Shapley Value of $rPSD_N$ on a:

$$SV_N(a) = \sum_{a \in X \subseteq \Sigma} \frac{(|X| - 1)!(|\Sigma| - |X|)!}{|\Sigma|!} \big(rPSD_N(X) - rPSD_N(X \setminus \{a\})\big).$$
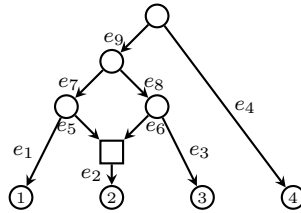

Fig. 1. The phylogenetic network used in Example 1.

Example 1. Consider the phylogenetic network N depicted in Figure 1 and let $w_i = \omega(e_i)$, for every $i = 1, \dots, 9$. Then:

$$\kappa(e_1) = \kappa(e_2) = \kappa(e_3) = \kappa(e_4) = \kappa(e_5) = \kappa(e_6) = 1$$
$$\kappa(e_7) = \kappa(e_8) = 2, \ \ \kappa(e_9) = 3$$

$$rPSD_N(1) = w_1 + w_7 + w_9$$
$$rPSD_N(2) = w_2 + w_5 + w_6 + w_7 + w_8 + w_9$$
$$rPSD_N(3) = w_3 + w_8 + w_9$$
$$rPSD_N(4) = w_4$$
$$rPSD_N(1, 4) = w_1 + w_4 + w_7 + w_9$$

4

$$rPSD_N(2,3) = w_2 + w_3 + w_5 + w_6 + w_7 + w_8 + w_9$$
$$rPSD_N(2,4) = w_2 + w_4 + w_5 + w_6 + w_7 + w_8 + w_9$$
$$rPSD_N(3,4) = w_3 + w_4 + w_8 + w_9$$
$$rPSD_N(1,2,3) = w_1 + w_2 + w_3 + w_5 + w_6 + w_7 + w_8 + w_9$$
$$rPSD_N(1,2,4) = w_1 + w_2 + w_4 + w_5 + w_6 + w_7 + w_8 + w_9$$
$$rPSD_N(1,3,4) = w_1 + w_3 + w_4 + w_7 + w_8 + w_9$$
$$rPSD_N(2,3,4) = w_2 + w_3 + w_4 + w_5 + w_6 + w_7 + w_8 + w_9$$

So, the Shapley Values of the leaves of N are:

$$SV_N(1) = \frac{1}{4}\big(rPSD_N(1) - rPSD_N(\emptyset)\big)$$

$$+\frac{1}{12}\big(rPSD_N(1,2) - rPSD_N(2) + rPSD_N(1,3) - rPSD_N(3)$$
$$+rPSD_N(1,4) - rPSD_N(4)\big)$$

$$+\frac{1}{12}\big(rPSD_N(1,2,3) - rPSD_N(2,3) + rPSD_N(1,2,4)$$
$$-rPSD_N(2,4) + rPSD_N(1,3,4) - rPSD_N(3,4)\big)$$

$$+\frac{1}{4}\big(rPSD_N(1,2,3,4) - rPSD_N(1,2,3)\big)$$

$$= \frac{1}{4}(w_1 + w_7 + w_9) + \frac{1}{12}(3w_1 + 2w_7 + w_9) + \frac{1}{12}(3w_1 + w_7) + \frac{1}{4}w_1$$

$$= w_1 + \frac{1}{2}w_7 + \frac{1}{3}w_9 = \frac{\omega(e_1)}{\kappa(e_1)} + \frac{\omega(e_7)}{\kappa(e_7)} + \frac{\omega(e_9)}{\kappa(e_9)} = FP_N(1)$$

$$SV_N(3) = w_3 + \frac{1}{2}w_8 + \frac{1}{3}w_9 = \frac{\omega(e_3)}{\kappa(e_3)} + \frac{\omega(e_8)}{\kappa(e_8)} + \frac{\omega(e_9)}{\kappa(e_9)} = FP_N(3)$$

(by symmetry)

$$SV_N(2) = \frac{1}{4}\big(rPSD_N(2) - rPSD_N(\emptyset)\big)$$

$$+\frac{1}{12}\big(rPSD_N(1,2) - rPSD_N(1) + rPSD_N(2,3)$$
$$-rPSD_N(3) + rPSD_N(2,4) - rPSD_N(4)\big)$$

$$+\frac{1}{12}\big(rPSD_N(1,2,3) - rPSD_N(1,3) + rPSD_N(1,2,4)$$
$$-rPSD_N(1,4) + rPSD_N(2,3,4) - rPSD_N(3,4)\big)$$

$$+\frac{1}{4}\big(rPSD_N(1,2,3,4) - rPSD_N(1,3,4)\big)$$

5

$$= \frac{1}{4}(w_2 + w_5 + w_6 + w_7 + w_8 + w_9)$$

$$+ \frac{1}{12}(3w_2 + 3w_5 + 3w_6 + 2w_7 + 2w_8 + w_9)$$

$$+ \frac{1}{12}(3w_2 + 3w_5 + 3w_6 + w_7 + w_8) + \frac{1}{4}(w_2 + w_5 + w_6)$$

$$= w_2 + w_5 + w_6 + \frac{1}{2}w_7 + \frac{1}{2}w_8 + \frac{1}{3}w_9$$

$$= \frac{\omega(e_2)}{\kappa(e_2)} + \frac{\omega(e_5)}{\kappa(e_5)} + \frac{\omega(e_6)}{\kappa(e_6)} + \frac{\omega(e_7)}{\kappa(e_7)} + \frac{\omega(e_8)}{\kappa(e_8)} + \frac{\omega(e_9)}{\kappa(e_9)} = FP_N(2)$$

$$SV_N(4) = \frac{1}{4}\big(rPSD_N(4) - rPSD_N(\emptyset)\big)$$

$$+ \frac{1}{12}\big(rPSD_N(1,4) - rPSD_N(1) + rPSD_N(2,4)$$

$$- rPSD_N(2) + rPSD_N(3,4) - rPSD_N(3)\big)$$

$$+ \frac{1}{12}\big(rPSD_N(1,2,4) - rPSD_N(1,2) + rPSD_N(1,3,4)$$

$$- rPSD_N(1,3) + rPSD_N(2,3,4) - rPSD_N(2,3)\big)$$

$$+ \frac{1}{4}\big(rPSD_N(1,2,3,4) - rPSD_N(1,2,3)\big)$$

$$= \frac{1}{4}w_4 + \frac{1}{12}\cdot 3w_4 + \frac{1}{12}\cdot 3w_4 + \frac{1}{4}w_4 = w_4 = \frac{\omega(e_4)}{\kappa(e_4)} = FP_N(4)$$

In the simple phylogenetic network considered in the previous example, the subnet Shapley Value of each leaf was equal to its Fair Proportion. Next theorem establishes that it is always the case.

Theorem 1. *For every weighted phylogenetic network N on Σ and for every $a \in \Sigma$,*

$$FP_N(a) = SV_N(a).$$

Proof. Set $|\Sigma| = n$. For every $X \subseteq \Sigma$ containing a,

$$rPSD_N(X) - rPSD_N(X \setminus \{a\}) = \sum_{e:\, X \cap C(e) \neq \emptyset} \omega(e) - \sum_{e:\, (X\setminus\{a\}) \cap C(e) \neq \emptyset} \omega(e)$$

$$= \sum_{e:\, X \cap C(e) = \{a\}} \omega(e)$$

Then,

$$SV_N(a) = \sum_{k=1}^{n} \frac{(k-1)!(n-k)!}{n!} \sum_{|X|=k,\, a \in X} \big(rPSD_N(X) - rPSD_N(X \setminus \{a\})\big)$$

6

where

$$\sum_{|X|=k,a\in X}(rPSD_N(X)-rPSD_N(X\setminus\{a\}))=\sum_{|X|=k,a\in X}\sum_{e:\,X\cap C(e)=\{a\}}\omega(e)$$

$$=\sum_{e:\,a\in C(e)}|\{Y\subseteq\Sigma\setminus C(e)\mid |Y|=k-1\}|\cdot\omega(e)$$

$$=\sum_{e:\,a\in C(e)}\binom{n-\kappa(e)}{k-1}\omega(e)$$

and therefore

$$SV_N(a)=\sum_{k=1}^{n}\left(\frac{(k-1)!(n-k)!}{n!}\sum_{e:\,a\in C(e)}\binom{n-\kappa(e)}{k-1}\omega(e)\right)$$

$$=\sum_{e:\,a\in C(e)}\left(\omega(e)\sum_{k=1}^{n}\frac{(k-1)!(n-k)!}{n!}\binom{n-\kappa(e)}{k-1}\right)$$

$$=\sum_{e:\,a\in C(e)}\left(\omega(e)\sum_{j=0}^{n-1}\frac{j!(n-j-1)!}{n!}\binom{n-\kappa(e)}{j}\right)=\sum_{e:\,a\in C(e)}\frac{\omega(e)}{\kappa(e)}$$

where the last equality is a consequence of Lemma 6.15 in [14], which establishes that, for every $1\leqslant m\leqslant n$,

$$\sum_{j=0}^{n-1}\frac{j!(n-j-1)!}{n!}\binom{n-m}{j}=\frac{1}{m}.$$

$\square$

Remark 1. A *multilabelled tree* (a *MUL-tree*, for short) on Σ is a rooted tree with its leaves labelled in Σ. The difference with usual phylogenetic trees is that the leaf labelling in a MUL-tree need not be bijective and, thus, more than one leaf may be assigned the same label. MUL-trees include *area cladograms* [6] and *gene trees* [7]. Given a MUL-tree T, if, for every label $a\in\Sigma$ assigned to more than one leaf, we remove all leaves labelled with a and the arcs ending in them and we add a new reticulation h_a, a new leaf labelled with a, new arcs from the parents of former leaves labelled with a to h_a and a new arc (h_a,a), we obtain a phylogenetic network uniquely determined by T, which we dub *associated* to T. For instance, the phylogenetic network in Figure 1 is the associated to the MUL-tree depicted in Figure 2.

This representation of MUL-trees as phylogenetic networks allows us to translate to their setting the concepts developed so far. So, let T be
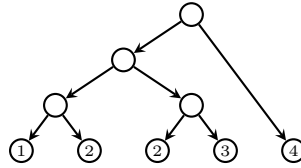
Fig. 2. A multilabelled tree.

a weighted MUL-tree. For every arc e in it, let $C(e)$ be the set of labels of its descendant leaves and $\kappa(e) = |C(e)|$ the number of different labels assigned to descendant leaves of e. For every label $a \in \Sigma$, we define its *Fair Proportion* in T as $FP_T(a) = \sum_{e:a\in C(e)} \omega(e)/\kappa(e)$: notice that now we split each $\omega(e)$ equally among the different labels of e's descendant leaves, without taking into account their multiplicities, that is, how many leaves have any given label. Then, if, for every $X \subseteq \Sigma$, we define its *MUL-Phylogenetic Diversity* in T as

$$mPSD_T(X) = \sum_{e:\, X\cap C(e)\neq\emptyset} \omega(e),$$

Theorem 1 applied to the phylogenetic network associated to T implies that FP_T is the Shapley Value of $mPSD_T$.

4 The unrooted subnet Shapley Value on a rooted phylogenetic network

Consider the following two further coallitional games asssociated to a phylogenetic network N on Σ: for every $X \subseteq \Sigma$,

– The *Cophenetic Value* $CV_N(X)$ (cf. [1,13]) is 0 if $X = \emptyset$ and the sum of the weights of the arcs that are ancestors of all leaves in X otherwise:

$$CV_N(\emptyset) = 0 \text{ and } CV_N(X) = \sum_{e:\, X\subseteq C(e)} \omega(e) \text{ if } X \neq \emptyset.$$

– The *unrooted Phylogenetic Subnet Diversity* $uPSD_N(X)$ is the difference

$$uPSD_N(X) = rPSD_N(X) - CV_N(X) = \sum_{\substack{e:\, X\cap C(e)\neq\emptyset \\ X\nsubseteq C(e)}} \omega(e).$$

8

So, if T is a phylogenetic tree, then CV_T is equal to the usual cophenetic value of a set X of leaves, that is, the total weight of the path going from the root of T to $LCA_T(X)$, and $uPSD_T$ is equal to Faith's unrooted Phylogenetic Diversity uPD_T as recalled in §2.

For every $a \in \Sigma$, let its *unrooted subnet Shapley Value* in N be the Shapley Value of $uPSD_N$ on a,

$$uSV_N(a) = \sum_{a \in X \subseteq \Sigma} \frac{(|X| - 1)!(|\Sigma| - |X|)!}{|\Sigma|!} \left(uPSD_N(X) - uPSD_N(X \backslash \{a\}) \right),$$

and its *cophenetic Shapley Value* in N, the Shapley Value of CV_N on a,

$$cSV_N(a) = \sum_{a \in X \subseteq \Sigma} \frac{(|X| - 1)!(|\Sigma| - |X|)!}{|\Sigma|!} \left(CV_N(X) - CV_N(X \backslash \{a\}) \right),$$

By the additivity of Shapley Values, $rPSD_N = uPSD_N + CV_N$ implies that

$$SV_N = uSV_N + cSV_N.$$

Our goal is to obtain an expression for uSV_N that generalizes to rooted phylogenetic networks the expression for uPD_N on phylogenetic trees established in [8]. We do it using Theorem 1 and the following expression for cSV_N.

Lemma 1. *For every weighted phylogenetic network N on Σ and for every $a \in \Sigma$,*

$$cSV_N(a) = \frac{1}{n} rPSD_N(\Sigma) - \sum_{e : a \notin C(e)} \frac{\omega(e)}{n - \kappa(e)}.$$

Proof. Set $|\Sigma| = n$. To simplify the notations, we shall omit the subscripts N in CV_N and cSV_N. For every $\{a\} \subsetneq X \subseteq \Sigma$

$$CV(X) - CV(X \backslash \{a\}) = \sum_{e : X \subseteq C(e)} \omega(e) - \sum_{e : (X \backslash \{a\}) \subseteq C(e)} \omega(e) = - \sum_{\substack{e : (X \backslash \{a\}) \subseteq C(e) \\ a \notin C(e)}} \omega(e)$$

while

$$CV(\{a\}) - CV(\emptyset) = \sum_{e : a \in C(e)} \omega(e) = rPSD_N(\Sigma) - \sum_{e : a \notin C(e)} \omega(e)$$

9

Then,

$$cSV(a) = \sum_{k=1}^{n} \frac{(k-1)!(n-k)!}{n!} \sum_{|X|=k, a \in X} \big(CV(X) - CV(X \setminus \{a\})\big)$$

$$= \frac{1}{n}(CV(\{a\}) - CV(\emptyset))$$

$$+ \sum_{k=2}^{n} \frac{(k-1)!(n-k)!}{n!} \sum_{|X|=k, a \in X} \big(CV(X) - CV(X \setminus \{a\})\big)$$

where, for every $k \geqslant 2$,

$$\sum_{|X|=k, a \in X} (CV(X) - CV(X \setminus \{a\})) = - \sum_{|X|=k, a \in X} \sum_{\substack{e:\, (X \setminus \{a\}) \subseteq C(e) \\ a \notin C(e)}} \omega(e)$$

$$= - \sum_{e:\, a \notin C(e)} |\{Y \subseteq C(e) \mid |Y| = k-1\}| \cdot \omega(e) = - \sum_{e:\, a \notin C(e)} \binom{\kappa(e)}{k-1} \omega(e)$$

and therefore

$$cSV(a) = \frac{1}{n}\left(rPSD_N(\Sigma) - \sum_{e:\, a \notin C(e)} \omega(e) \right)$$

$$- \sum_{k=2}^{n} \left(\frac{(k-1)!(n-k)!}{n!} \sum_{e:\, a \notin C(e)} \binom{\kappa(e)}{k-1} \omega(e) \right)$$

$$= \frac{1}{n} rPSD_N(\Sigma) - \sum_{k=1}^{n} \left(\frac{(k-1)!(n-k)!}{n!} \sum_{e:\, a \notin C(e)} \binom{\kappa(e)}{k-1} \omega(e) \right)$$

$$= \frac{1}{n} rPSD_N(\Sigma) - \sum_{e:\, a \notin C(e)} \left(\omega(e) \sum_{k=1}^{n} \frac{(k-1)!(n-k)!}{n!} \binom{\kappa(e)}{k-1} \right)$$

$$= \frac{1}{n} rPSD_N(\Sigma) - \sum_{e:\, a \notin C(e)} \frac{\omega(e)}{n - \kappa(e)}$$

using again Lemma 6.15 in [14]. $\qquad\square$

Replacing the expressions for cSV_N and $rCSV_N$ given in Theorem 1 and the last lemma, respectively, in $uCSV_N = rCSV_N - cSV_N$, we obtain the following result:

Theorem 2. *For every weighted phylogenetic network N on Σ and for every $a \in \Sigma$,*

$$uSV_N(a) = \frac{1}{n} \sum_{e:\, a \in C(e)} \frac{n - \kappa(e)}{\kappa(e)} \cdot \omega(e) + \frac{1}{n} \sum_{e:\, a \notin C(e)} \frac{\kappa(e)}{n - \kappa(e)} \cdot \omega(e).$$

10

It is not difficult to check that this expression agrees with the one given in [8, Thm. 4] when N is a rooted phylogenetic network.

5 Conclusions

In this note we have generalized to rooted phylogenetic networks two results on Shapley Values for phylogenetic trees: the equality of the rooted phylogenetic Shapley Value to the Fair Proportion and the simple expression of the unrooted phylogenetic Shapley Value in terms of the weights and the number of descendant leaves of arcs.

We would like to call the reader's attention on the fact that Theorem 1 is easily generalized to coallitional games $W : \mathcal{P}(\Sigma) \to \mathbb{R}$ for which there exist a set E and two mappings $C : E \to \mathcal{P}(\Sigma)$ and $\omega : E \to \mathbb{R}$ such that

$$W(X) = \sum_{e:\, X \cap C(e) \neq \emptyset} \omega(e).$$

For such a game W, the proof of Theorem 1 *mutatis mutandis* shows that its Shapley value on $a \in \Sigma$ is simply

$$SV_a(W) = \sum_{e:\, a \in C(e)} \frac{\omega(e)}{|C(e)|}.$$

For instance, a Shapley Value of this type can be used to assess the importance of a question in an exam, one of the main goals of Item Response Theory [3], as follows. Let Σ be the set of questions in an exam and let E be the set of students taking this exam. We assume all questions in the exam to be worth the same score. For every student e, let $C(e)$ be the set of questions correctly answered in her exam and set $\omega(e) = 1/|E|$. For every set of questions X, let $W(X) = \sum_{e:\, X \cap C(e) \neq \emptyset} \omega(e)$, which is equal to the fraction of students that answered correctly some question in X. Then, as we have just seen, the Shapley Value of this game on a given question a is

$$SV_a(W) = \frac{1}{|E|} \sum_{e:\, a \in C(e)} \frac{1}{|C(e)|}.$$

This Shapley Value measures the contribution of question a to the global success in the exam; it increases with the number of students who answered the question correctly, but decreases with the grades they obtained.

If different questions may have different scores, then it would be sensible to take as $\omega(e)$ the total score of the exam divided by $|E|$, in which

case $W(X)$, for a set of questions X, would be the average grade of the students who answered correctly some question in X. For another, recent use of the Shapley Value in the classification of items in an exam, see [10].

References

1. G. Cardona, A. Mir, L. Rotger, F. Rosselló, D. Sánchez. "Cophenetic metrics for phylogenetic trees, after Sokal and Rohlf." BMC Bioinformatics (2013) 14:3
2. J. Diniz. "Phylogenetic diversity and conservation priorities under distinct models of phenotypic evolution." Conservation Biology 18 (2004), 698–704.
3. S. Embretson, S. Reise. *Item response theory*. Psychology Press (2013).
4. D. Faith. "Conservation evaluation and phylogenetic diversity." Biological Conservation 61 (1992), 1–10.
5. M. Fuchs, E. Y. Jin. "Equality of Shapley value and fair proportion index in phylogenetic trees." Journal of Mathematical Biology 71 (2015), 1133–1147.
6. G. Ganapathy, B. Goodson, R. Jansen, H. Le, V. Ramachandran, T. Warnow. "Pattern identification in biogeography." IEEE/ACM Transactions on Computational Biology and Bioinformatics 3 (2006), 334–346.
7. W. Gregg, S. Ather, M. Hahn. "Gene-tree reconciliation with mul-trees to resolve polyploidy events." Systematic Biology, 66 (2017), 1007–1018.
8. C.-J. Haake, A. Kashiwada, F. E. Su. "The Shapley value of phylogenetic trees." Journal of Mathematical Biology 56 (2008), 479–497.
9. D. Huson, R. Rupp, C. Scornavacca. *Phylogenetic Networks: Concepts, Algorithms and Applications*. Cambridge University Press (2010).
10. M. Lutsenko, N. Shadrinceva. "Shapley values of test items" (in Russian). Bulletin of St. Petersburg University, Series 13 (2017), 300–312.
11. D. Redding, A. Mooers. "Incorporating evolutionary measures into conservation prioritization." Conservation Biology 20 (2006), 1670–1678.
12. L. Shapley. "A value for n-person games." *Contributions to the Theory of Games Vol. II* (H. Kuhn and A. Tucker, eds.), Annals of Mathematical Studies 28 (Princeton University Press, 1953), 307–317.
13. R. Sokal, F. Rohlf. "The Comparison of Dendrograms by Objective Methods." Taxon 11 (1962), 33–40.
14. M. Steel. *Phylogeny: Discrete and random processes in evolution*. SIAM (2016).
15. K. Wicke, M. Fischer. "Phylogenetic diversity and biodiversity indices on phylogenetic networks.". arXiv preprint arXiv:1706.05279 (2017).