Journal of mathematical biology manuscript No.
(will be inserted by the editor)

Biodiversity, Shapely Value and Phylogenetic Trees: Some Remarks

Hubert Stahn

Received: date / Accepted: date

Abstract This paper explores the main differences between the Shapley Values of a set of taxa introduced by Haake et al. (2007) and Fuchs and Jin (2015), the latter having been found identical to the Fair Proportion Index (Redding and Mooers 2006). In line with Shapley (1953), we identify the cooperative game basis for each of these two classes of phylogenetic games and use them (i) to construct simple formulas for these two Shapley values and (ii) to compare these different approaches. Using the set of weights of a phylogenetic tree as a parameter space, we then discuss the conditions under which these two values coincide and, if they are not the same, revisit Hartman's (2013) convergence result. Finally, we compare the species ranking induced by these two values. Considering the Kendal and the Spearman rank correlation coefficient, simulations show that these rankings are strongly correlated.

Keywords Biodiversity · Phylogenetic trees · Shapley Value · Fair Proportion index

Mathematics Subject Classification (2010) 92B99 · 91A12 · 05C05

1 Introduction

In recent years, biodiversity conservation programs have been placing increasing emphasis on a notion from cooperative game theory: the Shapley value[1] .

Finacial support of the Labex AMSE (ANR-11-IDEX-0001-02) and the ANR GREEN-Econ (ANR-16-CE03-0005) are gratefully acknowledged.

Hubert Stahn
Aix-Marseille University (Aix-Marseille School of Economics), CNRS, EHESS & ECM
AMSE, Chateau Lafarge, 50 Chemin du chateau Lafarge, F-13290 Les Milles, France
Tel :+33 685 483 636
E-mail: hubert.stahn@univ-amu.fr

[1] For recent applications see for instance Cadotte et al. (2010), Martyn et al. (2012), Redding et al. (2014), Volkman et al. (2014) , or Jensen et al. (2016).

Lloyd Shapley in 1953 asked a very simple question: assuming that a group of individuals shares a common goal whose outcome is measurable, how can we evaluate the contribution of each individual to this specific objective? Where conservation programs are concerned, this question becomes: how can we evaluate the contribution of a species to overall biodiversity and how can we organize a conservation policy under a limited budget (see Weitzman 1998) by targeting specific species?

To answer this question from a game theoretical point of view, Shapley (1953) looks at the outcome that each sub-group can achieve by itself, i.e. the characteristic function, and deduces, under several axioms, a unique imputation rule which specifies the individual contributions. These axioms depict a set of acceptable restrictions on the imputation rule. He assumes that (i) the outcome of the largest group is attributed to the members (efficiency), (ii) two individuals who contribute in the same way to each particular subgroup receive the same reward (symmetry) and (iii) an individual with a zero contribution to each subgroup receives nothing (null-player). He also introduces a more technical, but nevertheless important, assumption (iv) which states that the imputation rule is additive with respect to the characteristic functions. This means that the rewards obtained in a game resulting from the sum of two characteristic functions is the sum of the rewards linked to each of these characteristic functions.

If we now move back to biodiversity conservation problems, this Shapley metric provides an estimate of the contribution of one particular species to the overall phylogenetic diversity of a set of taxa. Its application simply requires the construction of the characteristic function of this game, i.e. the biodiversity index that a subset of taxa achieves by itself. The answer to this question can be found in Faith's (1992) seminal contribution. Each set of taxa has a polygenetic diversity index whose definition can be applied recursively in order to obtain the biodiversity index of each subgroup. The Shapley axiomatic does the rest. It provides a measure of the contribution of each species to the global biodiversity index.

This biodiversity measure, which gives rise to several applications, was also studied from a theoretical point of view, especially by Haake et al. (2007), Hartman (2013) and Fuchs and Jin (2015). However, comparing the early contribution of Haake et al. (2007) with the more recent work of Fuchs and Jin (2015) yields the impression that they are working with two different Shapley values. Of course, the first paper considers unrooted trees while the second introduces rooted trees. But this difference is not crucial, since any rooted binary tree can always be transformed into an unrooted tree by including an additional leaf, the root, with a zero weight. Actually, if we turn back to Shapley's work, the uniqueness of the imputation rule suggests that these two papers do not use the same characteristic function. Notably, they do not use the same definition of the phylogenetic diversity index of a subgroup. Haake et al. (2007) consider the subtree spanned by the sub-group of taxa, while Fuchs and Jin (2015) include the origin in this sub-group. This clearly raises several

questions: (i) what is the real difference between these two Shapley values and under which restrictions do they coincide? (ii) if not the same, when used in a prioritization problem do they predict drastically different rankings or are they reasonably correlated? (iii) since Fuchs and Jin (2015) show that their Shapley value is equivalent to the Fair Proportion index (Redding and Mooers 2006), is the Shapley value introduced by Haake et al. (2007) close to the Fair Proportion index, as suggested by Hartman (2013), especially concerning species ranking?

This paper attempts to answer these questions. To do so, we perform a preliminary step in which we explicitly provide simple formulas for these two Shapley values, as in Haake et al. (2007) but unlike Fuchs and Jin (2015), who simply show the equivalence to the Fair Proportion index. While this step might appear redundant, our approach, based on Shapley's original proof, provides a unified method which simplifies comparisons. Actually, there are three steps to his argument. He first constructs a linearly independent family of cooperative games which spans the set of all games. He then shows that each member of this family induces, under axioms (i) to (iii), a unique distribution of the individual contributions. Finally, he extends his result to all games by observing, under axiom (iv), that the imputation rule is a linear operator. Our approach uses the same rationale. We identify the basis of the subspace for the phylogenetic game, compute the contribution to biodiversity of each element of this basis under the Shapley assumptions and extend, by linearity, this measure to the set of all phylogenetic games. This method has several advantages.

It first provides a simple way to compute these two Shapley values by pointing out the differences between the two approaches. Haake et al.(2007) use unrooted trees and therefore do not, by construction, include the root in the computation of the phylogenetic diversity index of a coalition, while Fuchs and Jin (2015), who use rooted trees, include it. This basic distinction leads to two different subsets of potential characteristic functions, hence to different basis and distinct Shapley values. This computation exercise also leads to finding equality between the Shapley value and the Fair Proportion index.

Our method also provides a natural parametrization of the two sets of potential characteristic functions. In fact, the two basis that we identify are mainly related to the split structure of the tree, while the potential characteristic functions are obtained by a linear combination of these vectors using the weights for each edge of a phylogenetic tree under consideration. This means, under the linearity assumption, that the two Shapley values are linear in weight, which incidentally provides a natural way to compute the difference and to compare the two values. We notably identify the linear subspace of weights for which the two Shapley values are the same. When they are different, we revisit the result of Hartman (2013), who suggests that the contribution of each edge to individual biodiversity converges, as the number of taxa increases, to the contribution of each edge to the Fair Proportion index.

Finally, direct computation of these two Shapley values also enables us to examine the species ranking induced by these values, especially their degree of correlation when the two metrics are known to be different. This problem is more empirical and requires simulations. To gain some intuitions, we first consider a symmetric tree, randomly choose the weights of the edges in the subset in which the two Shapley values are known to be different and compute, in each case, the rank correlation coefficients between the species ranking induced by the two Shapley values. We especially consider the Tau statistic constructed by Kendall (1938) and the Rho statistic introduced by Spearman (1904). The distributions of these correlation coefficients suggest that the rankings induced by the two Shapley values are strongly correlated, even though the number of species under consideration is low. These simulation results are then extended to a random choice of the split structure of the tree.

Our argument will be organized as follows. Section 2 reviews the main notations concerning phylogenetic trees and recalls Shapley's result. Sections 3 and 4 are devoted to an explicit derivation of the Shapley value for phylogenetic trees when, respectively, the phylogenetic index is or is not independent of the origin. Section 5 compares the two Shapley values. Section 6 explores the impact of the species ranking provided by the two metrics. Section 7 concludes.

2 Notations and preliminary results

In this section, we briefly review the notion of phylogenetic trees and recall the main results related to the notion of Shapley value.

2.1 Phylogenetic trees

A phylogenetic tree of n taxa is a *weighted binary tree* whose set $I = \{1, \dots n\}$ of leaves is identified with the n taxa. Its graph $(V, E, w(\cdot))$ is composed of (i) a set V of vertices, $v \in V$, including the n taxa, (ii) a set $E \subset V \times V$ of edges which describes the adjacent vertices $e = (v, v')$ with $v \neq v'$, and (iii) a map, $w : E \to \mathbb{R}$, which assigns a weight to each edge $e \in E$. Being a tree, this graph is connected and free of cycles so that there exists a unique path $\{v \to v'\} \subset E$ between two vertices. The number of vertices adjacent to vertex v is called the the degree of v, $\deg(v)$. The degree of each taxon, i.e. of an external vertex, is 1. If the degree of all the other (internal) vertices is 3, the tree is said to be unrooted. If there exists one vertex of degree 2, called the origin, 0, the tree is said to be rooted.

Each *unrooted* binary tree is usually non-oriented and contains $(2n-3)$ edges. This tree can be split into two subtrees by removing one edge, e. This induces a partition of the set of leaves/taxa, $s_e = \left\{I_e \mid \bar{I}_e\right\}$. Reciprocally, an unrooted binary tree can be characterized by its pairwise compatible split structure $S_e = \{s_e\}_{e \in E}$. In contrast, a *rooted* tree is usually oriented, since there exists an origin, and contains $(2n-2)$ edges. The orientation makes it

possible to associate with each edge, e, the set, I_e, of leaves/taxa which follow this edge. This set is called the clade of edge e.

Each rooted phylogenetic tree can be transformed into an unrooted tree by adding an additional edge, e_0, with zero weight, i.e. $w(e_0) = 0$, which links an additional leaf, the origin 0, to the rest of the tree. The introduction of these *unrooted rooted trees* gives us the opportunity to compare the two Shapley values. In this case, the set of taxa is no longer identified with the leaves but rather with the leaves minus 0, and the set of relevant edges E is obtained by excluding e_0 from the set of edges, so that $|E| = 2n-2$. By removing a relevant edge, we now obtain a split $s_e = \left\{I_e \mid \bar{I}_e \cup \{0\}\right\}$, wherein, by convention, leaf 0 is in the second subset. It follows that I_e can be identified with the clade associated with edge e.

2.2 Cooperative games and Shapley value

Let us now consider sub-groups of the set, $I = \{1, \dots n\}$, of taxa. Each sub-group or coalition, C, belongs to 2^I, the set of 2^n subsets of I. By convention, $\emptyset$ and I are respectively called the *empty* and the *grand* coalition. Cooperative games assign a score to each coalition, i.e. an additive measure, which evaluates the benefit that this group can achieve by itself. This *characteristic function* $v : 2^I \rightarrow \mathbb{R}$ associates each coalition C with a real $v(C)$. By convention, the image of the empty set is zero, i.e. $v(\emptyset) = 0$. Given this characteristic function, these games aim to propose an *imputation* $\phi_v : I \rightarrow \mathbb{R}$ measuring the contribution, $\phi(i)$, of each individual to the score, $v(I)$, of the grand coalition. Since the set of potential imputations is large, Shapley (1953) adds additional restrictions on this mapping. He first requires that this imputation distributes the wealth obtained by the grand coalition, implying that nothing is lost. This efficiency axiom says that:

Axiom 1 (Efficiency) $\sum_{i \in I} \phi_v(i) = v(I)$.

In addition to this axiom, Shapley (1953) requires that two individuals who contribute in the same way to every coalition obtain the same imputation. This symmetry axiom is given by:

Axiom 2 (Symmetry) $\forall C \subset I \backslash \{i, j\}$ *if* $v(C \cup \{i\}) = v(C \cup \{j\})$ *then* $\phi_v(i) = \phi_v(j)$.

He also assumes that an individual contributing to no coalition receives no payment. In fact, he says:

Axiom 3 (Null Player) *If* $\forall C \subset I \backslash \{i\}$, $v(C \cup \{i\}) = v(C)$ *then* $\phi_v(i) = 0$.

Finally, Shapley (1953) adds a somewhat more technical assumption. He requires that the imputation obtained from the sum of two characteristic functions should simply be the sum of the two initial imputations:

Axiom 4 (Additivity) $\forall v_1, v_2$, $\forall i \in I$, $\phi_{(v_1+v_2)}(i) = \phi_{v_1}(i) + \phi_{v_2}(i)$.

Under these four axioms, he shows that:

Proposition (Shapley (1953)) *There exists a unique imputation rule, $Sh_v(i)$, called the Shapley value which satisfies these four axioms.*

His proof is essentially based on the idea that the space of characteristic functions can be spanned by a family of linearly independent games (winning coalition) with the property that each of these characteristic functions induces, under Axioms 1-3, a unique imputation. Since these imputations also satisfy a scalar multiplication property, the additivity Axiom extends this preliminary result, in a unique way, from this independent family of games to the set of all characteristic functions. Finally, he supplements his main result by providing one of the most popular ways to compute this value[2] using an average value of the marginal contribution of each individual to the different coalitions, i.e.:

$$\forall i \in I, \quad Sh_v(i) = \sum_{C \subset I, i \in C} \frac{(|C|-1)!(n-|C|)!}{n!} \left(v\left(C\right) - v\left(C \backslash \{i\}\right)\right) \tag{1}$$

The probability distribution used in this computation is based on the idea that individuals are randomly ordered. Given his rank, player i forms a coalition with the previously ranked players and receives a gain corresponding to his marginal contribution to the coalition formed. The probability that when he enters he will find coalition $C\backslash\{i\}$ there already is $\frac{(|C|-1)!(n-|C|)!}{n!}$.

3 Shapley value for unrooted trees (Haake et al. 2007)

In this case, the phylogenetic diversity index, $PD^u(C)$, of a coalition $C \in 2^I$ is given by the sum of the weights of the edges contained in the subtree spanned by coalition C. Since there exists, for any tree, a unique path, $\{i \to j\} \subset E$ which joins each pair $\{i, j\}$ of taxa, the edges of this subtree are simply given by $E_c^u = \cup_{\{i,j\} \subset C} \{i \to j\}$, the union of all paths joining two taxa in coalition C, and the phylogenetic diversity index of a coalition C becomes:

$$\forall C \in 2^I, \quad PD^u(C) = \textstyle\sum_{e \in E_c^u} w(e) \tag{2}$$

Let us now identify the function $PD^u(C)$ as the characteristic function of a cooperative game and denote by $\mathcal{P}^u$ the set of all characteristic functions obtained by changing the weights while keeping the tree unchanged. This set, $\mathcal{P}^u$, is obviously a subset of the set of all characteristic functions of cooperative games explored by Shapley (1953). We even claim that $\mathcal{P}^u$ is a linear subspace of dimension $(2n-3)$ corresponding to the number of edges of an unrooted binary tree with n leaves. The intuition behind this result is quite simple. Since the tree structure remains unchanged, we can introduce a family of $(2n-3)$ new phylogenetic trees. Each of these trees, say of type e, is associated with a given edge $e \in E$ and has the property that the weight of edge e is $w(e) = 1$

[2] For other expressions of the Shapley value see for instance Kleinberg and Weiss (1985) or Rothblum (1988).

while for all $e' \neq e$, $w(e') = 0$. With this property, the phylogenetic diversity index, $b_e^u(C)$, of a coalition C associated with tree e is either 0 or 1 and, from Eq.(2), it is immediate that this quantity is 1 if and only if there exist two taxa in C which are connected by a path meeting edge e. More formally, we can say that:

$$\forall C \in 2^I, \quad b_e^u(C) = \begin{cases} 1 \text{ if } \exists \{i,j\} \subset C,\ e \in \{i \rightarrow j\} \\ 0 \text{ otherwise} \end{cases} \tag{3}$$

Moreover, we know that the removal of edge, e, from the tree caused the tree to split into two subtrees isolating two subsets of taxa. This partition $s_e = \left\{I_e \mid \bar{I}_e\right\}$ produces a new interpretation of Eq.(3) mainly based on the set of taxa. In fact, if there exists a path between two elements of C which meets e, this also means that coalition C meets both subsets of the partition s_e, *i.e.* Eq.(3) becomes:

$$\forall C \in 2^I, \quad b_e^u(C) = \begin{cases} 1 \text{ if } I_e \cap C \neq \emptyset \text{ and } \bar{I}_e \cap C \neq \emptyset \\ 0 \text{ otherwise} \end{cases} \tag{4}$$

It is also immediate that the $(2n-3)$ family, $\{b_e^u\}_{e \in E} \subset \mathcal{P}^u$, of characteristic functions has the property that:

$$\forall C \in 2^I, \quad PD^u(C) = \textstyle\sum_{e \in E} w(e) b_e^u(C) \tag{5}$$

It therefore remains to be shown that this family is linearly independent in order to say:

Lemma 1 *$\{b_e^u\}_{e \in E} \subset \mathcal{P}^u$ is a basis of dimension $(2n-3)$ of the linear set $\mathcal{P}^u$.*

Proof Assume that the family $\{b_e^u(\cdot)\}$ is not independent. In this case, there exists, $(\alpha_e)_{e \in E} \neq 0$, a vector of scalars in $\mathbb{R}^{2n-3}$ which satisfies

$$\forall C, \quad \textstyle\sum_{e \in E} \alpha_e b_e^u(C) = 0 \tag{6}$$

Now select an edge e whose induced slit is $\left\{I_e \mid \bar{I}_e\right\}$ and construct the partition $\{E_1 \mid e \mid E_2\}$ of the set of edges which isolates e and the sets of edges, E_1 and E_2, associated with the subtrees, respectively, spanned by I_e and $\bar{I}_e$. In this case, we observe that: (i) for each $e' \in E_1$, the associated split $\left\{I_{e'} \mid \bar{I}_{e'}\right\}$ has the property that $I_{e'} \subsetneqq I_e$ so that, by Eq.(4), $b_{e'}^u(I_e) = 1$, (ii) for each $e' \in E_2$ the associated split $\left\{\bar{I}_{e'} \mid \bar{I}_{e'}\right\}$ verifies $I_e \subsetneqq I_{e'}$ so that $b_{e'}^u(I_e) = 0$ and finally (iii) $b_e^u(I_e) = 0$ since I_e cannot meet $\bar{I}_e$. This implies, if the family $\{b_e^u(\cdot)\}$ is not independent, that:

$$\textstyle\sum_{e' \in E} \alpha_{e'} b_{e'}^u(I_e) = \sum_{e' \in E_1} \alpha_{e'} = 0 \tag{7}$$

By a symmetric argument, we can also say that:

$$\textstyle\sum_{e' \in E} \alpha_{e'} b_e^u(\bar{I}_e) = \sum_{e' \in E_2} \alpha_{e'} = 0 \tag{8}$$

Moreover, if we consider the grand coalition, Eq.(4) says that $\forall e \in E$, $b_e^u(I) = 1$. It follows:

$$\sum_{e\in E}\alpha_e b_e^u(I) = \sum_{e\in E}\alpha_e = 0 \tag{9}$$

By combining Eqs (7), (8) and (9), we obtain that $\alpha_e = 0$ and by repeating the argument for each $e \in E$, it is impossible that $(\alpha_e)_{e\in E} \neq 0$. □

The next step in our analysis to construct the Shapley value, $Sh_\tau^u(i)$, of a taxon $i \in I$, using the decomposition of $PD^u(\cdot)$ provided by Eq.(5). This requires, in a preliminary step, the computation of the Shapley value, $Sh_{b_e^u}(i)$, of a taxon $i \in I$ for each characteristic function, $b_e^u(\cdot)$. By the standard Shapley formula (see Eq.(1)), this quantity is:

$$\forall i \in I, \quad Sh_{b_e^u}(i) = \sum_{C\subset I, i\in C} \frac{(|C|-1)!(n-|C|)!}{n!}\left(b_e^u(C) - b_e^u(C\backslash\{i\})\right) \tag{10}$$

Let us now concentrate on the marginal contribution $(b_e^u(C) - b_e^u(C\backslash\{i\}))$ of taxa i to coalition C. This quantity belongs, *a priori*, to $\{-1, 0, 1\}$. But $b_e^u(C) = 0$ obviously implies by Eq.(4) that $b_e^u(C\backslash\{i\}) = 0$, meaning that this marginal contribution is either 0 or 1. Moreover, using Eq.(4) again, if $i \in I_e$, this quantity is equal to 1 if and only if $C\backslash\{i\}$ is non-empty and belongs to $\bar{I}_e$ while if $i \in \bar{I}_e$ this occurs if and only if the non-empty set $(C\backslash\{i\}) \subset I_e$. We now need to use combinatorial analysis to identify the number of occurrences of these cases, concluding after simplification that:

Lemma 2 *If $\{I_e \mid \bar{I}_e\}$ denotes the split associated with edge e, the Shapley value of taxa $i \in I$ for the characteristic function $b_e^u(\cdot)$ is:*

$$\forall i \in I, \quad Sh_{b_e^u}(i) = \left(\mathbf{I}_{i\in I_e}\frac{|\bar{I}_e|}{n\,|I_e|} + (1-\mathbf{I}_{i\in I_e})\frac{|I_e|}{n\,|\bar{I}_e|}\right) \tag{11}$$

$$\text{with } \mathbf{I}_{i\in I_e} = \begin{cases} 1 \text{ if } i \in I_e \\ 0 \text{ otherwise} \end{cases}$$

Proof Let us first observe that Eq.(10) can also be written as:

$$Sh_{b_e^u}(i) = \sum_{C\subset I\backslash\{i\}} \frac{|C|!(n-|C|-1)!}{n!} \underbrace{\left(b_e^u(C\cup\{i\}) - b_e^u(C)\right)}_{=\Delta_e} \tag{12}$$

Let us now assume that $i \in I_e$. In this case Eq.(4) says that $\Delta_e = 1$ iff $C \subseteqq \bar{I}_e$ and $C \neq \emptyset$. To verify this point, note that (i) if $C \neq \emptyset$ and $C \subseteqq \bar{I}_e$, then $b_e^u(C) = 0$ and $b_e^u(C\cup\{i\}) = 1$ since $i \in I_e$ so that $\Delta_e = 1$ and (ii) if not, either C meets both I_e and $\bar{I}_e$ so that $b_e(C) = b_e(C\cup\{i\}) = 1$, or $C \subset I_e$ which implies that $b_e^u(C) = b_e^u(C\cup\{i\}) = 0$, i.e. in both cases $\Delta_e = 0$. This preliminary observation clearly says that the terms composing $Sh_{b_e^u}(i)$ are different from 0 iff $C \subseteqq \bar{I}_e$ and $C \neq \emptyset$. Let us now observe that (i) there exists $\binom{|\bar{I}_e|}{|C|}$ potential choices of a subset of $c = |C|$ elements in $\bar{I}_e$ and (ii) the

size of this subset, C, can be $c = 1, \ldots, \left|\bar{I}_e\right|$. It follows that the Shapley value becomes:

$$Sh_{b_e^u}(i) = \sum_{c=1}^{\left|\bar{I}_e\right|} \frac{c!(n-c-1)!}{n!} \binom{\left|\bar{I}_e\right|}{c} = \frac{\left|\bar{I}_e\right|!}{n!} \sum_{c=1}^{\left|\bar{I}_e\right|} \frac{(n-c-1)!}{(\bar{I}_e - c)!} \tag{13}$$
$$= \frac{\left|\bar{I}_e\right|!\left(n-\left|\bar{I}_e\right|-1\right)!}{n!} \sum_{c=1}^{\left|\bar{I}_e\right|} \binom{n-c-1}{n-\left|\bar{I}_e\right|-1}$$

Moreover, an iterative use of the Pascal formula gives:

$$\binom{n-1}{n-\left|\bar{I}_e\right|} = \binom{n-2}{n-\left|\bar{I}_e\right|-1} + \binom{n-2}{n-\left|\bar{I}_e\right|} \tag{14}$$
$$= \binom{n-2}{n-\left|\bar{I}_e\right|-1} + \left(\binom{n-3}{n-\left|\bar{I}_e\right|-1} + \binom{n-3}{n-\left|\bar{I}_e\right|}\right)$$
$$= \ldots = \sum_{c=1}^{\left|\bar{I}_e\right|-1} \binom{n-c-1}{n-\left|\bar{I}_e\right|-1} + \binom{n-\left|\bar{I}_e\right|}{n-\left|\bar{I}_e\right|} = \sum_{c=1}^{\left|\bar{I}_e\right|} \binom{n-c-1}{n-\left|\bar{I}_e\right|-1}$$

This leads to:

$$Sh_{b_e^u}(i) = \frac{\left|\bar{I}_e\right|!\left(n-\left|\bar{I}_e\right|-1\right)!}{n!} \binom{n-1}{n-\left|\bar{I}_e\right|}$$
$$= \frac{\left|\bar{I}_e\right|}{n\left(n-\left|\bar{I}_e\right|\right)} = \frac{\left|\bar{I}_e\right|}{n\left|I_e\right|} \text{ (since } \left|I_e\right| + \left|\bar{I}_e\right| = n) \tag{15}$$

Finally, if $i \in \bar{I}_e$ it simply remains to apply the same argument by replacing $\bar{I}_e$ by I_e. This gives $Sh_{b_e^u}(i) = \frac{|I_e|}{n\left|\bar{I}_e\right|}$ and, combining both results, Eq.(11) follows. □

To go a step further, let us observe that Eq.(10) is homogeneous of degree 1 with respect to the characteristic function $b_e^u(\cdot)$, meaning that by multiplying $b_e^u(\cdot)$ by any scalar $\alpha \in \mathbb{R}$, one gets $\forall i \in I$, $Sh_{\alpha \cdot b_e^u}(i) = \alpha \cdot Sh_{\alpha \cdot b_e^u}(i)$. Since the construction of the Shapley value also requires additivity (see axiom 4), it follows that:

$$\forall i \in I, \quad Sh^u_{PD^u}(i) = Sh^u\left(\sum_{e \in E} w(e) b_e^u\right)(i) = \sum_{e \in E} w(e) Sh^u_{b_e^u}(i) \tag{16}$$

Since $\{b_e^u\}_{e \in E}$ is a basis of $\mathcal{P}^u$, we even say that this imputation rule is unique, meaning that there exists a unique function which attributes to each phylogenetic characteristic function $PD^u(\cdot)$ individual contributions to biodiversity.[3] To conclude this discussion, we can say that:

[3] This result is not in contradiction with Haake et al. (2007) Th.6., stating that several weighting structures can induce the same Shapley value. Our uniqueness result simply says that the imputation rule is unique but not that this map is injective

Proposition 1 *Let τ be an unrooted phylogenetic tree with n leaves whose split structure is given by $\mathcal{S}_\tau = \left\{\left\{I_e \mid \bar{I}_e\right\}_{e\in E}\right\}$. Under the standard Shapley approach based on axioms 1, 2, 3 and 4, the Shapley value of species i is given by:*

$$\forall i \in I, \quad Sh^u_\tau(i) = \sum_{e\in E} w(e) \underbrace{\left(\mathbf{I}_{i\in I_e} \frac{\left|\bar{I}_e\right|}{n\left|I_e\right|} + (1-\mathbf{I}_{i\in I_e}) \frac{\left|I_e\right|}{n\left|\bar{I}_e\right|}\right)}_{m_{i,e}} \tag{17}$$

$$\text{with } \mathbf{I}_{i\in I_e} = \begin{cases} 1 \text{ if } i \in I_e \\ 0 \text{ otherwise} \end{cases}$$

The previous proposition also shows that the Shapley operator $Sh^u_\tau(\cdot)$ is linear in weights, meaning that $Sh^u_\tau(\cdot) = M \cdot w(\cdot)$ with M a $(n, 2n-3)$ matrix whose generic term is $m_{i,e}$ for all $i \in I$ and all $e \in E$. Moreover, if $c(i,e)$ denotes the number of taxa that are on the same side of a split, s_e, as taxa i and $f(i,e)$ denotes its complement (i.e. $f(i,e) = n- c(i,e)$), we obtain, with a different proof, the formula proposed by Haake et al. (2007). In other words, we can say:

Corollary 1 (Haake et al. (2007) Th.4) *The Shapley operator is linear in weights, i.e. $Sh^u_\tau(\cdot) = M \cdot w(\cdot)$ and the (i,e)th entry of the Shapley transformation $(n, 2n-3)$ matrix, M, is given by:*

$$\forall i \in I, \forall e \in E, \qquad m_{i,e} = \frac{f(i,e)}{n \times c(i,e)} \tag{18}$$

4 Shapley value for rooted trees (Fuchs and Jin 2015)

The definition of the phylogenetic diversity index of a coalition changes. It remains the sum of the weights along a subtree, but is now computed on the subtree spanned by coalition C and the origin, 0. This means that the relevant edges are those which belong to any path, $\{i \to 0\}$, from a taxon $i \in C$ to the origin. If $E^r_c = \cup_{i\subset C} \{i \to 0\}$ denotes this set of edges, this new characteristic function writes:

$$\forall C \in 2^I, \quad PD^r(C) = \textstyle\sum_{e\in E^r_c} w(e) \tag{19}$$

This function, $PD^r(\cdot)$, as well as the subset, $\mathcal{P}^r$, of characteristic functions obtained by changing the weights are completely different. For instance, any coalition, $\{i\}$, formed by a single taxon $i \in I$ now has a non-zero phylogenetic diversity index. Moreover $\mathcal{P}^r$ is expected to be a linear subspace of dimension $(2n-2)$ since any rooted binary tree with n leaves contains $(2n-2)$ edges.

The derivation of the Shapley value, however, remains the same. We now consider a $(2n-2)$ family of rooted trees each associating a unit weight with a specific edge and zero elsewhere. But the definition of the characteristic functions $\{b^r_e(\cdot)\}_{e\in E}$ induced by these trees changes. For a coalition C, $b^r_e(C)$

is now equal to 1 if and only if at least one path from an element $i \in C$ to the origin meets edge $e \in E$, i.e. :

$$\forall C \in 2^I, \quad b_e^r(C) = \begin{cases} 1 \text{ if } \exists i \in C,\, e \in \{i \to 0\} \\ 0 \text{ otherwise} \end{cases} \tag{20}$$

Since the tree is rooted, each edge $e \in E$ can also be associated with the clade $I_e \subset I$, which contains the subset of all taxa descending from this edge. This observation again leads to a simpler definition of the family $\{b_e^r(\cdot)\}_{e\in E}$ since the existence of a path from a taxon $i \in C$ to the origin meeting edge e is equivalent to claiming that coalition C meets clade I_e. It follows that $b_e^r(C)$ becomes:

$$\forall C \in 2^I, \quad b_e^r(C) = \begin{cases} 1 \text{ if } C \cap I_e \neq \emptyset \\ 0 \text{ otherwise} \end{cases} \tag{21}$$

and it can be shown that:

Lemma 3 *$\{b_e^r\}_{e\in E} \subset \mathcal{P}^r$ is a basis of dimension $(2n-2)$ of the linear set $\mathcal{P}^r$.*

Proof Let us again assume that there exists, $(\alpha_e)_{e\in E} \neq 0$, a vector of scalars in $\mathbb{R}^{2n-2}$ such that $\forall C$, $\sum_{e\in E}\alpha_e b_e^r(C) = 0$. To show that this assertion is wrong, we proceed by induction by starting from the leaves and by moving back to the origin of the tree. So let us note by e_i the external edge leading to leaf $i \in I$ and let I_e be the clade induced by each edge $e \in E$. From Eq.(21), $b_e^r(I \setminus \{i\}) = 0$ iff $(I \setminus \{i\}) \cap I_e = \emptyset$. This only occurs if $I_e = \{i\}$, that is for the clade associated with edge e_i. We can therefore say, under absence of independence, that:

$$\sum_{e\in E}\alpha_e b_e^r(I \setminus \{i\}) = \sum_{e\in E\setminus\{e_i\}} \alpha_e = 0 \tag{22}$$

Since we also know (see Eq.(9)) that, without independence, $\sum_{e\in E}\alpha_e = 0$, we immediately conclude that $\alpha_{e_i} = 0$ for all $b_{e_i}^r(\cdot)$ associated with an external edge, e_i.

Now let us consider an internal edge e whose clade is I_e, let us note by E_s the subset of edges contained in the subtree which follows edge e, and let us assume that $\forall e' \in E_s$, $\alpha_{e'} = 0$. To extend our result by induction, we now need to compute $b_{e'}^r\left(\bar{I}_e\right)$ for all $e' \in E$. From Eq.(21), $b_{e'}^r\left(\bar{I}_e\right) = 0$ iff $\left(\bar{I}_e\right) \cap I_{e'} = \emptyset$. This situation only occurs if $I_{e'} \subsetneqq I_e$, i.e. for clades associated with edges $e' \in E_s$. We can therefore say that:

$$\sum_{e'\in E}\alpha_{e'} b_{e'}^r\left(\bar{I}_e\right) = \sum_{e'\in E\setminus\{E_s\cup\{e\}\}} \alpha_{e'} = 0 \tag{23}$$

Moreover, by Eq.(9), we can say that $\sum_{e\in E}\alpha_e = 0$ and since we have assumed that $\forall e_s \in E_s$, $\alpha_{e_s} = 0$, we get $\sum_{e'\in E\setminus E_s} \alpha_{e'} = 0$. Comparing with Eq.(23), we conclude that $\alpha_e = 0$.

It remains to use this argument by induction to say that $(\alpha_e)_{e\in E} = 0$ which is the desired contradiction. □

The next step is computating the Shapley values for each characteristic function $b_e^r(\cdot)$ as in Eq.(10) where $b_e^r(\cdot)$ replaces $b_e^u(\cdot)$. It is immediate that the marginal contribution $(b_e^r(C) - b_e^r(C\backslash\{i\}))$ of agent i to coalition C is again either 0 or 1 since $b_e^r(C) = 0$ always implies that $b_e^r(C\backslash\{i\}) = 0$. Moreover, if $i \notin I_e$, the path $\{i \to 0\}$ never meets edge e, meaning that i's marginal contribution to any coalition is always 0. In the opposite case, i.e. $i \in I_e$, this taxon makes a real marginal contribution if and only if the rest of coalition C is included in $\bar{I}_e$ (the subset $C\backslash\{i\}$ could be empty). It remains, by a combinatorial argument, to identify the number of cases in which this last situation occurs and to reach the conclusion that:

Lemma 4 *If I_e denotes the clade associated with edge e, the Shapley value of taxon $i \in I$ for the characteristic function $b_e^r(\cdot)$ is:*

$$\forall i \in I, \quad Sh_{b_e^r}(i) = \mathbf{I}_{i\in I_e}\frac{1}{|I_e|} \text{ with } \mathbf{I}_{i\in I_e} = \begin{cases} 1 \text{ if } i \in I_e \\ 0 \text{ otherwise} \end{cases} \tag{24}$$

Proof As in the proof of Lemma 2, it remains to identify the cases in which Δ_e of Eq.12 is equal to one (when, of course, $b_e^u(\cdot)$ is replaced by $b_e^r(\cdot)$).
Let us first assume that $i \in I_e$. From Eq.(21), (i) $b_e^r(C \cup \{i\}) = 1$ for any $C \subset I\backslash\{i\}$ and (ii) $b_e^r(C) = 0$ iff $C \subsetneqq \bar{I}_e$ (C being possibly empty), meaning that $\Delta_e = 1$ iff $C \subsetneqq \bar{I}_e$. Moreover, there again exist $\binom{|\bar{I}_e|}{|C|}$ potential choices of a subset of $c = |C|$ elements in $\bar{I}_e$ but now $c = 0, \ldots, |\bar{I}_e|$ since C can be empty. The Shapley value is therefore:

$$\begin{aligned} Sh_{b_e^r}(i) &= \sum_{c=0}^{|\bar{I}_e|} \frac{c!(n-c-1)!}{n!}\binom{|\bar{I}_e|}{c} = \frac{|\bar{I}_e|!}{n!}\sum_{c=0}^{|\bar{I}_e|}\frac{(n-c-1)!}{(\bar{I}_e-c)!} \\ &= \frac{|\bar{I}_e|!(n-|\bar{I}_e|-1)!}{n!}\sum_{c=0}^{|\bar{I}_e|}\binom{n-c-1}{n-|\bar{I}_e|-1} \end{aligned} \tag{25}$$

By adapting the previous iterative Pascal formula (see Eq.(14)), we obtain:

$$\begin{aligned} Sh_{b_e^u}(i) &= \frac{|\bar{I}_e|!(n-|\bar{I}_e|-1)!}{n!}\left(\binom{n-1}{n-|\bar{I}_e|} + \binom{n-1}{n-|\bar{I}_e|-1}\right) \\ &= \frac{1}{n}\left(\frac{|\bar{I}_e|}{(n-|\bar{I}_e|)} + 1\right) = \frac{1}{|I_e|} \text{ (since } |I_e| + |\bar{I}_e| = n) \end{aligned} \tag{26}$$

Now assume that $i \notin I_e$. In this case either $b_e^r(C) = 1$ or 0 depending whether $C \cap I_e \neq \emptyset$ or not, but in any case adding i to C makes no difference since $i \notin I_e$. It follows from the null player axiom (see Axiom 3) that $Sh_{b_e^r}(i) = 0$. By combining both results, we get Eq.(24). □

We can observe, with the same argument as in the preceding section, that the Shapley value is still a linear operator on $\mathcal{P}^r$ and we can conclude that:

Proposition 2 *Let τ be a rooted phylogenetic tree with n leaves whose clade structure is given by $\mathcal{C}_\tau = \{I_e\}_{e\in E}$. Under the standard Shapley approach based on axioms 1, 2, 3 and 4, the Shapley value of species i is given by:*

$$Sh^r_\tau(i) = \sum_{e\in E} \mathbf{I}_{i\in I_e} \frac{w(e)}{|I_e|} \text{ with } \mathbf{I}_{i\in I_e} = \begin{cases} 1 \text{ if } i \in I_e \\ 0 \text{ otherwise} \end{cases} \tag{27}$$

This expression can even be simplified. Due to the presence of the indicator function $\mathbf{I}_{i\in I_e}$, the sum only runs over the set of edges, e, whose clade, I_e, contains taxon i. Moreover, it is straightforward to verify that this set can be identified with the set of all edges contained in the path $\{i \to 0\}$. This means that the Shapley value is the sum along $\{i \to 0\}$ of the weights associated with these edges divided by $D_e = |I_e|$, the number of taxa descending from each of these edges. But this is the definition of the Fair Proportion index introduced by Redding and Mooers (2006).

Corollary 2 (Fuchs and Jin (2015) Th.1) *The Shapley value of an individual in a rooted binary tree is equal to the Fair Proportion index, i.e.*

$$Sh^r_\tau(i) = \sum_{e\in\{i\to 0\}} \frac{w(e)}{D_e} = FP^r_\tau(i) \tag{28}$$

5 Comparing the two Shapley values

Since any comparison calls for identical trees, we now move to *unrooted rooted trees.* For these trees, the two previous propositions remain true even if the proof of the different Lemmas should be slightly adjusted.

Let us begin to examine the difference between these two Shapley values. This quantity is given by:

$$\begin{aligned} &\forall i \in I,\ \Delta_{sh}(i) = Sh^u_\tau(i) - Sh^r_\tau(i) \\ &= \tfrac{1}{n} \sum_{e\in E} w(e) \left(-\mathbf{I}_{i\in I_e} + (1 - \mathbf{I}_{i\in I_e}) \frac{|I_e|}{|\bar{I}_e|} \right) \end{aligned} \tag{29}$$

Since both Shapley operators are linear with respect to the weight vector, $w = (w(e))_{e\in E}$, the difference, $\Delta_{sh} = (\Delta_{sh}(i))_{i\in I}$, shares the same property and can be written as:

$$\Delta_{sh} = \tfrac{1}{n} A \cdot w, \text{ the generic term of } A \text{ being } a_{i,e} = \begin{cases} -1 \text{ if } i \in I_e \\ \frac{|I_e|}{|\bar{I}_e|} \text{ otherwise} \end{cases} \tag{30}$$

where A is a $(n, 2n-2)$ matrix.

The properties of this matrix are crucial for the comprehension of the difference between these two Shapley values. Note that both values satisfy

efficiency (see axiom 1). As a consequence, $\sum_{i\in I}\Delta_{sh}(i)=0$ independently of the weight vector, w. We can therefore say, if ε_n denotes the unit vector of $\mathbb{R}^n$, that $\varepsilon_n'\cdot A=0$, or, in other words, that A is at most of rank $(n-1)$ since its image, $\langle A\rangle$, is orthogonal to ε_n. This also suggests that one component of Δ_{sh} is redundant. We dismiss the last component of this vector and call $\tilde{\Delta}_{sh}$ the new difference vector. If we now denote by E_{ext} and E_{int} the sets of relevant external and internal edges and by e_n the edge reaching taxa n, $\tilde{\Delta}_{sh}$ can be written as:

$$\left(\tilde{\Delta}_{sh}(i)\right)_{i=I\setminus\{n\}}= \tag{31}$$

$$\frac{1}{n}\left[\underbrace{[a_{i,e}]_{i=I\setminus\{n\},e\in E_{ext}\setminus\{e_n\}}}_{=B_{ext}}\mid\underbrace{[a_{i,e_n}]_{i=I\setminus\{n\}}}_{=B_{e_n}}\mid\underbrace{[a_{i,e}]_{i=I\setminus\{n\},e\in E_{int}}}_{=B_{int}}\right]\cdot w$$

It remains to prove that the $(n-1,n-1)$ matrix B_{ext} is invertible in order to claim that:

Lemma 5 *A is a linear mapping of rank $(n-1)$. Its kernel which is also of dimension $(n-1)$ is given by the set of weights, $w\in\mathbb{R}^{2n-2}$ which satisfies:*

$$(w(e))_{e\in E_{ext}\setminus\{e_n)}=$$

$$\tfrac{n-1}{n}\left(I_{n-1}+\varepsilon_{n-1}\cdot(\varepsilon_{n-1})'\right)\cdot[B_{e_n}\mid B_{int}]\cdot\begin{pmatrix}w(e_n)\\(w(e))_{e\in E_{int}}\end{pmatrix} \tag{32}$$

Proof The kernel, $\ker(A)$, of A satisfies $Aw=0$. But we know that, for the unit vector $\varepsilon_n\in\mathbb{R}^n$, $(\varepsilon_n)'\cdot A=0$, meaning that the last equation in $Aw=0$ is redundant. It follows from Eq.(31) that:

$$w\in\ker(A)\Leftrightarrow[B_{ext}\mid B_{e_n}\mid B_{int}]\cdot\begin{pmatrix}(w(e))_{e\in E_{ext}\setminus\{e_n\}}\\w(e_n)\\(w(e))_{e\in E_{int}}\end{pmatrix}=0 \tag{33}$$

Let us now concentrate on the $(n-1,n-1)$ matrix B_{ext}. To construct this matrix, we need to come back to the early definition of the generic term a_{ie} of the matrix A (see Eq.(30) and observe, since we are looking at external edges, that $I_e=\{i\}$. It follows that:

$$B_{ext}=\begin{bmatrix}-1&\frac{1}{n-1}&\cdots&\frac{1}{n-1}\\\frac{1}{n-1}&\ddots&\ddots&\vdots\\\vdots&\ddots&\ddots&\frac{1}{n-1}\\\frac{1}{n-1}&\cdots&\frac{1}{n-1}&-1\end{bmatrix}=-\tfrac{n}{n-1}\left(I_{n-1}-\tfrac{1}{n}\varepsilon_{n-1}\cdot(\varepsilon_{n-1})'\right) \tag{34}$$

with I_{n-1} the identity matrix of $\mathbb{R}^{n-1}$ and ε_{n-1} the unit vector of $\mathbb{R}^{n-1}$. Moreover, since $(\varepsilon_{n-1})'\cdot\varepsilon_{n-1}=n-1$, a simple computation shows that:

$$\left(I_{n-1}-\tfrac{1}{n}\varepsilon_{n-1}\cdot(\varepsilon_{n-1})'\right)\cdot\left(I_{n-1}+\varepsilon_{n-1}\cdot(\varepsilon_{n-1})'\right)=I_{n-1} \tag{35}$$

It follows that $(B_{ext})^{-1} = -\frac{n-1}{n}\left(I_{n-1} - \varepsilon_{n-1}\cdot(\varepsilon_{n-1})'\right)$. Eq.(32) of Lemma 5 directly follows from the block decomposition provided by Eq.(33). This also shows that $\dim(\ker(A)) = n-1$. □

The fact that ε_n is orthogonal to the image $\langle A\rangle$ has an additional consequence. We can claim that it is impossible to find a system of weights, w , which verifies either $Aw < 0$ or $Aw > 0$, otherwise there is a contradiction with $\varepsilon_n' \cdot A = 0$. In other words, it is impossible to have $\forall i$, $Sh_\tau^u(i) \geq Sh_\tau^r(i)$ or $Sh_\tau^u(i) \leq Sh_\tau^r(i)$ with at least one strict inequality, meaning that neither the unrooted nor the rooted Shapley value dominates the other.
The next proposition summarizes this first discussion.

Proposition 3 *For a given tree structure, we can say that:*
(i) the two Shapley values are identical on a $(n-1)$ linear subset of weights satisfying Eq.(32),
(ii) for all other weights, there exist at least two taxa with different Shapley values,
(iii) neither the unrooted nor the rooted Shapley value induces a higher evaluation of the contribution to biodiversity for all taxa,
(iv) if for one taxon, i, $Sh_\tau^u(i) > Sh_\tau^r(i)$ then there always exists another taxon, j, for which $Sh_\tau^u(j) < Sh_\tau^r(j)$.

In line with Hartman (2013), let us now compare the contribution of the weight of an edge to each Shapley value. This contribution is straightforward for the rooted Shapley value. Since the latter is equal to the Fair Proportion index, each edge only contributes to its descendants, I_e, in a proportion, $\frac{1}{|I_e|}$, inverse to their number. For an unrooted Shapley value (see Eq.(17)), the sharing rule for the weight associated with an edge is totally different. Each edge contributes to the Shapley value of each taxon under a two step sharing rule. First, the weight is shared out between the descendents and the non-descendents: the descendents as a whole group receive a quantity proportional to the number of non-descendents, $\frac{n-|I_e|}{n}$, and vice-versa. These overall amounts are then uniformly shared out among the members of each group. To summarize, we can say:

Proposition 4 *The contribution of the weight, $w(e)$, of an edge, $e \in E$, to the two Shapley values of taxon $i \in I$ is summarized in the following:*

	$i \in I_e$	$i \notin I_e$
$Sh_\tau^u(i)$	$\frac{1}{\lvert I_e\rvert}\left(\frac{n-\lvert I_e\rvert}{n}\right)$	$\frac{1}{n-\lvert I_e\rvert}\left(\frac{\lvert I_e\rvert}{n}\right)$
$Sh_\tau^r(i)$	$\frac{1}{\lvert I_e\rvert}$	0

As a consequence, we immediately observe that:

Corollary 3 (Hartman (2013) Th.1) *As the number of taxa $n \to \infty$, the contribution of an edge e to the rooted and unrooted Shapley values becomes the same.*

6 Comparing the ordering induced by the two Shapley values

Up to now, proposition 3 tells us that the two Shapley values are equal, and in fact equal to the Fair Proportion index, on a linear subset of weights of dimension, this subset being largely related to the split structure of the tree under consideration. But this makes us wonder what happens when weight structures do not belong to this set. In particular, if we consider prioritization programs, do these two Shapley values induce drastically different species rankings or are they fairly similar ? To answer this question, we proceed by simulations[4] . For configurations where these two Shapley values are known to be different, we randomly choose weights and/or tree structures, compute the two Shapley values and look at the Kendall (1938) and the Spearman (1904) rank correlation coefficient. The first statistic essentially captures the inversions of elements in the two rankings, while the second highlights differences in the position of an element in each ranking. But in any case, if both statistics are close to 1, the two rankings are considered strongly (positively) correlated.

We perform these simulations in two steps. In the first step, we arbitrarily fix a symmetric tree inducing 16, 32 and 64 species and randomly select weights that do not belong to the subset where these Shapley values are identical. In the second step, we control for the symmetry assumption. We choose randomly not only the weights but also the split structure of the tree, nevertheless keeping the number of leaves constant.

In the first set of simulations, we look at symmetric binary trees with 16, 32 and 64 leaves and consider 5000 randomly selected edge weights. This selection satisfies two constraints. Naturally, the weights are chosen such that the Shapley values are different, but we also normalize the sum of the weights at 1. This second restriction follows from the fact that the two Shapley values are linear in weighting, meaning that if we multiply the weighting structure by any $\lambda > 0$ we obtain the same ranking. The normalization rule eliminates this redundancy. For each of these symmetric trees and each random weighting structure, we compute the two Shapley values, look at the induced ranking and compute the Kendall and Spearman correlation coefficients. We obtain 2×3 sets of 5000, respectively, Kendall and Spearman rank correlation coefficients associated with symmetric trees with respectively, 16, 32 and 64 and leaves. The main characteristics of these distributions are summarized in Tab. 1.

Table 1 Main characteristics of the Kendall and Spearman coef.: the symmetric tree case

	Kendall 16 species	Spearman 16 species	Kendall 32 species	Spearman 32 species	Kendall 64 species	Spearman 64 species
Mean	.9441	.9861	.9688	.9956	.9825	.9988
Median	.9500	.9912	.9718	.9967	.9841	.9989
St. dev.	.0402	.0137	.0167	.0031	.0073	.0007

[4] The Matlab R2017a codes are available upon request

These results clearly suggest that, at least in the symmetric case, the two rankings induced by the two Shapley values are largely correlated and, as expected (see Hartman 2013), the degree of correlation increases with the number of species. This intuition is confirmed by a kernel estimation of the density of these different distributions (see Fig.1).

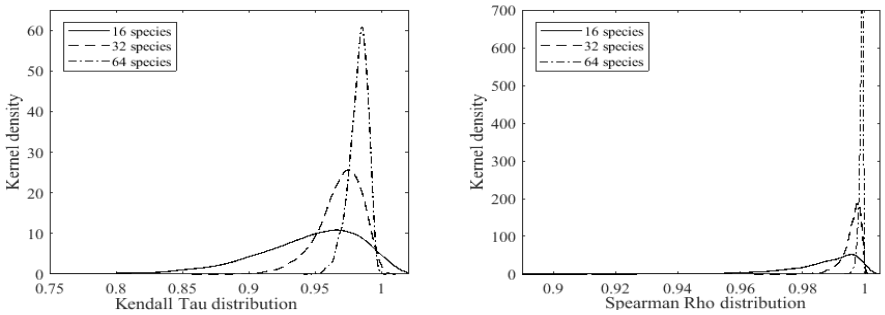

Fig. 1 Kernel Density estimates: the symmetric tree case

In the second set of simulations, we again look at trees with 16, 32 and 64 leaves, but we not only randomly select the weights as in the previous set of simulations, we also randomly construct a compatible split structure of these trees by keeping the number of leaves fixed. We again obtain 2×3 sets of 5000 respectively Kendall and Spearman rank correlation coefficients. The Kernel estimates of these densities are given in Fig. 2 while the main characteristics are summarized in Tab.2

So even though the introduction of random trees decreases the mean of each distribution and increases the standard deviation, especially when the number of taxa is small, the rankings induced by these two Shapley values remain largely correlated.

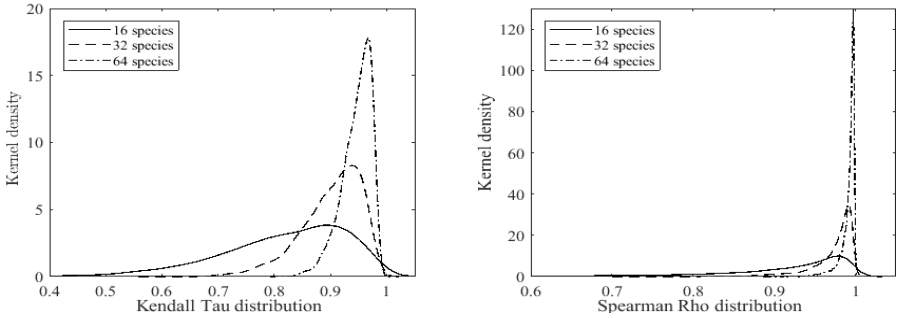

Fig. 2 Kernel Density estimates: the random tree case

Table 2 Main characteristics of the Kendall and Spearman coef.: the random tree case

	Kendall 16 species	Spearman 16 species	Kendall 32 species	Spearman 32 species	Kendall 64 species	Spearman 64 species
Mean	.8182	.9121	.9013	.9720	.9464	.9913
Median	.8333	.9441	.9113	.9824	.9514	.9948
St. dev.	.1125	.0919	.0539	.0303	.0268	.0099

7 Concluding remarks

The purpose of this paper was to explore the main differences between the Shapley Values for a taxon introduced by Haake et al. (2007) and Fuchs and Jin (2015). Although these two metrics are based on the same Shapley axiomatic, they do not use the same definition of Phylogenetic Diversity: the former does not include the root in the subtree while the latter does. This induces two different subsets of potential characteristic functions and different Shapley values. To illustrate this, we explicitly compute the Shapley values in both cases by identifying, in line with Shapley, the basis of the two different sets of characteristic functions. This gives us the opportunity to formally compare the two values and to identify for each phylogenetic tree a set of weights for which these quantities are identical. This clearly raises a second question where prioritization problems are concerned. Although the two Shapley values are different, do they induce a similar ranking for the different species? To answer this question, we simulate alternative situations and show that the Kendall and Spearman rank correlation coefficients are both close to 1. Bearing in mind that the Shapley value introduced by Fuchs and Jin (2015) is equal to the Fair Proportion index constructed by Redding and Mooers (2006), this suggests that the latter, simpler index can be used in prioritization problems.

References

1. Cadotte M. W., Jonathan Davies, T., Regetz J., Kembel S. W., Cleland E., Oakley T. H. (2010), Phylogenetic diversity metrics for ecological communities: integrating species richness, abundance and evolutionary history. Ecol. Lett., 13:96–105. https://doi.org/10.1111/j.1461-0248.2009.01405.x
2. Faith D. P. (1992) Conservation evaluation and phylogenetic diversity. Biol. Conserv. 61:1 10. https://doi.org/10.1016/0006-3207(92)91201-3
3. Fuchs, M., and Jin, E. Y.,(2015) Equality of shapley value and fair proportion index in phylogenetic trees. J. Math. Biol. 71:1133–1147. https://doi.org/10.1007/s00285-014-0853-0
4. Haake C.-J., Kashiwada A., Su F. E., (2007) The Shapley value of phylogenetic trees. J. Math. Biol. 56:479–497. https://doi.org/10.1007/s00285-007-0126-2
5. Hartmann K. (2013) The equivalence of two phylogenetic biodiversity measures: the shapley value and fair proportion index. J. Math. Biol. 67:1163–1170. https://doi.org/10.1007/s00285-012-0585-y
6. Jensen E.L., Mooers A.Ø., Caccone A., Russello M.A. (2016) I-HEDGE: determining the optimum complementary sets of taxa for conservation using evolutionary isolation. PeerJ 4:e2350 https://doi.org/10.7717/peerj.2350

7. Kendall, M., (1938) A New Measure of Rank Correlation. Biomet. 30:81–89 https://doi.org/10.2307/2332226
8. Kleinberg, N. L., Weiss J. H., (1985) A new formula for the Shapley value. Econ. Lett. 18 : 311 - 315. https://doi.org/10.1016/0165-1765(85)90249-6
9. Martyn I., Kuhn T.S., Mooers A.Ø., Moulton V. and Spillner A., (2012) Computing evolutionary distinctiveness indices in large scale analysis. Algorithm. Mol. Biol. 7:6. https://doi.org/10.1186/1748-7188-7-6
10. Redding D.W., Mooers A.Ø. (2006) Incorporating evolutionary measures into conservation prioritization", Conserv. Biol. 20:1670–1678. https://doi.org/10.1111/j.1523-1739.2006.00555.x
11. Redding D.W., Mazel F., Mooers A.Ø., (2014) Measuring evolutionary isolation for conservation. PLoS ONE (2014) 9(12): e113490 https://doi.org/10.1371/journal.pone.0113490
12. Rothblum, U. G. (1988) Combinatorial representations of the Shapley value based on average relative payoffs in: Roth A. E. (ed), The Shapley value: Essays in honor of Lloyd S. Shapley, Cambridge University Press, Cambridge pp 121-126
13. Shapley, L.S. (1953) A Value for n-Person Games, in H. W. Kuhn and A. W. Tucker (eds) Contributions to the Theory of Games, vol. II, Ann. Math. Studies 28, Princeton University Press, Princeton, New Jersey, pp 307-17
14. Spearman C., (1904) The proof and measurement of association between two things. Am. J. Psychol. 15: 72–101. https://doi.org/10.2307/1412159
15. Volkmann L., Martyn I., Moulton V., Spillner A., Mooers A.Ø., (2014) Prioritizing Populations for Conservation Using Phylogenetic Networks PLoS ONE 9(2): e88945. https://doi.org/10.1371/journal.pone.0088945
16. Weitzman M.L.(1998) The Noah's Ark problem. Econometrica 66:1279-1298. https://doi.org/10.2307/2999617