

1 **PHYLOGENETIC DIVERSITY INDICES FROM AN AFFINE AND PROJECTIVE**  
2 **VIEWPOINT**

3 V. MOULTON, A. SPILLNER, AND K. WICKE

ABSTRACT. Phylogenetic diversity indices are commonly used to rank the elements in a collection of species or populations for conservation purposes. The derivation of these indices is typically based on some quantitative description of the evolutionary history of the species in question, which is often given in terms of a phylogenetic tree. Both rooted and unrooted phylogenetic trees can be employed, and there are close connections between the indices that are derived in these two different ways. In this paper, we introduce more general phylogenetic diversity indices that can be derived from collections of subsets (clusters) and collections of bipartitions (splits) of the given set of species. Such indices could be useful, for example, in case there is some uncertainty in the topology of the tree being used to derive a phylogenetic diversity index. As well as characterizing some of the indices that we introduce in terms of their special properties, we provide a link between cluster-based and split-based phylogenetic diversity indices that uses a discrete analogue of the classical link between affine and projective geometry. This provides a unified framework for many of the various phylogenetic diversity indices used in the literature based on rooted and unrooted phylogenetic trees, generalizations and new proofs for previous results concerning tree-based indices, and a way to define some new phylogenetic diversity indices that naturally arise as affine or projective variants of each other or as generalizations of tree-based indices.

4 1. INTRODUCTION

5 Evolutionary isolation metrics or phylogenetic diversity indices provide quantitative measures  
6 of biodiversity and are increasingly popular tools to prioritize species for conservation (Isaac  
7 et al., 2007; Redding et al., 2008, 2014; Redding and Mooers, 2006; Tucker et al., 2016; Vane-  
8 Wright et al., 1991). These indices quantify the importance of a species to overall biodiversity  
9 by assessing its unique and shared evolutionary history as indicated by its placement in an un-  
10 derlying phylogeny. Preserving phylogenetic diversity and the “Tree of Life” has become an  
11 integral component of conservation considerations (see, e.g., the “Phylogenetic Diversity Task  
12 Force”<sup>1</sup> initiated by the IUCN). Indeed, conservation initiatives like the EDGE of Existence  
13 programme<sup>2</sup> (Gumbs et al., 2023; Isaac et al., 2007) incorporate phylogenetic diversity indices  
14 in their identification of species that are both evolutionary distinct and globally endangered.

---

*Date:* June 20, 2024.

School of Computing Sciences, University of East Anglia, UK.

Merseburg University of Applied Sciences, Germany.

Department of Mathematical Sciences, New Jersey Institute of Technology, USA.

<sup>1</sup><https://www.pdtf.org/>

<sup>2</sup><https://www.edgeofexistence.org/>

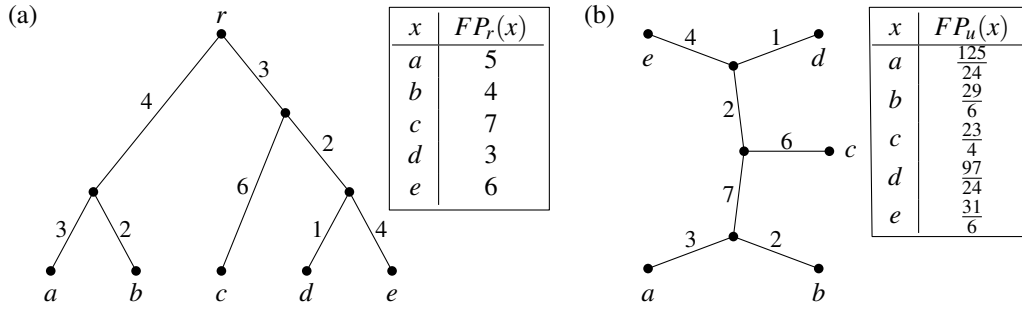


FIGURE 1. (a) A rooted phylogenetic tree on the set  $X = \{a, b, c, d, e\}$  of species. The root vertex is  $r$  and all edges are weighted. The table gives the value  $FP_r(x)$  of the fair proportion index on this rooted tree for each  $x \in X$ . (b) The unrooted phylogenetic tree with weighted edges on the same set  $X$  of species obtained by suppressing the root of the tree in (a). The table gives the value  $FP_u(x)$  of the fair proportion index on this unrooted tree for each  $x \in X$ .

15 Moreover, the “guide to phylogenetic metrics for conservation, community ecology and macroe-  
 16 cology” by Tucker et al. (2016) has been cited more than 700 times since its publication, thus  
 17 demonstrating an even more widespread interest and application of phylogenetic tools, and in  
 18 particular different phylogenetic diversity indices, within conservation settings.

19 Mathematically, with a multitude of phylogenetic diversity indices at hand, there is now an  
 20 increasing interest in understanding how the different indices relate to each other. Much of the  
 21 previous work in this direction has focused on comparing and analyzing different indices de-  
 22 rived from rooted phylogenetic trees (Bordewich and Semple, 2024; Manson, 2024; Manson and  
 23 Steel, 2023; Wicke and Steel, 2020). Phylogenetic diversity indices have also been defined for  
 24 unrooted trees (Haake et al., 2008; Wicke and Steel, 2020), and an exploration of the relationship  
 25 between indices derived via rooted and unrooted phylogenetic trees is presented by Wicke and  
 26 Steel (2020).

27 As one might expect, phylogenetic diversity indices for rooted and unrooted trees are closely  
 28 related. To illustrate this, consider the much studied *fair proportion index* (Isaac et al., 2007;  
 29 Redding, 2003). For the rooted phylogenetic tree with edge weights in Figure 1(a), the value  
 30  $FP_r(x)$  of the rooted fair proportion index for a species  $x \in X$  (here and throughout this manu-  
 31 script,  $X$  denotes a non-empty finite set of taxa or species) is computed by adding, over all edges  
 32 that are contained in the path from the root  $r$  to the leaf labeled by  $x$ , the weight of the edge  
 33 divided by the total number of species for which the path from the root to the leaf labeled by that  
 34 species also contains that edge. For example, for species  $e$  there are three edges in the path from  
 35  $r$  to  $e$  and we obtain

$$36 \quad (1) \quad FP_r(e) = \frac{3}{3} + \frac{2}{2} + \frac{4}{1} = 6.$$

37 In Wicke and Steel (2020) the fair proportion index has also been defined for unrooted phy-  
 38 logenetic trees. Consider the unrooted phylogenetic tree with edge weights in Figure 1(b). The  
 39 removal of an edge breaks the tree into two subtrees. The value  $FP_u(x)$  of the unrooted fair

40 proportion index for a species  $x \in X$  is one half of the value obtained by adding, over all edges  
 41 in the unrooted tree, the weight of the edge divided by the number of species that lie in the same  
 42 subtree as  $x$  after removal of the edge. For example, for species  $e$  we obtain

$$43 \quad (2) \quad FP_u(e) = \frac{1}{2} \cdot \left( \frac{3}{4} + \frac{2}{4} + \frac{7}{3} + \frac{6}{4} + \frac{2}{2} + \frac{1}{4} + \frac{4}{1} \right) = \frac{31}{6}.$$

44 As can be seen in Figure 1,  $\sum_{x \in X} FP_r(x) = \sum_{x \in X} FP_u(x) = 25$ , which is the total weight of  
 45 the edges of the phylogenetic tree from which the values are computed. Among other natu-  
 46 ral requirements, this property called *completeness* (formally defined in Section 2), should be  
 47 preserved when relating phylogenetic diversity indices for rooted and unrooted trees.

48 To better understand how this can be systematically achieved, in this paper we consider indices  
 49 from the viewpoint of affine and projective clustering. This way of thinking about clustering has  
 50 its origins in (Dress, 1997), and since then has become a useful tool in phylogenetic combina-  
 51 torics (see, e.g., Dress 2012, Ch. 9 and Kleinman et al. 2013). More specifically, in this paper  
 52 we extend the study of phylogenetic diversity indices into the more general setting of collections  
 53 of *clusters* (subsets of a set) and collections of *splits* (bipartitions of a set). These settings corre-  
 54 spond to affine and projective viewpoints of clustering, respectively (see Section 5). Considering  
 55 collections of clusters and splits in general can be beneficial since it allows for the representation  
 56 of data that is not tree-like or where it is difficult to determine the correct topology for a phylo-  
 57 genetic tree. Indeed, phylogenetic diversity indices have already been introduced for collections  
 58 of splits (see, e.g., Abhari et al. 2024).

59 To illustrate this way of thinking, as hinted above, collections of clusters naturally arise when  
 60 computing the rooted fair proportion index. In particular, clusters arise from rooted phylogenetic  
 61 trees by taking, for each edge, the subset of species for which the path from the root to that  
 62 species contains the edge (e.g., in Figure 1(a) the edge with weight 3 next to the root gives rise to  
 63 the cluster  $\{c, d, e\}$ ). Thus, the sum used to compute the fair proportion index of  $e$  in Equation (1)  
 64 is just the sum of the values  $\frac{\omega(C)}{|C|}$  taken over all clusters  $C$  that contain  $e$ , where  $\omega(C)$  is the weight  
 65 of the edge giving rise to cluster  $C$  and  $|C|$  denotes the number of species in  $C$ . Similarly, we can  
 66 interpret Equation (2) in terms of splits, using the fact that splits arise from unrooted phylogenetic  
 67 trees by taking, for each edge, the split obtained by removing the edge and considering the subsets  
 68 of species in the two resulting subtrees (e.g., in Figure 1(b) the edge with weight 7 gives rise to  
 69 the split  $\{\{a, b\}, \{c, d, e\}\}$ ). Then the sum used to compute the unrooted fair proportion index of  
 70  $e$  in Equation (2) is just the sum of the values  $\frac{\lambda(S)}{2|A|}$  taken over all splits  $S$  coming from the tree,  
 71 where  $\lambda(S)$  is the weight of the edge giving rise to  $S$  and  $A$  is the part in  $S$  that contains  $e$ . More  
 72 generally, the sums used to compute  $FP_r$  and  $FP_u$  can be applied to *any* collection of weighted  
 73 clusters or splits, respectively (for example, the values for  $FP_u$  computed for a collection of  
 74 weighted splits visualized by the network in Figure 2(a) are shown in the second column in  
 75 Figure 2(b)).

76 Thinking about phylogenetic diversity indices in an affine and projective way, leads us to two  
 77 key questions that we will consider in this paper:

- 78 (i) How do properties of tree-based phylogenetic diversity indices extend to indices defined  
 79 via collections of clusters and splits?

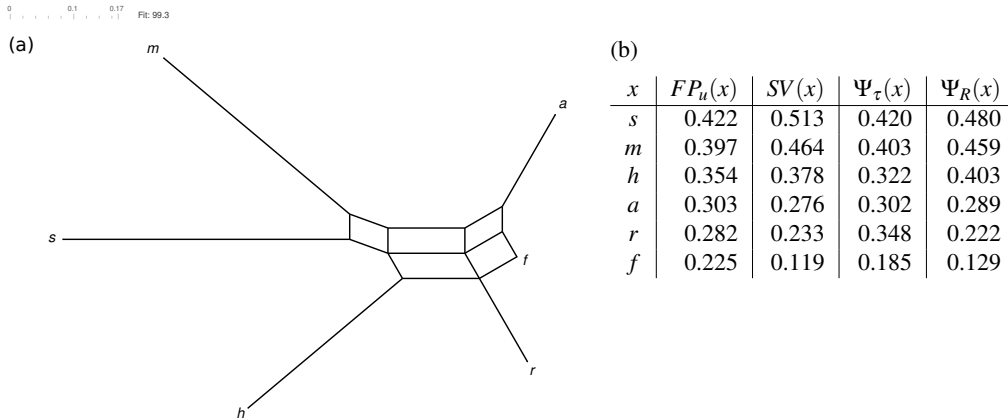


FIGURE 2. (a) A network visualizing a collection  $\mathcal{S}$  of weighted splits on the set  $X = \{a, f, h, m, r, s\}$  of six owl populations (see Figure 12 in the appendix for more details on this data set). Each band of parallel edges in this network corresponds to a split of  $X$  and the length of the edges in the band corresponds to the weight of the split. (b) The values of four different phylogenetic diversity indices  $FP_u$ ,  $SV$ ,  $\Psi_\tau$  and  $\Psi_R$  considered in this paper, computed for the six owl populations. The populations are ranked by the values of  $FP_u$ .

80 (ii) How can the relationships between collections of clusters and collections of splits be  
81 exploited to relate cluster- and split-based phylogenetic diversity indices?

82 In this contribution, we give answers to both of these questions, introducing the concept of phylo-  
83 genetic diversity indices based on collections of clusters and splits, and giving characterizations  
84 for some of these indices in terms of their special properties. We also present a general frame-  
85 work to systematically relate cluster- and split-based phylogenetic diversity indices via a process  
86 that is commonly used in phylogenetic combinatorics. This provides concise proofs for general-  
87 izations of previous results for trees as well as ways to define new indices.

88 The rest of this paper is structured as follows. We first illustrate our new concepts and results  
89 by focusing on a few well-known tree-based phylogenetic diversity indices, namely the fair pro-  
90 portion index, the Shapely value (Haake et al., 2008; Shapley, 1953), and the equal splits index  
91 (Redding and Mooers, 2006), before we look into some new split-based phylogenetic diversity  
92 indices. More specifically, in Section 2 we formally define cluster-based phylogenetic diversity  
93 indices and present some key properties that such indices may have. Then, in Section 3, we  
94 present a characterization of the general cluster-based fair proportion index. In Section 4 we  
95 consider the Shapley value,  $SV$  (the values of  $SV$  for the six owl populations considered in Fig-  
96 ure 2(a) are given in Figure 2(b)). In particular, we present a characterization of the Shapley  
97 value and use its relationship to the fair proportion index to describe the first building block of  
98 our framework. In Section 5 we then give the complete framework, and illustrate some of its  
99 applications in Section 6 using the fair proportion index and a split-based phylogenetic diversity  
100 index,  $\Psi_\tau$ , related to the equal splits index as examples. Then, in Section 7, we introduce a  
101 family of new split-based phylogenetic diversity indices,  $\Psi_R$ , that generalize the phylogenetic

102 diversity index for unrooted phylogenetic trees given by Wicke and Steel (2020, Sec. 5.2) (the  
 103 values of the indices  $\Psi_\tau$  and  $\Psi_R$  for the six owl populations considered in Figure 2(a) are also  
 104 given in Figure 2(b)). We conclude in Section 8 discussing some potential interesting directions  
 105 for future work.

106 **2. CLUSTER-BASED INDICES**

107 Let  $X$  be a non-empty finite set. We denote the power set of  $X$  by  $\mathcal{P}(X)$ . We call a non-empty  
 108 subset  $C \subseteq X$  a *cluster* on  $X$  and call a non-empty collection  $\mathcal{C} \subseteq \mathcal{P}(X) \setminus \{\emptyset\}$  a *cluster system*  
 109 on  $X$ . In this section we introduce the concept of a phylogenetic diversity index on a cluster  
 110 system, and illustrate some basic properties of these indices by considering a generalization of  
 111 the fair proportion index for rooted trees that we discussed in the introduction.

112 To motivate the definition of a phylogenetic diversity index on a cluster system, we briefly look  
 113 again at rooted phylogenetic trees. Fixing a rooted phylogenetic tree  $\mathcal{T}$  on a set  $X$  of species, a  
 114 phylogenetic diversity index  $\Phi$  on  $\mathcal{T}$  assigns, to each weighting<sup>3</sup>  $\omega$  of the edges in  $\mathcal{T}$ , a vector  
 115  $\Phi(\omega) \in \mathbb{R}^X$ . To give an example, let  $\Phi$  be the fair proportion index on the rooted phylogenetic  
 116 tree in Figure 1(a). Then, for the weighting  $\omega$  of its edges given in Figure 1(a), we can write

117 (3) 
$$\Phi(\omega) = (5, 4, 7, 3, 6),$$

118 or, in more detail,  $(\Phi(\omega))(a) = 5$ ,  $(\Phi(\omega))(b) = 4, \dots$ ,  $(\Phi(\omega))(e) = 6$ .

119 As described in the introduction, each edge in a rooted phylogenetic tree on  $X$  is associated  
 120 with a cluster on  $X$ . In Figure 3(a) the clusters associated with the edges of the rooted phylo-  
 121 genetic tree in Figure 1(a) are given, where each cluster is weighted by the length of the corre-  
 122 sponding edge. Note that this cluster system  $\mathcal{C}$  has a special property, namely it is a *hierarchy*,  
 123 that is,  $C \cap C' \in \{\emptyset, C, C'\}$  holds for all  $C, C' \in \mathcal{C}$ . In particular, as we see in this example, hi-  
 124 erarchies are essentially those cluster systems that can be represented by a rooted phylogenetic  
 125 tree on  $X$  (see, e.g., Semple and Steel 2003, Thm. 3.5.2 for a more precise statement of this fact  
 126 using the concept of a rooted  $X$ -tree).

127 Bearing these facts in mind, for an arbitrary cluster system  $\mathcal{C}$  on  $X$ , we consider the space  $\mathbb{L}(\mathcal{C})$   
 128 consisting of all weightings  $\omega : \mathcal{C} \rightarrow \mathbb{R}$ . We then define a *phylogenetic diversity index* on  $\mathcal{C}$  to  
 129 be a map  $\Phi : \mathbb{L}(\mathcal{C}) \rightarrow \mathbb{R}^X$ . For example, following the intuitive description in the introduction,  
 130 we define the *fair-proportion index* on a cluster system  $\mathcal{C}$  on  $X$  by putting, for each  $\omega \in \mathbb{L}(\mathcal{C})$   
 131 and all  $x \in X$ ,

132 (4) 
$$(FP(\omega))(x) = \sum_{C \in \mathcal{C}: x \in C} \frac{\omega(C)}{|C|}.$$

133 It can then be checked that (4) applied to the weighted cluster system in Figure 3(a) yields  
 134 precisely the vector we saw in (3).

135 We now introduce three key properties of cluster-based indices which generalize properties  
 136 of tree-based indices described in the literature. We will illustrate these properties for the fair

---

<sup>3</sup>In biological applications, weights are usually assumed to be non-negative. Our framework also applies when this assumption is violated.

(a)	$C$	$\omega(C)$	(b)	$a$	$b$	$c$	$d$	$e$
	$\{a\}$	3		1	0	0	0	0
	$\{b\}$	2		0	1	0	0	0
	$\{c\}$	6		0	0	1	0	0
	$\{d\}$	1		0	0	0	1	0
	$\{e\}$	4		0	0	0	0	1
	$\{a,b\}$	4		$\frac{1}{2}$	$\frac{1}{2}$	0	0	0
	$\{d,e\}$	2		0	0	0	$\frac{1}{2}$	$\frac{1}{2}$
	$\{c,d,e\}$	3		0	0	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

FIGURE 3. (a) The weighted clusters on  $X$  corresponding to the edges of the rooted phylogenetic tree in Figure 1(a). (b) The matrix  $\Gamma$  from Equation (5) for the fair proportion index on  $\mathcal{C}$ , where  $\mathcal{C}$  is the cluster system consisting of the clusters given in (a).

137 proportion index and, as we shall see, these properties are also shared by some of the other  
 138 phylogenetic diversity indices that we consider later on.

139 Let  $\mathcal{C}$  be a cluster system on  $X$ . A phylogenetic diversity index  $\Phi$  on  $\mathcal{C}$  is *additive* if

140 (A)  $\Phi(\omega_1 + \omega_2) = \Phi(\omega_1) + \Phi(\omega_2)$  for all  $\omega_1, \omega_2 \in \mathbb{L}(\mathcal{C})$ ,

141 and  $\Phi$  is *homogeneous* if

142 (H)  $\Phi(a \cdot \omega) = a \cdot \Phi(\omega)$  for all  $\omega \in \mathbb{L}(\mathcal{C})$  and all  $a \in \mathbb{R}$ .

143 Properties (A) and (H) together mean that  $\Phi$  is a *linear map*, in which case we call  $\Phi$  *linear*.  
 144 Phylogenetic diversity indices considered in the literature are usually linear. This may be due to  
 145 useful consequences of linearity such as, for example, that applying a linear phylogenetic diver-  
 146 sity index to a weighting obtained by taking the average over several different edge weightings  
 147 of a fixed rooted phylogenetic tree amounts to averaging the values of the phylogenetic diver-  
 148 sity index. In this paper, most (but not all) results assume linearity of the phylogenetic diversity  
 149 indices involved. To avoid any confusion, we will always explicitly state which properties we  
 150 assume.

151 Note that every linear phylogenetic diversity index  $\Phi$  on  $\mathcal{C}$  corresponds to a  $|\mathcal{C}| \times |X|$ -matrix  
 152  $\Gamma = \Gamma_\Phi = (\gamma_{(C,x)})$  such that

153 (5) 
$$(\Phi(\omega))(x) = \sum_{C \in \mathcal{C}} \omega(C) \cdot \gamma_{(C,x)}$$

154 for all  $\omega \in \mathbb{L}(\mathcal{C})$  and all  $x \in X$ . The entries of the matrix  $\Gamma$  are usually assumed to be non-  
 155 negative (see, e.g., Manson and Steel 2023, Def. 1). Again, our framework also applies when  
 156 this assumption is violated. In Section 7 we will come back to this point. In Figure 3(b) we give,  
 157 as an example, the matrix  $\Gamma$  corresponding to the fair proportion index on the cluster system in  
 158 Figure 3(a).

159 Finally, we call a phylogenetic diversity index  $\Phi$  on  $\mathcal{C}$  *complete* if

160 (C)  $\sum_{x \in X} (\Phi(\omega))(x) = \sum_{C \in \mathcal{C}} \omega(C)$  holds for all  $\omega \in \mathbb{L}(\mathcal{C})$ .

(a)	<table style="border-collapse: collapse; width: 100%;"> <tr> <th style="border-bottom: 1px solid black; padding: 2px 5px;"><math>C</math></th> <th style="border-bottom: 1px solid black; padding: 2px 5px;"><math>\omega(C)</math></th> </tr> <tr> <td style="padding: 2px 5px;"><math>\{a\}</math></td> <td style="padding: 2px 5px;">1</td> </tr> <tr> <td style="padding: 2px 5px;"><math>\{b\}</math></td> <td style="padding: 2px 5px;"><math>\frac{1}{2}</math></td> </tr> <tr> <td style="padding: 2px 5px;"><math>\{c\}</math></td> <td style="padding: 2px 5px;"><math>\frac{1}{2}</math></td> </tr> <tr> <td style="padding: 2px 5px;"><math>\{d\}</math></td> <td style="padding: 2px 5px;">1</td> </tr> <tr> <td style="padding: 2px 5px;"><math>\{a,b\}</math></td> <td style="padding: 2px 5px;">1</td> </tr> <tr> <td style="padding: 2px 5px;"><math>\{b,c\}</math></td> <td style="padding: 2px 5px;">2</td> </tr> <tr> <td style="padding: 2px 5px;"><math>\{b,c,d\}</math></td> <td style="padding: 2px 5px;">3</td> </tr> </table>	$C$	$\omega(C)$	$\{a\}$	1	$\{b\}$	$\frac{1}{2}$	$\{c\}$	$\frac{1}{2}$	$\{d\}$	1	$\{a,b\}$	1	$\{b,c\}$	2	$\{b,c,d\}$	3	(b)	<table style="border-collapse: collapse; width: 100%;"> <tr> <th style="border-bottom: 1px solid black; padding: 2px 5px;"><math>a</math></th> <th style="border-bottom: 1px solid black; padding: 2px 5px;"><math>b</math></th> <th style="border-bottom: 1px solid black; padding: 2px 5px;"><math>c</math></th> <th style="border-bottom: 1px solid black; padding: 2px 5px;"><math>d</math></th> </tr> <tr> <td style="padding: 2px 5px;"><math>\{a\}</math></td> <td style="padding: 2px 5px;">1</td> <td style="padding: 2px 5px;">0</td> <td style="padding: 2px 5px;">0</td> </tr> <tr> <td style="padding: 2px 5px;"><math>\{b\}</math></td> <td style="padding: 2px 5px;">0</td> <td style="padding: 2px 5px;">1</td> <td style="padding: 2px 5px;">0</td> </tr> <tr> <td style="padding: 2px 5px;"><math>\{c\}</math></td> <td style="padding: 2px 5px;">0</td> <td style="padding: 2px 5px;">0</td> <td style="padding: 2px 5px;">1</td> </tr> <tr> <td style="padding: 2px 5px;"><math>\{d\}</math></td> <td style="padding: 2px 5px;">0</td> <td style="padding: 2px 5px;">0</td> <td style="padding: 2px 5px;">0</td> </tr> <tr> <td style="padding: 2px 5px;"><math>\{a,b\}</math></td> <td style="padding: 2px 5px;"><math>\frac{1}{2}</math></td> <td style="padding: 2px 5px;"><math>\frac{1}{2}</math></td> <td style="padding: 2px 5px;">0</td> </tr> <tr> <td style="padding: 2px 5px;"><math>\{b,c\}</math></td> <td style="padding: 2px 5px;">0</td> <td style="padding: 2px 5px;"><math>\frac{1}{2}</math></td> <td style="padding: 2px 5px;"><math>\frac{1}{2}</math></td> </tr> <tr> <td style="padding: 2px 5px;"><math>\{b,c,d\}</math></td> <td style="padding: 2px 5px;">0</td> <td style="padding: 2px 5px;"><math>\frac{1}{3}</math></td> <td style="padding: 2px 5px;"><math>\frac{1}{3}</math></td> </tr> </table>	$a$	$b$	$c$	$d$	$\{a\}$	1	0	0	$\{b\}$	0	1	0	$\{c\}$	0	0	1	$\{d\}$	0	0	0	$\{a,b\}$	$\frac{1}{2}$	$\frac{1}{2}$	0	$\{b,c\}$	0	$\frac{1}{2}$	$\frac{1}{2}$	$\{b,c,d\}$	0	$\frac{1}{3}$	$\frac{1}{3}$
$C$	$\omega(C)$																																																		
$\{a\}$	1																																																		
$\{b\}$	$\frac{1}{2}$																																																		
$\{c\}$	$\frac{1}{2}$																																																		
$\{d\}$	1																																																		
$\{a,b\}$	1																																																		
$\{b,c\}$	2																																																		
$\{b,c,d\}$	3																																																		
$a$	$b$	$c$	$d$																																																
$\{a\}$	1	0	0																																																
$\{b\}$	0	1	0																																																
$\{c\}$	0	0	1																																																
$\{d\}$	0	0	0																																																
$\{a,b\}$	$\frac{1}{2}$	$\frac{1}{2}$	0																																																
$\{b,c\}$	0	$\frac{1}{2}$	$\frac{1}{2}$																																																
$\{b,c,d\}$	0	$\frac{1}{3}$	$\frac{1}{3}$																																																

FIGURE 4. (a) A weighted cluster system  $\mathcal{C}$  on  $X = \{a, b, c, d\}$  that is not a hierarchy. (b) The matrix  $\Gamma$  from Equation (5) for the fair proportion index  $\Phi$  on  $\mathcal{C}$ .

161 For tree-based phylogenetic diversity indices, completeness is often required as part of their  
 162 definition (see, e.g., Bordewich and Sempel 2024; Wicke and Steel 2020). For example, we  
 163 have seen in the introduction for the fair proportion index on a rooted phylogenetic tree that  
 164  $\sum_{x \in X} FP_r(x)$  equals the total weight of the edges in the tree. Property (C) expresses this fact in  
 165 terms of clusters. Note that a linear phylogenetic diversity index  $\Phi$  on  $\mathcal{C}$  is complete if and only  
 166 if  $\sum_{x \in X} \gamma_{(C,x)} = 1$  for all  $C \in \mathcal{C}$  (cf. Wicke and Steel 2020, Eq. (2) and Wicke 2020, Eq. (1)),  
 167 where  $\Gamma = (\gamma_{(C,x)})$  is the matrix from Equation (5).

168 We next show that the fair proportion index satisfies all three of the above properties.

169 **Lemma 2.1.** *The fair proportion index is a complete, linear phylogenetic diversity index on  $\mathcal{C}$*   
 170 *for any cluster system  $\mathcal{C}$  on  $X$ .*

171 *Proof:* As we have seen in the example in Figure 3, the fair proportion index can be described  
 172 by a matrix  $\Gamma = (\gamma_{(C,x)})$  where the row associated with a cluster  $C \in \mathcal{C}$  contains  $|C|$  entries equal  
 173 to  $\frac{1}{|C|}$  and  $|X| - |C|$  entries equal to 0. ■

174  
 175 We conclude this section with an example of a weighted cluster system  $\mathcal{C}$  on  $X = \{a, b, c, d\}$   
 176 that is not a hierarchy and which illustrates the possible consequences of restricting  $\mathcal{C}$  to some  
 177 hierarchy. The cluster system  $\mathcal{C}$  and the weighting  $\omega$  are given in Figure 4(a). Figure 4(b)  
 178 gives the matrix  $\Gamma = \Gamma_\Phi$  corresponding to the fair proportion index  $\Phi$  on  $\mathcal{C}$ . We have  $\Phi(\omega) =$   
 179  $(\frac{3}{2}, 3, \frac{5}{2}, 2)$ . Now consider the hierarchies  $\mathcal{C}_1 = \mathcal{C} \setminus \{\{a, b\}\}$  and  $\mathcal{C}_2 = \mathcal{C} \setminus \{\{b, c\}, \{b, c, d\}\}$  and  
 180 the fair proportion index  $\Phi_i$  on  $\mathcal{C}_i$ ,  $i \in \{1, 2\}$ . Note that the matrix  $\Gamma_{\Phi_i}$  is obtained from  $\Gamma$  by  
 181 removing the rows corresponding to clusters in  $\mathcal{C} \setminus \mathcal{C}_i$  and the weighting  $\omega_i$  is just the restriction  
 182 of  $\omega$  to  $\mathcal{C}_i$ . This yields  $\Phi_1(\omega_1) = (\frac{3}{2}, 1, \frac{1}{2}, 1)$  and  $\Phi_2(\omega_2) = (1, \frac{5}{2}, \frac{5}{2}, 2)$ . As can be seen, the  
 183 rankings of the elements in  $X$  obtained by  $\Phi_1$  and  $\Phi_2$  are different and, thus, need not coincide  
 184 with the ranking obtained by considering the whole cluster system  $\mathcal{C}$ .

### 185 3. A CHARACTERIZATION OF THE FAIR PROPORTION INDEX

186 In general, it is of interest to characterize phylogenetic diversity indices in terms of their key  
 187 properties, as this can help to understand better how they are related to one another. In this  
 188 section, as an illustration for cluster-based indices, we shall present a characterization of the fair

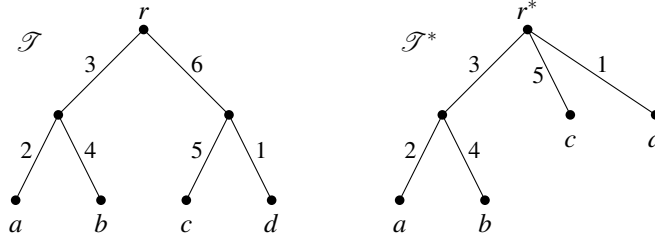


FIGURE 5. Collapsing the edge with weight 6 in the rooted phylogenetic tree  $\mathcal{T}$  on  $X = \{a, b, c, d\}$  yields the rooted phylogenetic tree  $\mathcal{T}^*$  on  $X$ .

189 proportion index. This generalizes the characterization of the fair proportion index on rooted  
 190 phylogenetic trees given by Manson and Steel (2023, Thm. 6).

191 Our characterization will require three properties. The first two properties concern linear phy-  
 192 logenetic diversity indices  $\Phi$  on a cluster system  $\mathcal{C}$  on  $X$ , and are given in terms of the matrix  
 193 corresponding to  $\Phi$ . For all  $C \in \mathcal{C}$ , let  $ch(C)$  denote the set of those  $C' \in \mathcal{C}$  with  $C' \subsetneq C$  such  
 194 that there is no  $C'' \in \mathcal{C}$  with  $C' \subsetneq C'' \subsetneq C$ . We emphasize that even though a cluster in  $\mathcal{C}$   
 195 may receive the weight 0, it is still considered as present in  $\mathcal{C}$  and, therefore, the sets  $ch(C)$  for  $C \in \mathcal{C}$   
 196 do not change when such a weighting is encountered.

197 We say that  $\Phi$  satisfies the *neutrality condition* if

198 (NC) the entries of the matrix  $\Gamma_\Phi$  in Equation (5) are such that  $\gamma_{(C,x)} = \gamma_{(C,y)}$  holds for all  $C \in \mathcal{C}$   
 199 with  $ch(C) = \emptyset$  and all  $x, y \in C$ .

200 A property similar to (NC) was introduced by Manson and Steel (2023) for rooted  $X$ -trees. In  
 201 addition, we say that  $\Phi$  is a *descendant diversity index* if

202 (DD)  $\Phi$  is complete, all entries of the matrix  $\Gamma_\Phi$  in Equation (5) are non-negative and, for all  
 203  $C \in \mathcal{C}$ ,  $\gamma_{(C,x)} = 0$  if  $x \notin C$ .

204 Property (DD) was introduced by Bordewich and Semple (2024) for the special case where the  
 205 cluster system  $\mathcal{C}$  is a hierarchy (using the equivalent description of hierarchies in terms of rooted  
 206  $X$ -trees).

207 The third property is a bit more complicated, and thus we first motivate it using rooted trees  
 208 as in Manson and Steel (2023). Let  $\mathcal{T}$  be a rooted phylogenetic tree on  $X$  with edge weights  
 209 and let  $\mathcal{T}^*$  be the rooted phylogenetic tree on  $X$  obtained by collapsing one of the edges of  $\mathcal{T}$ .  
 210 This is illustrated in Figure 5. In addition, let  $\Phi$  and  $\Phi^*$  be phylogenetic diversity indices on  
 211  $\mathcal{T}$  and  $\mathcal{T}^*$ , respectively. Both  $\Phi$  and  $\Phi^*$  yield a vector in  $\mathbb{R}^X$  for all weightings of the edges  
 212 of  $\mathcal{T}$  and  $\mathcal{T}^*$ , respectively. The topology of the rooted phylogenetic trees, however, may have  
 213 an impact on how the weights of the edges are used to compute these vectors by  $\Phi$  and  $\Phi^*$ ,  
 214 respectively. Therefore, since the topologies of  $\mathcal{T}$  and  $\mathcal{T}^*$  differ, the vector in  $\mathbb{R}^X$  that we obtain  
 215 by  $\Phi^*$  for  $\mathcal{T}^*$  will usually not coincide with the vector that we obtain by  $\Phi$  in the limit, as the  
 216 weight of the edge in  $\mathcal{T}$  tends to 0 (keeping the weights of all other edges in  $\mathcal{T}$  in constant).

217 With this in mind, let  $\mathcal{C}$  be a cluster system on  $X$  and let  $C \in \mathcal{C}$  be such that  $\mathcal{C}^* = \mathcal{C} \setminus \{C\}$   
 218 is non-empty. A phylogenetic diversity index  $\Phi$  on  $\mathcal{C}$  is *downward continuous* with respect to a  
 219 phylogenetic diversity index  $\Phi^*$  on  $\mathcal{C}^*$  if



220 (DC) for all  $\omega \in \mathbb{L}(\mathcal{C})$  we have

221 (6) 
$$\lim_{\omega(C) \rightarrow 0} \Phi(\omega) = \Phi^*(\omega^*),$$

222 where  $\omega^* \in \mathbb{L}(\mathcal{C}^*)$  is the weighting with  $\omega^*(D) = \omega(D)$  for all  $D \in \mathcal{C}^*$ . Note that when forming  
 223 the cluster system  $\mathcal{C}^*$  we remove the cluster  $C$  from the cluster system  $\mathcal{C}$  but not the elements in  
 224  $C$  from the set  $X$ . In particular, both  $\Phi$  and  $\Phi^*$  yield vectors in  $\mathbb{R}^X$ .

225 With the properties (NC), (DD) and (DC) in hand, we now present our characterization of the  
 226 fair proportion index.

227 **Theorem 3.1.** *Suppose we have, for each cluster system  $\mathcal{C}$  on  $X$ , a phylogenetic diversity index*  
 228  *$\Phi_{\mathcal{C}}$  on  $\mathcal{C}$ . Then the following are equivalent:*

- 229 (i) *For all cluster systems  $\mathcal{C}$  on  $X$ ,  $\Phi_{\mathcal{C}}$  is the fair proportion index on  $\mathcal{C}$ .*  
 230 (ii) *For all cluster systems  $\mathcal{C}$  on  $X$ ,  $\Phi_{\mathcal{C}}$  is a descendant diversity index that satisfies the*  
 231 *neutrality condition and is downward continuous with respect to  $\Phi_{\mathcal{C} \setminus \{C\}}$  for all  $C \in \mathcal{C}$*   
 232 *such that  $\mathcal{C} \setminus \{C\} \neq \emptyset$ .*

233 Before proving this theorem, to provide some intuition for its statement, consider the clus-  
 234 ter system  $\mathcal{C} = \{\{a, b\}, \{a, b, c, d\}, \{a, b, c, d, e\}\}$  on  $X = \{a, b, c, d, e\}$ , and let  $\Phi$  be the linear  
 235 phylogenetic diversity index with the following matrix  $\Gamma_{\Phi}$

236

	$a$	$b$	$c$	$d$	$e$
$\{a, b\}$	1/2	1/2	0	0	0
$\{a, b, c, d\}$	1/4	1/4	1/4	1/4	0
$\{a, b, c, d, e\}$	2/5	2/5	1/15	1/15	1/15.

237 In addition, let  $C = \{a, b\}$ ,  $\mathcal{C}^* = \mathcal{C} \setminus \{C\}$ , and let  $\Phi^*$  be the linear phylogenetic diversity index  
 238 on  $\mathcal{C}^*$  whose matrix  $\Gamma_{\Phi^*}$  is obtained by deleting the row corresponding to  $C$  from  $\Gamma_{\Phi}$ . Then both  
 239  $\Phi$  and  $\Phi^*$  satisfy properties (NC) and (DD). Moreover,  $\Phi$  is downward continuous with respect  
 240 to  $\Phi^*$ . But, clearly,  $\Phi$  is not the fair proportion index on  $\mathcal{C}$ . Hence, it is not enough to look at a  
 241 phylogenetic diversity index  $\Phi$  on a cluster system  $\mathcal{C}$  and phylogenetic diversity indices  $\Phi^*$  on  
 242 cluster systems  $\mathcal{C} \setminus \{C\}$  for some  $C \in \mathcal{C}$ . Instead we need to look at *all* cluster systems on  $X$ .

243 *Proof of Theorem 3.1:* We first show that (i) implies (ii). Consider a cluster system  $\mathcal{C}$  on  $X$  and  
 244 put  $\Phi = \Phi_{\mathcal{C}}$ . By assumption,  $\Phi$  is the fair proportion index on  $\mathcal{C}$ . Thus, in view of Lemma 2.1,  
 245  $\Phi$  is linear and complete. Moreover, as illustrated by the example in Figure 3(b), it follows  
 246 immediately from the definition of the fair proportion index in (4) that  $\Phi$  is a descendant diversity  
 247 index and satisfies the neutrality condition.

248 It remains to establish downward continuity. Consider a cluster  $C \in \mathcal{C}$  and assume that  $\mathcal{C}^* =$   
 249  $\mathcal{C} \setminus \{C\} \neq \emptyset$ . Put  $\Phi^* = \Phi_{\mathcal{C}^*}$ . Let  $\Gamma = \Gamma_{\Phi}$  and  $\Gamma^* = \Gamma_{\Phi^*}$  be the matrices whose entries satisfy  
 250 Equation (5) for  $\Phi$  and  $\Phi^*$ , respectively. By assumption,  $\Phi$  is the fair proportion index on  $\mathcal{C}$  and  
 251  $\Phi^*$  is the fair proportion index on  $\mathcal{C}^*$ . Therefore, it follows again from the definition of the fair  
 252 proportion index in (4) that deleting the row corresponding to the cluster  $C$  from the matrix  $\Gamma$   
 253 yields the matrix  $\Gamma^*$ . But this immediately implies that Equation (6) holds for all  $\omega \in \mathbb{L}(\mathcal{C})$ , as  
 254 required.

255 Next we show that (ii) implies (i). Let  $\mathcal{C}$  be a cluster system on  $X$ . By assumption,  $\Phi = \Phi_{\mathcal{C}}$   
 256 is a descendant diversity index and, therefore, linear. Let  $\Gamma = \Gamma_{\Phi}$  be the matrix whose entries

257 satisfy Equation (5) for  $\Phi$ . In view of the definition of the fair proportion index in (4), it suffices  
 258 to show that the entries of  $\Gamma$  satisfy

$$259 \quad \gamma_{(C,x)} = \begin{cases} \frac{1}{|C|} & \text{for } x \in C \\ 0 & \text{for } x \notin C \end{cases}$$

260 for all  $C \in \mathcal{C}$  and all  $x \in X$ . We use induction on  $|\mathcal{C}|$  to show this.

261 To establish the base case of the induction, assume  $|\mathcal{C}| = 1$ . Consider  $C \in \mathcal{C}$  and  $x \in X$ . In view  
 262 of  $|\mathcal{C}| = 1$  we have  $ch(C) = \emptyset$ . Thus, in view of the assumption that  $\Phi$  is a descendant diversity  
 263 index and satisfies the neutrality condition, we have  $\gamma_{(C,x)} = \frac{1}{|C|}$  for all  $x \in C$  and  $\gamma_{(C,x)} = 0$  for all  
 264  $x \in X \setminus C$ , as required.

265 Next assume  $|\mathcal{C}| \geq 2$ . Consider  $C \in \mathcal{C}$  and put  $\mathcal{C}^* = \mathcal{C} \setminus \{C\}$ . By the assumption that  $\Phi$   
 266 is downward continuous with respect to  $\Phi^* = \Phi_{\mathcal{C}^*}$ , the matrix  $\Gamma^* = \Gamma_{\Phi^*}$  whose entries satisfy  
 267 Equation (5) for  $\Phi^*$  is obtained by deleting the row corresponding to cluster  $C$  from  $\Gamma$ . Thus, by  
 268 induction, we have

$$269 \quad \gamma_{(D,x)} = \begin{cases} \frac{1}{|D|} & \text{for } x \in D \\ 0 & \text{for } x \notin D \end{cases}$$

270 for all  $D \in \mathcal{C} \setminus \{C\}$  and all  $x \in X$ . Since this holds for all  $C \in \mathcal{C}$ , this finishes the inductive  
 271 proof.  $\square$

272

#### 4. THE SHAPELY VALUE

273 The Shapely value is a well-known phylogenetic diversity index that can be computed using  
 274 rooted phylogenetic trees and that has its origins in game theory. Interestingly, to understand a  
 275 generalization of this index in the cluster setting, it is necessary to consider mappings on slightly  
 276 more general spaces than those used in the definition of cluster-based phylogenetic diversity  
 277 indices in Section 2. In this section, we shall explain this, and then give a characterization of a  
 278 cluster-based version of the Shapely value.

279 As before, let  $X$  be a finite non-empty set. A *game* is a map  $g : \mathcal{P}(X) \rightarrow \mathbb{R}$ . The elements of  
 280  $X$  are referred to as the *players* in this context and the value  $g(C)$  for some  $C \in \mathcal{P}(X)$  can be  
 281 interpreted as the gain when the players in  $C$  form a coalition. One aspect of analyzing such a  
 282 game is to quantify, for each player  $x \in X$ , the value  $v(x) \in \mathbb{R}$  of the player with respect to the  
 283 game (see, e.g., Branzei et al. 2008 for a more detailed exposition of these concepts).

284 Formally speaking, we are thus interested in maps  $v$  from  $\mathbb{R}^{\mathcal{P}(X)}$  to  $\mathbb{R}^X$ , and the *Shapely value*  
 285 is one specific such map  $v$  given by

$$286 \quad (7) \quad (v(g))(x) = \frac{1}{|X|!} \cdot \sum_{M \in \mathcal{P}(X): x \in M} [(|M| - 1)! \cdot (|X| - |M|)! \cdot (g(M) - g(M \setminus \{x\}))].$$

287 This map was originally proposed by Shapley (1953).

288 In a biological context, the players of Shapley's game are species and from a rooted phylo-  
 289 genetic tree  $\mathcal{T}$  on  $X$  with edge weights we obtain a game  $g$  by setting  $g(M) = PD(M)$  for each  
 290  $M \in \mathcal{P}(X)$ , where  $PD(M)$  is the *phylogenetic diversity* of  $M$ . The value  $PD(M)$  is defined as

291 the total weight of those edges in  $T$  that lie on a path from the root to some species in  $M$  (Faith,  
292 1992). For example, for the rooted phylogenetic tree in Figure 1(a) we obtain

$$293 \quad PD(\{a, b, d\}) = 3 + 2 + 4 + 3 + 2 + 1 = 15.$$

294 We now explain a way to generalize these considerations to cluster systems  $\mathcal{C}$  on  $X$ . First we  
295 need to define the phylogenetic diversity of a subset of  $X$  relative to a weighted cluster system.  
296 Let  $\omega \in \mathbb{L}(\mathcal{C})$ . Then the *phylogenetic diversity* of a subset  $M$  of  $X$  with respect to  $\omega$  is defined  
297 as

$$298 \quad (8) \quad PD(M) = PD_\omega(M) = \sum_{C \in \mathcal{C}: M \cap C \neq \emptyset} \omega(C).$$

299 To further explore properties of the Shapley value in the context of our work, it will be convenient  
300 to consider the set

$$301 \quad \mathbb{PD}(\mathcal{C}) = \{g \in \mathbb{R}^{\mathcal{P}(X)} : \text{there exists } \omega \in \mathbb{L}(\mathcal{C}) \text{ with } g = PD_\omega\},$$

302 that is, the set of games in  $\mathbb{R}^{\mathcal{P}(X)}$  for which there is some  $\omega \in \mathbb{L}(\mathcal{C})$  which gives rise to this  
303 game.

304 The following lemma states two key structural properties of the set  $\mathbb{PD}(\mathcal{C})$  for any cluster  
305 system  $\mathcal{C}$  on  $X$ . To prove this lemma, we define, for all  $C \in \mathcal{P}(X)$ , the game  $g_C : \mathcal{P}(X) \rightarrow \mathbb{R}$   
306 obtained by putting

$$307 \quad (9) \quad g_C(M) = \begin{cases} 1 & \text{if } C \cap M \neq \emptyset \\ 0 & \text{if } C \cap M = \emptyset. \end{cases}$$

308 **Lemma 4.1.** *Let  $\mathcal{C}$  be a cluster system on  $X$ . Then  $\mathbb{PD}(\mathcal{C})$  is a linear subspace of  $\mathbb{R}^{\mathcal{P}(X)}$  that*  
309 *has dimension  $|\mathcal{C}|$ .*

310 *Proof:* In view of (8),  $\mathbb{PD}(\mathcal{C})$  is the linear span of the games  $g_C$  for  $C \in \mathcal{C}$  defined in (9):

$$311 \quad PD_\omega(M) = \sum_{C \in \mathcal{C}: M \cap C \neq \emptyset} \omega(C) = \sum_{C \in \mathcal{C}} \omega(C) \cdot g_C(M)$$

312 Thus, it suffices to show that the games  $g_C$ ,  $C \in \mathcal{C}$ , are linearly independent. To see this, con-  
313 sider the square matrix  $A$  whose rows and columns are each in one-to-one correspondence with  
314 the elements of  $\mathcal{P}(X) \setminus \{\emptyset\}$ . For all  $C, M \in \mathcal{P}(X) \setminus \{\emptyset\}$  the entry of  $A$  in the row corresponding  
315 to  $C$  and the column corresponding to  $M$  is 1 if  $C \cap M \neq \emptyset$  and is 0 otherwise.  $A$  is the so-called  
316 *intersection matrix* of  $\mathcal{P}(X) \setminus \{\emptyset\}$  and it is known that  $A$  has full rank (see, e.g., Jukna 2011, p.  
317 216). Thus, in particular, the rows corresponding to  $C \in \mathcal{C}$  are linearly independent.  $\blacksquare$

318

319 Now, as explained above, for a cluster system  $\mathcal{C}$  on  $X$ , we restrict in (7) to games  $g = PD$   
320 in  $\mathbb{PD}(\mathcal{C})$ . More specifically, we define the Shapley value *relative to the cluster system  $\mathcal{C}$*  as the  
321 map  $SV : \mathbb{PD}(\mathcal{C}) \rightarrow \mathbb{R}^X$  obtained by putting

$$322 \quad (10) \quad (SV(PD))(x) = \frac{1}{|X|!} \cdot \sum_{M \in \mathcal{P}(X): x \in M} [(|M| - 1)! \cdot (|X| - |M|)! \cdot (PD(M) - PD(M \setminus \{x\}))]$$

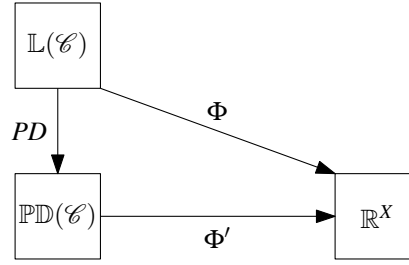


FIGURE 6. This diagram depicts the relationship between a phylogenetic diversity index  $\Phi$  on  $\mathbb{L}(\mathcal{C})$  for a cluster system  $\mathcal{C}$  on  $X$  and a phylogenetic diversity index  $\Phi'$  on  $\mathbb{PD}(\mathcal{C})$  as described by Equation (11).

323 for all  $PD \in \mathbb{PD}(\mathcal{C})$  and all  $x \in X$ . Note that, in view of Lemma 4.1,  $\mathbb{PD}(\mathcal{C})$  may be a proper  
 324 subspace of  $\mathbb{R}^{\mathcal{P}(X)}$  (i.e., the set of all games). As we will see below, any characterization of the  
 325 Shapley value relative to a cluster system must take this into account (see also Dubey 1975 for a  
 326 more general discussion of this aspect).

327 The sharp-eyed reader will have noticed that the Shapley value relative to a cluster system  $\mathcal{C}$   
 328 is *not* a phylogenetic diversity index on  $\mathcal{C}$ , as the latter is defined as a map from  $\mathbb{L}(\mathcal{C})$  to  $\mathbb{R}^X$ .  
 329 However, we can resolve this issue by slightly generalizing our cluster-based definition of phy-  
 330 logenetic diversity indices. Let  $\mathbb{L}$  be a linear subspace of  $\mathbb{R}^{\mathcal{P}(X)}$ . Then we define a phylogenetic  
 331 diversity index *on*  $\mathbb{L}$  to be a map  $\Phi : \mathbb{L} \rightarrow \mathbb{R}^X$ . This encompasses then the Shapley value as a  
 332 phylogenetic diversity index on  $\mathbb{L} = \mathbb{PD}(\mathcal{C})$  for all cluster systems  $\mathcal{C}$  on  $X$ . Moreover, viewing  
 333  $\mathbb{L}(\mathcal{C})$  as the linear subspace

$$334 \quad \mathbb{L} = \{\omega \in \mathbb{R}^{\mathcal{P}(X)} : \omega(C) = 0 \text{ for all } C \notin \mathcal{C}\},$$

335 it also encompasses phylogenetic diversity indices on  $\mathcal{C}$  as defined in Section 2. In fact, we can  
 336 say even more about these relationships, which we will return to in the next section.

337 For the remainder of this section, we focus on giving a characterization of the Shapley value  
 338 relative to a cluster system. This will involve the following two properties. We say that a phylo-  
 339 genetic diversity index  $\Phi$  on a linear subspace  $\mathbb{L}$  of  $\mathbb{R}^{\mathcal{P}(X)}$  satisfies *Pareto efficiency* if

$$340 \quad (\text{PE}) \quad \sum_{x \in X} (\Phi(\omega))(x) = \omega(X) \text{ for all } \omega \in \mathbb{L}.$$

341 **Remark 4.2.** *The properties of completeness and Pareto efficiency are tightly linked. Let  $\mathcal{C}$  be*  
 342 *a cluster system on  $X$  and note that  $\sum_{C \in \mathcal{C}} \omega(C) = PD_{\omega}(X)$  holds for all  $\omega \in \mathbb{L}(\mathcal{C})$ . Therefore,*  
 343 *every complete phylogenetic diversity index  $\Phi$  on  $\mathbb{L}(\mathcal{C})$  corresponds to a phylogenetic diversity*  
 344 *index  $\Phi'$  on  $\mathbb{PD}(\mathcal{C})$  that satisfies Pareto efficiency, where  $\Phi'$  is obtained such that the diagram in*  
 345 *Figure 6 commutes, that is,*

$$346 \quad (11) \quad \Phi'(PD_{\omega}) = \Phi(\omega)$$

347 *for all  $\omega \in \mathbb{L}(\mathcal{C})$ .*

348 We say that a phylogenetic diversity index  $\Phi$  on a linear subspace  $\mathbb{L}$  of  $\mathbb{R}^{\mathcal{P}(X)}$  satisfies *group*  
 349 *proportionality* (cf. Haake et al. 2008) if

$$350 \quad (\text{GP}) \quad (\Phi(a \cdot g_C))(x) = \begin{cases} \frac{a}{|C|} & \text{if } x \in C \\ 0 & \text{if } x \notin C, \end{cases} \quad \text{for all } C \in \mathcal{P}(X) \setminus \{\emptyset\} \text{ and all } a \in \mathbb{R}$$

351 with  $g_C$  the game as defined in (9). Note that a similar characterization to that given in the  
 352 following theorem was established by Wicke and Steel (2020, Thm. 7) for the special case of  
 353 cluster systems that form a hierarchy.

354 **Theorem 4.3.** *Let  $\mathcal{C}$  be a cluster system on  $X$ . The Shapley value is the unique phylogenetic di-*  
 355 *versity index on  $\mathbb{PD}(\mathcal{C})$  that is additive and satisfies Pareto efficiency and group proportionality.*

356 *Proof:* Assume that  $\Phi'$  is the Shapley value on  $\mathbb{PD}(\mathcal{C})$ . It is known (see, e.g., Aumann 1994)  
 357 that  $\Phi'$  satisfies Pareto efficiency for all  $\omega \in \mathbb{R}^{\mathcal{P}(X)}$  and is additive for all  $\omega_1, \omega_2 \in \mathbb{R}^{\mathcal{P}(X)}$ . Thus,  
 358 these two properties hold, in particular, for all  $\omega, \omega_1, \omega_2 \in \mathbb{PD}(\mathcal{C}) \subseteq \mathbb{R}^{\mathcal{P}(X)}$ .

359 To establish that  $\Phi'$  also satisfies group proportionality, consider  $x \in X$ ,  $C \in \mathcal{C}$  and  $a \in \mathbb{R}$ . We  
 360 calculate the value  $(\Phi'(a \cdot g_C))(x)$  using Formula (10) (similar calculations are used in the proofs  
 361 of Haake et al. 2008, Thm. 4 and Coronado et al. 2018, Thm. 1):

362 If  $x \notin C$  we have  $g_C(M) - g_C(M \setminus \{x\}) = 0$  for all  $M \in \mathcal{P}(X)$ , implying  $(\Phi'(a \cdot g_C))(x) = 0$ , as  
 363 required. So assume that  $x \in C$ , put  $c = |C|$ ,  $m = |M|$ , and put  $n = |X|$ . Then, in view of the fact  
 364 that only  $M \in \mathcal{P}(X)$  with  $M \cap C = \{x\}$  contribute to  $(\Phi'(a \cdot g_C))(x)$ , we have

$$365 \quad (\Phi'(a \cdot g_C))(x) = \frac{a}{n!} \cdot \sum_{m=1}^{n-c+1} (m-1)! \cdot (n-m)! \cdot \binom{n-c}{m-1} = \frac{a \cdot (n-c)! \cdot (c-1)!}{n!} \cdot \sum_{j=c-1}^{n-1} \binom{j}{c-1}$$

$$366 \quad = \frac{a \cdot (n-c)! \cdot (c-1)!}{n!} \cdot \binom{n}{c} = \frac{a}{c},$$

367 as required, where we used the formula for the sum along a diagonal in Pascal's triangle to obtain  
 368 the first equality in the second line.

369 Uniqueness now follows from the fact that, in view of the proof of Lemma 4.1,  $\mathbb{PD}(\mathcal{C})$  is the  
 370 linear span of  $\{g_C : C \in \mathcal{C}\}$ . ■

371

372 Interestingly, as shown by Fuchs and Jin (2015), the vector in  $\mathbb{R}^X$  that results from computing  
 373 the Shapley value on the game  $PD$  obtained from an edge-weighted rooted phylogenetic tree  
 374 always coincides with the vector that we obtain by computing the fair proportion index on the  
 375 rooted phylogenetic tree. In fact, this is a particular instance of (11). The following Corollary of  
 376 Theorem 4.3 makes this more precise.

377 **Corollary 4.4.** *Let  $\mathcal{C}$  be a cluster system on  $X$ ,  $\Phi$  be the fair proportion index on  $\mathbb{L}(\mathcal{C})$ , and  $\Phi'$*   
 378 *be the Shapley value on  $\mathbb{PD}(\mathcal{C})$ . Then*

$$379 \quad \Phi(\omega) = \Phi' \left( \sum_{C \in \mathcal{C}} \omega(C) \cdot g_C \right) = \Phi'(PD_\omega)$$

380 *holds for all  $\omega \in \mathbb{L}(\mathcal{C})$ .*

381 *Proof:* This follows immediately from the definition of the fair proportion index together with  
 382 the fact that, by Theorem 4.3, the Shapley value is additive and satisfies group proportionality. ■

383

384 It is remarked in the discussion by Coronado et al. (2018) that Corollary 4.4 can also be derived  
 385 using arguments based on so-called phylogenetic networks (for more on the connection between  
 386 such networks and diversity indices see Section 8). Moreover, the fact that the Shapley value on  
 387  $\mathbb{PD}(\mathcal{C})$  satisfies Pareto efficiency means that it apportions the phylogenetic diversity of  $X$  among  
 388 the elements of  $X$ . In view of Corollary 4.4 this then also holds for the fair proportion index  
 389 on  $\mathbb{L}(\mathcal{C})$  and, in view of Remark 4.2, this corresponds to the fact that the fair proportion index is  
 390 complete, as can be seen in the example in Figure 3(a):

$$391 \quad PD_{\omega}(X) = \sum_{C \in \mathcal{C}} \omega(C) = 25 = \sum_{x \in X} (FP(\omega))(x).$$

## 392 5. AN AFFINE AND PROJECTIVE FRAMEWORK FOR PHYLOGENETIC DIVERSITY INDICES

393 As mentioned in the introduction, the notion of phylogenetic diversity indices has also been  
 394 considered on unrooted phylogenetic trees (Haake et al., 2008; Wicke and Steel, 2020) and, just  
 395 as rooted phylogenetic trees can be encoded by a collection of clusters, unrooted phylogenetic  
 396 trees on a set  $X$  of species can be encoded by a collection  $\mathcal{S}$  of bipartitions, or splits, of  $X$   
 397 (see, e.g., Steel 2016, Ch. 2). In the area of phylogenetic combinatorics, the interplay between  
 398 collections of clusters and collections of splits has been studied in terms of affine and projective  
 399 models of clustering, respectively, in analogy with the interplay between affine and projective  
 400 geometry in classical geometry (Dress 2012, p. 207; see also Dress 1997). One of the key ideas  
 401 that we will exploit from this theory is that we can map a collection  $\mathcal{S}$  of splits of  $X$  in a natural  
 402 way to a cluster system  $\mathcal{C}(\mathcal{S})$  on  $X$  (defined in (13) below) and, in this way, derive split-based  
 403 indices from cluster-based indices. In this section, we will make this more precise, and illustrate  
 404 the resulting framework using the fair proportion index and the Shapely value as examples.

405 First, we formally define the concepts mentioned above. A *split*  $S$  of  $X$  is a bipartition of  $X$   
 406 into two non-empty subsets  $A$  and  $B$ , that is,  $A \cup B = X$  and  $A \cap B = \emptyset$ . We denote such a split as  
 407 an unordered pair  $A|B = B|A$ . A *split system*  $\mathcal{S}$  on  $X$  is a non-empty set of splits of  $X$ . By  $\mathcal{S}(X)$   
 408 we denote the set of all splits of  $X$  and, for a split system  $\mathcal{S} \subseteq \mathcal{S}(X)$ , we denote by  $\mathbb{L}(\mathcal{S})$  the  
 409 set of all weightings  $\lambda : \mathcal{S}(X) \rightarrow \mathbb{R}$  with  $\lambda(S) = 0$  for all  $S \in \mathcal{S}(X) \setminus \mathcal{S}$ . In addition, we denote  
 410 by  $\mathbb{PD}(\mathcal{S})$  the set of all weightings  $PD : \mathcal{S}(X) \rightarrow \mathbb{R}$  that can be written as

$$411 \quad (12) \quad PD(M) = PD_{\lambda}(M) = \sum_{A|B \in \mathcal{S} : A \cap M \neq \emptyset, B \cap M \neq \emptyset} \lambda(A|B)$$

412 for some  $\lambda \in \mathbb{L}(\mathcal{S})$ . The value  $PD_{\lambda}(M)$  is usually called the *phylogenetic diversity* of  $M$  with  
 413 respect to the weighting  $\lambda$  of the splits in  $\mathcal{S}$  (see, e.g., Spillner et al. 2008).

414 Figure 7 gives an overview of the various spaces we shall consider and the maps between  
 415 them. In addition to the maps already introduced in Figure 6 in Section 4, we also consider, for  
 416 split systems  $\mathcal{S}$  on  $X$ , maps  $\tau$  from  $\mathbb{L}(\mathcal{S})$  to  $\mathbb{L}(\mathcal{C})$  where  $\mathcal{C}$  is the cluster system

$$417 \quad (13) \quad \mathcal{C}(\mathcal{S}) = \bigcup_{S \in \mathcal{S}} S$$

418 on  $X$  mentioned above. In particular, we are interested in maps  $\tau$  for which various parts of the  
 419 diagram in Figure 7 commute.

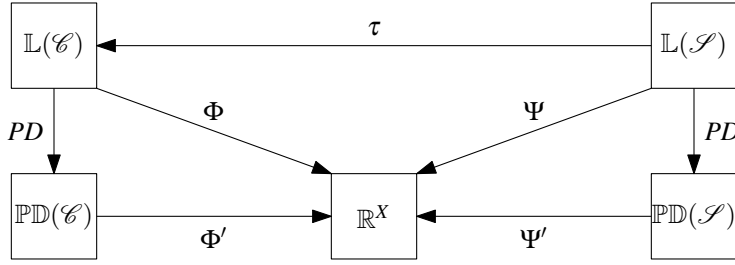


FIGURE 7. A diagram of the various maps we consider to study relationships between phylogenetic diversity indices. The left part of the diagram we have already seen in Figure 6. In analogy to this, the right part of the diagram depicts phylogenetic diversity indices  $\Psi$  and  $\Psi'$  on  $\mathbb{L}(\mathcal{S})$  and  $\mathbb{PD}(\mathcal{S})$ , respectively, where  $\mathcal{S}$  is a split system on  $X$ . Finally  $\tau$  associates with each weighting  $\lambda$  of the splits in  $\mathcal{S}$  a weighting  $\omega = \tau(\lambda)$  of the clusters in a cluster system  $\mathcal{C} = \mathcal{C}(\mathcal{S})$  that arises from  $\mathcal{S}$  by (13).

(a)	$S$	$\lambda(S)$	(b)	$C$	$\omega(C)$
	$\{a, b\}   \{c, d, e\}$	10		$\{c\}$	16
	$\{a, c\}   \{b, d, e\}$	5		$\{a, b\}$	6
	$\{c\}   \{a, b, d, e\}$	20		$\{a, c\}$	3
				$\{c, d, e\}$	4
				$\{b, d, e\}$	2
				$\{a, b, d, e\}$	4

FIGURE 8. (a) A split system  $\mathcal{S}$  on  $X = \{a, b, c, d, e\}$  with weighting  $\lambda$ . (b) The associated cluster system  $\mathcal{C}(\mathcal{S})$  on  $X$  as defined in (13) and the weighting  $\omega = \tau(\lambda)$  as defined in (15).

420 As an illustration of this setup, we now revisit the relationship between the fair proportion  
 421 index and the Shapely value. Let  $\mathcal{S}$  be a split system on  $X$ . Then the *Shapely value* on  $\mathbb{PD}(\mathcal{S})$   
 422 is defined as in (10). Equivalently, as shown by Haake et al. (2008) for trees and by Volkmann  
 423 et al. (2014) for split systems in general, the Shapely value on  $\mathbb{PD}(\mathcal{S})$  can also be computed as

$$424 \quad (14) \quad (SV(PD_\lambda))(x) = \sum_{A|B \in \mathcal{S}: x \in A} \frac{|B|}{|X| \cdot |A|} \cdot \lambda(A|B).$$

425 for all  $\lambda \in \mathbb{L}(\mathcal{S})$  and all  $x \in X$ .

426 Now consider the map  $\tau : \mathbb{L}(\mathcal{S}) \rightarrow \mathbb{L}(\mathcal{C}(\mathcal{S}))$  defined by putting, for  $\lambda \in \mathbb{L}(\mathcal{S})$ ,

$$427 \quad (15) \quad (\tau(\lambda))(A) = \frac{|B|}{|X|} \cdot \lambda(A|B) \quad \text{and} \quad (\tau(\lambda))(B) = \frac{|A|}{|X|} \cdot \lambda(A|B)$$

428 for all  $A, B \in \mathcal{C}(\mathcal{S})$  such that  $A|B$  is a split in  $\mathcal{S}$ . For example, consider the split system  $\mathcal{S}$   
 429 with weighting  $\lambda$  in Figure 8(a). Using Formula (14), we obtain  $SV(a) = \frac{11}{2}$  in this example and  
 430 we also have  $FP(a) = \frac{11}{2}$  for the fair proportion index as defined in (4) applied to the cluster

431 system  $\mathcal{C}(\mathcal{S})$  with weighting  $\omega = \tau(\lambda)$ . We conclude this section by showing that this is not a  
432 coincidence.

433 **Theorem 5.1.** *Let  $\mathcal{S}$  be a split system on  $X$ ,  $\Phi$  be the fair proportion index on  $\mathbb{L}(\mathcal{C}(\mathcal{S}))$  and*  
434  *$\Psi'$  be the Shapley value on  $\mathbb{PD}(\mathcal{S})$ . If  $\tau$  is as defined in (15), then*

$$435 \quad (16) \quad \Phi(\tau(\lambda)) = \Psi'(PD_\lambda)$$

436 for all  $\lambda \in \mathbb{L}(\mathcal{S})$ .

*Proof:* Let  $\lambda \in \mathbb{L}(\mathcal{S})$  and put  $\omega = \tau(\lambda)$ . Since the maps  $\Phi$ ,  $\tau$ ,  $\Psi'$  and  $PD$  are all linear, it suffices to show Equation (16) for the case that one element of  $\mathcal{S}$ , say  $S = A|B$  has weight 1 (i.e.  $\lambda(A|B) = 1$ ), whereas  $\lambda(S') = 0$  for all  $S' \neq S$ . Then we have  $\omega(A) = |B|/|X|$ ,  $\omega(B) = |A|/|X|$ , and  $\omega(C) = 0$  for all  $C \in \mathcal{C}(\mathcal{S})$  with  $C \neq A, B$ . Now let  $x \in X$ , and assume without loss of generality that  $x \in A$ . Then,

$$(\Phi(\omega))(x) = \sum_{C \in \mathcal{C}(\mathcal{S}): x \in C} \frac{\omega(C)}{|C|} = \frac{\omega(A)}{|A|} = \frac{\frac{|B|}{|X|}}{|A|} = \frac{|B|}{|X| \cdot |A|}.$$

On the other hand, in view of (14) we have

$$(\Psi'(PD_\lambda))(x) = \sum_{A'|B' \in \mathcal{S}: x \in A'} \frac{|B'|}{|X| \cdot |A'|} \lambda(S) = \frac{|B|}{|X| \cdot |A|}$$

437 as well. This completes the proof. ■

438

439

## 6. COMPLETE DIVERSITY INDICES

440 In this section we shall consider Figure 7 once again, considering an alternative definition for  
441 the map  $\tau$  that can be used to translate, for any split system  $\mathcal{S}$  on  $X$ , the property of complete-  
442 ness from a cluster-based index  $\Phi$  on  $\mathbb{L}(\mathcal{C}(\mathcal{S}))$  to an associated split-based index  $\Psi = \Psi_\tau(\Phi)$   
443 on  $\mathbb{L}(\mathcal{S})$ . In particular, we will see that this immediately implies the completeness of the fair  
444 proportion index on unrooted phylogenetic trees that was established by Wicke and Steel (2020)  
445 (for example, see Figure 1(b) in the introduction). In addition, we illustrate the application of  
446 these considerations to a generalization of the so-called equal splits index that appears in Wicke  
447 and Steel (2020).

448 We begin by proving a result concerning completeness. Let  $\mathcal{S}$  be a split system on  $X$ . A  
449 phylogenetic diversity index  $\Psi$  on  $\mathbb{L}(\mathcal{S})$  is *complete* if

$$450 \quad (C') \quad \sum_{x \in X} (\Psi(\lambda))(x) = \sum_{S \in \mathcal{S}} \lambda(S) \text{ holds for all } \lambda \in \mathbb{L}(\mathcal{S}).$$

451 Define the map  $\tau : \mathbb{L}(\mathcal{S}) \rightarrow \mathbb{L}(\mathcal{C}(\mathcal{S}))$  by putting

$$452 \quad (17) \quad (\tau(\lambda))(C) = \frac{1}{2} \cdot \lambda(C|(X-C))$$

453 for all  $C \in \mathcal{C}(\mathcal{S})$ . The basic idea is to distribute the weight  $\lambda(S)$  of a split  $S = A|B \in \mathcal{S}$   
454 evenly on the two corresponding clusters  $A, B \in \mathcal{C}(\mathcal{S})$ . The following results, however, also



455 hold if the weight is distributed non-evenly, that is, when putting  $(\tau(\lambda))(A) = p \cdot \lambda(A|B)$  and  
 456  $(\tau(\lambda))(B) = (1 - p) \cdot \lambda(A|B)$  for some  $0 < p < 1$  with  $p \neq \frac{1}{2}$ .

457 With the map  $\tau$  defined in (17), we obtain, for a phylogenetic diversity index  $\Phi$  on  $\mathbb{L}(\mathcal{C}(\mathcal{S}))$ ,  
 458 the phylogenetic diversity index  $\Psi = \Psi_\tau(\Phi)$  on  $\mathbb{L}(\mathcal{S})$  by putting  $\Psi(\lambda) = \Phi(\tau(\lambda))$  for all  $\lambda \in$   
 459  $\mathbb{L}(\mathcal{S})$ .

460 **Theorem 6.1.** *Let  $\mathcal{S}$  be a split system on  $X$  and  $\Phi$  a complete linear phylogenetic diversity index*  
 461 *on  $\mathbb{L}(\mathcal{C}(\mathcal{S}))$ . If  $\tau$  is as defined in (17), then  $\Psi_\tau(\Phi)$  is a complete linear phylogenetic diversity*  
 462 *index on  $\mathbb{L}(\mathcal{S})$ .*

463 *Proof:* Let  $\Phi$  be a complete linear phylogenetic diversity index on  $\mathbb{L}(\mathcal{C}(\mathcal{S}))$ . We first show that  
 464  $\Psi_\tau$  is linear. Let  $\lambda_1, \lambda_2 \in \mathbb{L}(\mathcal{S})$  and  $a \in \mathbb{R}$ . Then, noting that  $\tau$  is linear, we have

$$465 \quad (\Psi_\tau(\Phi))(a \cdot \lambda_1 + \lambda_2) = \Phi(\tau(a \cdot \lambda_1 + \lambda_2)) = \Phi(a \cdot \tau(\lambda_1) + \tau(\lambda_2))$$

$$466 \quad \quad \quad = a \cdot \Phi(\tau(\lambda_1)) + \Phi(\tau(\lambda_2)) = a \cdot (\Psi_\tau(\Phi))(\lambda_1) + (\Psi_\tau(\Phi))(\lambda_2),$$

467 as required.

468 It remains to show that  $\Psi_\tau$  is complete. Let  $\lambda \in \mathbb{L}(\mathcal{S})$ . Then we have

$$469 \quad \sum_{x \in X} ((\Psi_\tau(\Phi))(\lambda))(x) = \sum_{x \in X} (\Phi(\tau(\lambda)))(x) = \sum_{C \in \mathcal{C}(\mathcal{S})} (\tau(\lambda))(C)$$

$$470 \quad \quad \quad = \sum_{C \in \mathcal{C}(\mathcal{S})} \frac{1}{2} \cdot \lambda(C|X - C) = \sum_{S \in \mathcal{S}} \lambda(S),$$

471 as required. ■

472

473 The following Corollary 6.2 includes, as a special case, the completeness of the fair proportion  
 474 index on unrooted phylogenetic trees that was established by Wicke and Steel (2020, Thm. 10).  
 475 To see this, it suffices to consider, for an unrooted phylogenetic tree on  $X$ , the split system  
 476 consisting of those splits of  $X$  that can be obtained by removing an edge from the tree.

477 **Corollary 6.2.** *Let  $\mathcal{S}$  be a split system on  $X$  and  $\Phi$  be the fair proportion index on  $\mathbb{L}(\mathcal{C}(\mathcal{S}))$ .*  
 478 *If  $\tau$  is as defined in (17), then  $\Psi_\tau(\Phi)$  is a complete linear phylogenetic diversity index on  $\mathbb{L}(\mathcal{S})$*   
 479 *and we have*

$$480 \quad (18) \quad ((\Psi_\tau(\Phi))(\lambda))(x) = \sum_{A|B \in \mathcal{S}: x \in A} \frac{\lambda(S)}{2 \cdot |A|}$$

481 for all  $\lambda \in \mathbb{L}(\mathcal{S})$  and all  $x \in X$ .

482 *Proof:* In view of Lemma 2.1, Theorem 6.1 implies that  $\Psi_\tau(\Phi)$  is a complete linear phylogenetic  
 483 diversity index on  $\mathbb{L}(\mathcal{S})$ . Moreover, (18) follows from (4), (13), and (17). ■

484

485 We now turn our attention to a generalization of the *equal splits* index, a phylogenetic diversity  
 486 index that was introduced in the setting of rooted phylogenetic trees by Redding and Mooers  
 487 (2006). We first define our generalization for cluster systems  $\mathcal{C}$  on  $X$ . For all  $C \in \mathcal{C}$ , let  $cl(C)$

(a)	$C$	$\omega(C)$	$m_a(C)$	$m_b(C)$	$m_c(C)$	$m_d(C)$	$m_e(C)$	(b)	$x$	$ES(x)$
	$\{a\}$	2	1	0	0	0	0		$a$	4
	$\{b\}$	3	0	1	0	0	0		$b$	5
	$\{c\}$	1	0	0	1	0	0		$c$	7
	$\{d\}$	4	0	0	0	1	0		$d$	8
	$\{a, b, c\}$	5	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	0	0		$e$	1
	$\{c, d\}$	7	0	0	$\frac{1}{2}$	$\frac{1}{2}$	0			
	$X$	3	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{5}{18}$	$\frac{1}{6}$	$\frac{1}{3}$			

FIGURE 9. (a) A cluster system  $\mathcal{C}$  on  $X = \{a, b, c, d, e\}$  with weighting  $\omega$  and the quantities  $m_x(C)$  as defined for all  $x \in X$  and  $C \in \mathcal{C}$  in (19). (b) The equal splits index  $ES(x)$  for all  $x \in X$  obtained from  $\mathcal{C}$  and  $\omega$  by (20).

488 denote the set of those  $x \in C$  that are not contained in any cluster  $C' \in ch(C)$ . Then put

$$489 \quad (19) \quad m_x(C) = \begin{cases} 0 & \text{if } x \notin C, \\ \frac{1}{|ch(C)| + |cl(C)|} & \text{if } x \in cl(C), \\ \sum_{C' \in ch(C)} \frac{m_x(C')}{|ch(C)| + |cl(C)|} & \text{otherwise,} \end{cases}$$

490 for all  $x \in X$  and all  $C \in \mathcal{C}$ . Note that  $m_x(C) = 1/|C|$  if  $x \in C$  and  $ch(C) = \emptyset$  (as in this case  
491  $|ch(C)| = 0$  and  $|cl(C)| = |C|$ ). Also note that  $m_x(C)$  is defined recursively. In particular, when  
492 computing  $m_x(C)$  in the third case it is assumed that  $m_x(C')$  for each  $C' \in \mathcal{C}$  with  $C' \subsetneq C$  has  
493 been computed already. The equal splits index is then defined by putting

$$494 \quad (20) \quad (ES(\omega))(x) = \sum_{C \in \mathcal{C}} m_x(C) \cdot \omega(C)$$

495 for all  $\omega \in \mathbb{L}(\mathcal{C})$  and all  $x \in X$ . As an example, consider the cluster system  $\mathcal{C}$  with weighting  $\omega$   
496 in Figure 9(a). For the cluster  $C = X$  we have  $ch(X) = \{\{a, b, c\}, \{c, d\}\}$  and  $cl(X) = \{e\}$ , which  
497 yields, by (19),  $m_e(X) = \frac{1}{3}$ . The resulting values of the equal splits index are given in Figure 9(b).

498 The equal splits index on  $\mathbb{L}(\mathcal{C})$  is linear with the corresponding  $|\mathcal{C}| \times |X|$ -matrix  $\Gamma$  in Equa-  
499 tion (5) having the entries  $\gamma_{(C,x)} = m_x(C)$ . Moreover, as can be seen in the example in Figure 9(a),  
500 the sum of the entries in each row of  $\Gamma$  equals 1. The next theorem establishes that this is always  
501 the case.

502 **Theorem 6.3.** *For all cluster systems  $\mathcal{C}$  on  $X$  the equal splits index is a complete linear phylo-*  
503 *genetic diversity index on  $\mathbb{L}(\mathcal{C})$ .*

504 *Proof:* Let  $\mathcal{C}$  be a cluster system on  $X$ . We already noted above that the equal splits index is  
505 linear. Thus, it remains to establish that the equal splits index is complete. More specifically, it  
506 suffices to show that

$$507 \quad (21) \quad \sum_{x \in X} m_x(C) = 1$$

508 for all  $C \in \mathcal{C}$ . To this end, put  $desc(C) = |\bigcup_{C' \in ch(C)} C'|$  for all  $C \in \mathcal{C}$ , that is,  $desc(C)$  equals the  
 509 total number of elements in  $X$  contained in the clusters in  $ch(C)$ . We show (21) by induction on  
 510  $desc(C)$ .

511 For the base case  $desc(C) = 0$  we have  $ch(C) = \emptyset$  and thus  $cl(C) = C$ , implying  $|cl(C)| = |C|$   
 512 and

$$513 \quad \sum_{x \in X} m_x(C) = \sum_{x \in C} m_x(C) + \sum_{x \in X \setminus C} m_x(C) = |C| \cdot \frac{1}{|C|} + 0 = 1,$$

514 as required.

515 Next assume  $desc(C) > 0$ . By the definition of  $ch(C)$ , we have  $desc(C') < desc(C)$  for all  
 516  $C' \in ch(C)$ . Thus,

$$\begin{aligned} 517 \quad \sum_{x \in X} m_x(C) &= \sum_{x \in X \setminus C} m_x(C) + \sum_{x \in cl(C)} m_x(C) + \sum_{x \in C \setminus cl(C)} m_x(C) \\ 518 \quad &= 0 + \frac{|cl(C)|}{|ch(C)| + |cl(C)|} + \sum_{x \in C \setminus cl(C)} \sum_{C' \in ch(C)} \frac{m_x(C')}{|ch(C)| + |cl(C)|} \\ 519 \quad &= \frac{|cl(C)|}{|ch(C)| + |cl(C)|} + \sum_{C' \in ch(C)} \sum_{x \in C \setminus cl(C)} \frac{m_x(C')}{|ch(C)| + |cl(C)|} \\ 520 \quad &= \frac{|cl(C)|}{|ch(C)| + |cl(C)|} + \sum_{C' \in ch(C)} \sum_{x \in X} \frac{m_x(C')}{|ch(C)| + |cl(C)|} \\ 521 \quad &= \frac{|cl(C)|}{|ch(C)| + |cl(C)|} + |ch(C)| \cdot \frac{1}{|ch(C)| + |cl(C)|} \\ 522 \quad &= 1, \end{aligned}$$

523 where the equality in the fourth line holds in view of the fact that  $m_x(C') = 0$  for all  $x \in$   
 524  $X \setminus (C \setminus cl(C))$  and for all  $C' \in ch(C)$ , and the equality in the fifth line holds by induction. ■

525

526 Our final result in this section, which is an immediate consequence of Theorem 6.1 and Theorem  
 527 6.3, summarizes how we obtain, via the map  $\tau$  defined in (17), a complete linear split-based  
 528 phylogenetic diversity index from the cluster-based equal splits index.

529 **Corollary 6.4.** *Let  $\mathcal{S}$  be a split system on  $X$  and  $\Phi$  be the equal splits index on  $\mathbb{L}(\mathcal{C}(\mathcal{S}))$ . If  $\tau$*   
 530 *is as defined in (17), then  $\Psi_\tau(\Phi)$  is a complete linear phylogenetic diversity index on  $\mathbb{L}(\mathcal{S})$ .*

531 We conclude this section coming back to the biological example in Figure 2(a) and compute  
 532 the phylogenetic diversity index  $\Psi_\tau = \Psi_\tau(\Phi)$  from Corollary 6.4 for this example. From the  
 533 split system  $\mathcal{S}$  on  $X = \{a, f, h, m, r, s\}$  with weighting  $\lambda$  given in Figure 12 in the appendix we  
 534 first compute the cluster system  $\mathcal{C} = \mathcal{C}(\mathcal{S})$  on  $X$  with weighting  $\omega = \tau(\lambda)$  (in Figure 13 in the  
 535 appendix we present the Hasse diagram for the 20 clusters in  $\mathcal{C}$ , where the weight  $\omega(C)$  obtained  
 536 by (17) is given below each cluster  $C$  in the diagram). Then we compute the matrix  $\Gamma = \Gamma_\Phi$  for  
 537 the equal splits index  $\Phi$  on  $\mathbb{L}(\mathcal{C})$  (see Figure 14 in the appendix) from which we obtain the values  
 538 of the phylogenetic diversity index  $\Psi_\tau$  given in Figure 2(b). For comparison purposes, we also

539 compute the Shapley value  $SV$  as defined in (14) for the split system  $\mathcal{S}$  on  $X$  with weighting  $\lambda$   
 540 using the program SplitsTreeCE (Huson and Bryant, 2005).

541 As can be seen in Figure 2(b), the sum of the values of the index  $\Psi_\tau$  yields the total weight 1.980  
 542 of all splits in  $\mathcal{S}$ , as it should be for a complete phylogenetic diversity index. The ranking of the  
 543 six populations given by  $SV$  is the same as the ranking given by  $FP_u$  computed in the introduc-  
 544 tion. The ranking given by  $\Psi_\tau$  slightly deviates from it but also ranks populations  $s$  and  $m$  at the  
 545 top and population  $f$  at the bottom.

546 The fact that the network in Figure 2(a) is not a tree implies that using a tree-based phyloge-  
 547 netic diversity index necessarily involves a (potentially arbitrary) decision which of the splits in  
 548  $\mathcal{S}$  are used to compute the tree-based index. More formally, we would first need to restrict to  
 549 some subset  $\mathcal{S}' \subseteq \mathcal{S}$  such that any two splits  $A|B, C|D \in \mathcal{S}'$  are *compatible*, that is, at least one  
 550 of the intersections  $A \cap C$ ,  $A \cap D$ ,  $B \cap C$ , and  $B \cap D$  is empty. A collection of pairwise compati-  
 551 ble splits is called a *compatible* split system. Intuitively, compatible split systems correspond to  
 552 unrooted phylogenetic trees.

553 To illustrate that the choice of a compatible subset of  $\mathcal{S}$  really has an impact on the ranking  
 554 of the six owl populations, we consider two compatible subsets  $\mathcal{S}'_1$  and  $\mathcal{S}'_2$  that are maximal  
 555 with respect to set inclusion. Using the index  $\Psi_\tau$ , we obtain the ranking (from highest to lowest)  
 556  $s, m, h, r, a, f$  based on  $\mathcal{S}'_1$  and the ranking  $s, m, a, h, r, f$  based on  $\mathcal{S}'_2$  (for details see Figures 15  
 557 and 16 in the appendix). Clearly, these two rankings are different, and in fact they also differ from  
 558 the ranking obtained when considering all splits in  $\mathcal{S}$  (cf. 2(b)). Interestingly, in this example,  
 559 the ranking given in Figure 2(b) for  $FP_u$  and  $SV$  does not change when restricting to *any* maximal  
 560 compatible subset of  $\mathcal{S}$ . This could be due to the fact that the *trivial* splits in  $\mathcal{S}$  (i.e. splits  $A|B$   
 561 with  $|A| = 1$  or  $|B| = 1$ ) carry more weight than the *non-trivial* splits, and both  $FP_u$  and  $SV$  are  
 562 less heavily influenced by the non-trivial splits than  $\Psi_\tau$ . In future work, it could be interesting to  
 563 further investigate the differences in rankings obtained from these and other split-based diversity  
 564 indices.

## 565 7. A GENERALIZATION OF THE PAUPLIN INDEX

566 In (Wicke and Steel, 2020, Sec. 5.2) a phylogenetic diversity index for unrooted phylogenetic  
 567 trees is introduced that is related to the formula for the total weight of the edges given by Pauplin  
 568 (2000). In this section, we describe how the viewpoint suggested by our framework leads to a  
 569 new family of split-based phylogenetic diversity indices.

570 Let  $n = |X| \geq 3$  and  $\theta = x_0, x_1, \dots, x_{n-1}$  be an ordering of the elements in  $X$ . Define the split  
 571 system

$$572 \quad \mathcal{S}_\theta = \{ \{x_i, \dots, x_j\} | X \setminus \{x_i, \dots, x_j\} : 0 \leq i \leq j < n - 1 \}.$$

573 A split system  $\mathcal{S}$  on  $X$  is *circular* if there exists an ordering  $\theta$  of the elements in  $X$  such that  
 574  $\mathcal{S} \subseteq \mathcal{S}_\theta$ . If  $\mathcal{S} = \mathcal{S}_\theta$  for some ordering  $\theta$  of the elements in  $X$  we call  $\mathcal{S}$  a *full* circular  
 575 split system on  $X$ . Clearly, a circular split system contains at most  $\binom{n}{2}$  splits. Moreover, every  
 576 compatible split system is circular (Bandelt and Dress, 1992). Volkmann et al. (2014) considered  
 577 the Shapley value and another phylogenetic diversity index on weighted circular split systems  
 578 (see also the more recent work by Abhari et al. (2024)).

579 Let  $\binom{X}{2}$  denote the set of all 2-element subsets of  $X$ . For a circular split system  $\mathcal{S}$  on  $X$   
 580 we consider the  $|\mathcal{S}| \times \binom{n}{2}$ -matrix  $M(\mathcal{S})$  whose rows correspond to the splits in  $\mathcal{S}$  and whose  
 581 columns correspond to the 2-element subsets in  $\binom{X}{2}$ . The entry of  $M = M(\mathcal{S})$  corresponding to  
 582  $S = A|B \in \mathcal{S}$  and  $\{x, y\} \in \binom{X}{2}$  is defined as

$$583 \quad m_{(S, \{x, y\})} = \begin{cases} 1 & \text{if } |\{x, y\} \cap A| = |\{x, y\} \cap B| = 1 \\ 0 & \text{otherwise.} \end{cases}$$

584 The matrix  $M(\mathcal{S})$  describes how a weighting  $\lambda \in \mathbb{L}(\mathcal{S})$  gives rise to pairwise distances between  
 585 the elements in  $X$ :

$$586 \quad (22) \quad d_\lambda(\{x, y\}) = \sum_{S \in \mathcal{S}} \lambda(S) \cdot m_{(S, \{x, y\})},$$

587 or, more compactly,  $d_\lambda = \lambda \cdot M$ , where  $\lambda$  is viewed as a row vector with  $|\mathcal{S}|$  entries and  $d_\lambda$  as  
 588 a row vector with  $\binom{n}{2}$  entries. The rows of  $M(\mathcal{S})$  are linearly independent (Bandelt and Dress,  
 589 1992). Thus, every weighting  $\lambda$  yields a distinct  $d_\lambda$ .

590 Since the matrix  $M(\mathcal{S})$  has full rank, there exists an  $\binom{n}{2} \times |\mathcal{S}|$ -matrix  $R$  that is *right inverse*  
 591 to  $M(\mathcal{S})$ , that is,  $M(\mathcal{S}) \cdot R$  yields the  $|\mathcal{S}| \times |\mathcal{S}|$ -identity matrix. The matrix  $R$  need not be  
 592 unique, however. Any such matrix corresponds to a linear estimator of a weighting of the splits  
 593 in  $\mathcal{S}$  from pairwise distances between the elements in  $X$  (see e.g. Pardi and Gascuel (2012) for a  
 594 discussion of such estimators for unrooted phylogenetic trees). Fixing such a matrix  $R$ , we define  
 595 a phylogenetic diversity index  $\Psi_R$  on  $\mathbb{L}(\mathcal{S})$  by putting

$$596 \quad (\Psi_R(\lambda))(x) = \frac{1}{2} \cdot \sum_{y \in X \setminus \{x\}} \sum_{S \in \mathcal{S}} d_\lambda(\{x, y\}) \cdot r_{(\{x, y\}, S)}$$

597 for all  $x \in X$  and all  $\lambda \in \mathbb{L}(\mathcal{S})$ . In view of (22), we have  $\Psi_R(\lambda) = \lambda \cdot \Gamma_R$  for the  $|\mathcal{S}| \times |X|$ -matrix  
 598  $\Gamma = \Gamma_{\Psi_R}$  whose entries are

$$599 \quad \gamma_{(S, x)} = \frac{1}{2} \cdot \sum_{y \in X \setminus \{x\}} \sum_{S' \in \mathcal{S}} m_{(S, \{x, y\})} \cdot r_{(\{x, y\}, S')}.$$

600 Moreover, since  $R$  is right inverse to  $M(\mathcal{S})$ , we have

$$601 \quad \sum_{x \in X} \gamma_{(S, x)} = \frac{1}{2} \cdot \sum_{x \in X} \sum_{y \in X \setminus \{x\}} \sum_{S' \in \mathcal{S}} m_{(S, \{x, y\})} \cdot r_{(\{x, y\}, S')} = \sum_{\{x, y\} \in \binom{X}{2}} \sum_{S' \in \mathcal{S}} m_{(S, \{x, y\})} \cdot r_{(\{x, y\}, S')} = 1$$

602 for all  $S \in \mathcal{S}$ . The next lemma summarizes these basic facts about  $\Psi_R$ .

603 **Lemma 7.1.** *Let  $\mathcal{S}$  be a circular split system on  $X$  with  $|X| \geq 3$ . Then, for every matrix  $R$  that is*  
 604 *right inverse to the matrix  $M(\mathcal{S})$ ,  $\Psi_R$  is a complete linear phylogenetic diversity index on  $\mathbb{L}(\mathcal{S})$ .*

605 Consider, as an example, the split system  $\mathcal{S} = \{S_1, \dots, S_5\}$  on  $X = \{a, b, c, d\}$ , for which the  
 606 matrix  $M(\mathcal{S})$  is given in Figure 10(a). The split system  $\mathcal{S}$  is compatible and, thus, circular.  
 607 There are infinitely many matrices  $R$  that are right inverse to the matrix  $M(\mathcal{S})$  and they can be  
 608 described by five parameters  $p_1, \dots, p_5 \in \mathbb{R}$  as shown in Figure 10(b). The matrix  $\Gamma_{\Psi_R}$  for the  
 609 resulting complete linear phylogenetic diversity index  $\Psi_R$  is given in Figure 10(c). This index

(a)		$\{a, b\}$	$\{a, c\}$	$\{a, d\}$	$\{b, c\}$	$\{b, d\}$	$\{c, d\}$
$S_1 = \{a\} \{b, c, d\}$	1	1	1	0	0	0	0
$S_2 = \{b\} \{a, c, d\}$	1	0	0	1	1	0	0
$S_3 = \{c\} \{a, b, d\}$	0	1	0	1	0	1	1
$S_4 = \{d\} \{a, b, c\}$	0	0	1	0	1	1	1
$S_5 = \{a, b\} \{c, d\}$	0	1	1	1	1	1	0

(b)		$S_1$	$S_2$	$S_3$	$S_4$	$S_5$	(c)		$a$	$b$	$c$	$d$
$\{a, b\}$	$\frac{1}{2}$	$\frac{1}{2}$	0	0	$-\frac{1}{2}$	$-\frac{1}{2}$	$S_1$	$\frac{1}{2}$	$\frac{1}{4}$	$r$	$\frac{1}{4} - r$	
$\{a, c\}$	$p_1$	$p_2$	$p_3$	$p_4$	$p_5$	$p_5$	$S_2$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4} - r$	$r$	
$\{a, d\}$	$\frac{1}{2} - p_1$	$-\frac{1}{2} - p_2$	$-p_3$	$-p_4$	$\frac{1}{2} - p_5$	$\frac{1}{2} - p_5$	$S_3$	$r$	$\frac{1}{4} - r$	$\frac{1}{2}$	$\frac{1}{4}$	
$\{b, c\}$	$-p_1$	$-p_2$	$\frac{1}{2} - p_3$	$-\frac{1}{2} - p_4$	$\frac{1}{2} - p_5$	$\frac{1}{2} - p_5$	$S_4$	$\frac{1}{4} - r$	$r$	$\frac{1}{4}$	$\frac{1}{2}$	
$\{b, d\}$	$-\frac{1}{2} + p_1$	$\frac{1}{2} + p_2$	$-\frac{1}{2} + p_3$	$\frac{1}{2} + p_4$	$p_5$	$p_5$	$S_5$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	
$\{c, d\}$	0	0	$\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$						

FIGURE 10. (a) The matrix  $M(\mathcal{S})$  for the split system  $\mathcal{S} = \{S_1, \dots, S_5\}$  on  $X = \{a, b, c, d\}$ . (b) The matrices  $R$  that are right inverse to  $M(\mathcal{S})$ . (c) The matrix  $\Gamma_{\Psi_R}$  for the phylogenetic diversity index  $\Psi_R$  on  $\mathbb{L}(\mathcal{S})$ .

610 has only a single parameter  $r \in \mathbb{R}$  with  $r = \frac{1}{2}(p_1 + \dots + p_5)$ . If all entries in  $\Gamma_{\Psi_R}$  are required to  
611 be non-negative, we need to restrict this parameter to  $0 \leq r \leq \frac{1}{4}$ . The Pauplin index in (Wicke  
612 and Steel, 2020, Sec. 5.2) corresponds to  $r = \frac{1}{8}$ .

613 We now present the main result of this section which focuses on full circular split systems.

614 **Theorem 7.2.** *Let  $\mathcal{S}$  be a full circular split system on  $X$  with  $\mathcal{S} = \mathcal{S}_\theta$  for the ordering  $\theta =$   
615  $x_0, x_1, \dots, x_{n-1}$  of  $X$ . Then the matrix  $M(\mathcal{S})$  has a unique right inverse matrix  $R$  and the matrix  
616  $\Gamma = \Gamma_{\Psi_R}$  for the complete linear phylogenetic diversity index  $\Psi_R$  on  $\mathbb{L}(\mathcal{S})$  has the entry*

$$617 \quad \gamma_{(S,x)} = \begin{cases} \frac{1}{2} & \text{if } i = j = k \text{ or } i = 0, j = n - 2, k = n - 1 \\ \frac{1}{4} & \text{if } i \neq j, k \in \{i, j\} \text{ or} \\ & (i - 1) \bmod n \neq (j + 1) \bmod n, k \in \{(i + 1) \bmod n, (j + 1) \bmod n\} \\ 0 & \text{otherwise} \end{cases}$$

618 for the split  $S = \{x_i, \dots, x_j\} | X \setminus \{x_i, \dots, x_j\} \in \mathcal{S}$  and the element  $x = x_k \in X$ .

619 *Proof:* Since  $\mathcal{S}$  is a full circular split system, the matrix  $M = M(\mathcal{S})$  is a square matrix. Hence,  
620  $M$  has a unique right inverse matrix  $R$  which is just the usual inverse matrix of  $M$ . Moreover, as  
621 shown by Chepoi and Fichet (1998), the matrix  $R$  has the entry

$$622 \quad (23) \quad r_{(\{y,z\}, S')} = \begin{cases} \frac{1}{2} & \text{if } \{y, z\} = \{a, (b + 1) \bmod n\} \text{ or } \{y, z\} = \{(a - 1) \bmod n, b\} \\ -\frac{1}{2} & \text{if } \{y, z\} = \{a, b\} \text{ or } \{y, z\} = \{(a - 1) \bmod n, (b + 1) \bmod n\} \\ 0 & \text{otherwise} \end{cases}$$

623 for  $\{y, z\} \in \binom{X}{2}$  and  $S' = \{x_a, \dots, x_b\} | X \setminus \{x_a, \dots, x_b\} \in \mathcal{S}$ .

624 Consider the split  $S = \{x_i, \dots, x_j\} | X \setminus \{x_i, \dots, x_j\} \in \mathcal{S}$  and the element  $x = x_k \in X$ . By sym-  
625 metry, it suffices to consider the following three cases.

626 Case 1:  $i < k < j$ . Consider  $y \in X \setminus \{x\}$  with  $m_{(S, \{x, y\})} = 1$ . Then, by the definition of the  
627 matrix  $M$ , we have  $y \in X \setminus \{x_i, \dots, x_j\}$ . This implies, in view of (23), that there exist precisely  
628 two splits  $S' \in \mathcal{S}$  with  $r_{(\{x, y\}, S')} = \frac{1}{2}$  and precisely two splits  $S' \in \mathcal{S}$  with  $r_{(\{x, y\}, S')} = -\frac{1}{2}$ . Hence,  
629 we have

$$630 \quad \sum_{S' \in \mathcal{S}} m_{(S, \{x, y\})} \cdot r_{(\{x, y\}, S')} = 0,$$

631 implying that  $\gamma_{(S, x)} = \frac{1}{2} \cdot 0 = 0$ , as required.

632 Case 2:  $i = k < j$ . Consider again  $y \in X \setminus \{x\}$  with  $m_{(S, \{x, y\})} = 1$ . If  $y \neq x_{(i-1) \bmod n}$ , we have  
633  $\sum_{S' \in \mathcal{S}} m_{(S, \{x, y\})} \cdot r_{(\{x, y\}, S')} = 0$  using the same argument as in Case 1. Otherwise, there exist  
634 precisely two splits  $S' \in \mathcal{S}$  with  $r_{(\{x, y\}, S')} = \frac{1}{2}$  but only one split  $S' \in \mathcal{S}$  with  $r_{(\{x, y\}, S')} = -\frac{1}{2}$ ,  
635 implying

$$636 \quad \sum_{S' \in \mathcal{S}} m_{(S, \{x, y\})} \cdot r_{(\{x, y\}, S')} = \frac{1}{2}$$

637 and, thus,  $\gamma_{(S, x)} = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$ , as required.

638 Case 3:  $i = k = j$ . Then we have  $m_{(S, \{x, y\})} = 1$  for all  $y \in X \setminus \{x\}$ . If  $y \notin \{x_{(k-1) \bmod n}, x_{(k+1) \bmod n}\}$   
639 we again have  $\sum_{S' \in \mathcal{S}} m_{(S, \{x, y\})} \cdot r_{(\{x, y\}, S')} = 0$  by the argument used in Case 1. Otherwise, we have  
640  $\sum_{S' \in \mathcal{S}} m_{(S, \{x, y\})} \cdot r_{(\{x, y\}, S')} = \frac{1}{2}$  by the argument used in Case 2, and, thus,  $\gamma_{(S, x)} = \frac{1}{2} \cdot (\frac{1}{2} + \frac{1}{2}) = \frac{1}{2}$ ,  
641 as required. ■

642

643 While Theorem 7.2 focuses on full circular split systems, it also suggests two specific phy-  
644 logenetic diversity indices  $\Psi_1$  and  $\Psi_2$  for any circular split system that is not full. Consider,  
645 as an example, again the split system  $\mathcal{S}$  in Figure 10(a). We have  $\mathcal{S} \subseteq \mathcal{S}_\theta$  for the ordering  
646  $\theta = a, b, c, d$ . To obtain the complete linear phylogenetic diversity index  $\Psi_1$  on  $\mathbb{L}(\mathcal{S})$ , we re-  
647 strict the matrix  $\Gamma$  obtained for  $\mathcal{S}_\theta$  by Theorem 7.2 to those rows that correspond to splits in  $\mathcal{S}$ .  
648 The resulting matrix  $\Gamma_{\Psi_1}$  is then the matrix in Figure 10(c) with  $r = 0$ . To obtain the complete  
649 linear phylogenetic diversity index  $\Psi_2$  on  $\mathbb{L}(\mathcal{S})$  we restrict the matrix  $R$  obtained for  $\mathcal{S}_\theta$  by  
650 Theorem 7.2 to those columns that correspond to splits in  $\mathcal{S}$ . The resulting matrix  $R'$  is a matrix  
651 that is right inverse to  $M(\mathcal{S})$  and we put  $\Psi_2 = \Psi_{R'}$ . More specifically, the matrix  $R'$  equals the  
652 matrix in Figure 10(b) with  $p_1 = p_3 = \frac{1}{2}$  and  $p_2 = p_4 = p_5 = 0$  and the matrix  $\Gamma_{\Psi_2}$  is the matrix  
653 in Figure 10(c) with  $r = \frac{1}{2}$ .

654 We conclude this section looking again at the example of owl populations in Figure 2(a).  
655 The corresponding split system  $\mathcal{S}$  in Figure 12 in the appendix is circular with the ordering  
656  $\theta = a, f, r, h, s, m$  of the elements in  $X$ . In Figure 11 we give the matrices  $\Gamma_{\Psi_R}$  obtained from  
657 the right inverse matrices  $R$  of  $M(\mathcal{S})$ . They contain five parameters  $r_1, \dots, r_5 \in \mathbb{R}$ . Only by  
658 putting  $r_1 = r_4 = \frac{1}{4}$  and  $r_2 = r_3 = r_5 = 0$ , however, all entries of  $\Gamma_{\Psi_R}$  are non-negative and then  
659  $\Gamma_{\Psi_R}$  equals the restriction of the matrix  $\Gamma$  obtained for  $\mathcal{S}_\theta$  by Theorem 7.2 to those rows that  
660 correspond to splits in  $\mathcal{S}$ . Using the weighting  $\lambda$  of the splits given in Figure 12 in the appendix,  
661 we obtain the values of  $\Psi_R$  given in Figure 2(b), which yields the same ranking of the six owl  
662 populations as  $FP_u$  and  $SV$ .

	$a$	$f$	$h$	$m$	$r$	$s$
$S_1$	$\frac{1}{2}$	$r_1$	$r_2$	$\frac{1}{4}$	$\frac{1}{4} - r_1 - r_2$	$0$
$S_2$	$r_2$	$\frac{1}{4} - r_2 - r_4$	$\frac{1}{2}$	$r_3$	$r_4$	$\frac{1}{4} - r_3$
$S_3$	$\frac{1}{4}$	$-r_3 - r_5$	$r_3$	$\frac{1}{2}$	$r_5$	$\frac{1}{4}$
$S_4$	$\frac{1}{4} - r_1 - r_2$	$r_1 + r_2$	$r_4$	$r_5$	$\frac{1}{2}$	$\frac{1}{4} - r_4 - r_5$
$S_5$	$0$	$-\frac{1}{4} + r_3 + r_4 + r_5$	$\frac{1}{4} - r_3$	$\frac{1}{4}$	$\frac{1}{4} - r_4 - r_5$	$\frac{1}{2}$
$S_6$	$\frac{1}{2} - r_1$	$r_1$	$\frac{1}{4} - r_4$	$\frac{1}{4} - r_3 - r_5$	$\frac{1}{4}$	$-\frac{1}{4} + r_3 + r_4 + r_5$
$S_7$	$\frac{1}{4} + r_2$	$-r_2$	$\frac{1}{4}$	$\frac{1}{4} - r_3$	$\frac{1}{4}$	$r_3$
$S_8$	$\frac{1}{4}$	$-\frac{1}{4} + r_4$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4} - r_4$	$\frac{1}{4}$
$S_9$	$\frac{1}{4}$	$r_1 - r_3 - r_5$	$r_2 + r_3$	$\frac{1}{4}$	$\frac{1}{4} - r_1 - r_2 + r_5$	$\frac{1}{4}$
$S_{10}$	$\frac{1}{4}$	$-\frac{1}{4} + r_1 + r_4$	$\frac{1}{4} + r_2$	$0$	$\frac{1}{2} - r_1 - r_2 - r_4$	$\frac{1}{4}$

FIGURE 11. The matrices  $\Gamma_\Psi$  for the phylogenetic diversity indices  $\Psi = \Psi_R$  obtained from the matrices  $R$  that are right inverse to the matrix  $M(\mathcal{S})$  for the split system  $\mathcal{S}$  on  $X = \{a, f, h, m, r, s\}$  represented by the network in Figure 2(a). The splits  $S_1, \dots, S_{10}$  are listed in Figure 12 in the appendix.

663

## 8. CONCLUSION

664 In this paper, we have presented a framework for phylogenetic diversity indices defined on  
 665 linear spaces coming from weighted cluster and split systems. Using some examples of popular  
 666 tree-based phylogenetic diversity indices from the literature, we have shown that this framework  
 667 can yield generalizations of these indices for cluster and split systems as well as allowing us to  
 668 gain a better understanding of their interrelationships.

669 Note that in our framework presented in Figure 7, by associating to any split system  $\mathcal{S}$  on  $X$   
 670 the cluster system  $\mathcal{C}(\mathcal{S})$  on  $X$  and then considering maps  $\tau$ , we have focused on deriving split-  
 671 based indices from cluster-based indices. In the affine and projective clustering approach, how-  
 672 ever, there are also ways to associate to any cluster system  $\mathcal{C}$  on  $X$  a split system  $\mathcal{S}(\mathcal{C})$  on  $X$   
 673 (see, e.g., Dress 2012, Sec. 9.3). Thus, it could be interesting to investigate how this fact might  
 674 be used to derive cluster-based indices from split-based indices.

675 In our results, we have considered cluster and split systems in general, special examples of  
 676 which include hierarchical cluster systems, compatible split systems (which correspond to rooted  
 677 and unrooted phylogenetic trees, respectively) and circular split systems. There are, however,  
 678 various other special classes of cluster and split systems that could be interesting to investigate  
 679 within our framework. For example, it would also be interesting to consider diversity indices  
 680 coming from *weak hierarchies*, a special type of cluster system introduced by Bandelt and Dress  
 681 (1989). The advantage of considering such specialized cluster and split systems is that they can  
 682 be efficiently computed from biological data, making them potentially more useful for applica-  
 683 tions.

684 In the literature, various approaches have been proposed to generalize tree-based phylogenetic  
 685 diversity indices using phylogenetic networks, a graph-theoretical generalization of phylogenetic  
 686 trees (Coronado et al., 2018; Volkmann et al., 2014; Wicke and Fischer, 2018). Such networks  
 687 are essentially directed, acyclic, graphs with a single root and whose set of leaves corresponds



688 to some collection of species. The fair proportion index, for example, is generalized in terms of  
689 such networks by Coronado et al. (2018). In general, phylogenetic networks give rise to cluster  
690 systems (see, e.g., Steel 2016, Sec. 10.3.4) by, for example, taking the set of leaves that lie  
691 below a vertex or an edge in the network (just as with rooted phylogenetic trees). Thus, it could  
692 be interesting to explore how phylogenetic diversity indices defined in terms of phylogenetic  
693 networks, such as, for example, those considered by Wicke and Fischer (2018), fit into our  
694 cluster based framework. This could also be interesting to investigate for *split networks* such  
695 as the one presented in Figure 2(a), which are a certain type of undirected phylogenetic network  
696 (see, e.g., Dress 2012, Sec. 4.4).

697 With the different ways of defining diversity indices via clusters and splits and translating  
698 between the two viewpoints, it could also be interesting to analyze under which circumstances  
699 different indices give exactly the same score and thus also identical rankings of the taxa. For  
700 example, Wicke and Steel (2020) characterized precisely when the fair proportion index and the  
701 equal splits index for rooted binary phylogenetic trees coincide. Thus, it would be interesting  
702 to characterize which conditions a weighted cluster system (resp. weighted split system) has to  
703 satisfy in order to obtain similar results for pairs of cluster- or split-based indices.

704 In another direction, it could be interesting to apply our framework to establish properties and  
705 generalizations of other tree-based phylogenetic diversity indices that we did not consider in this  
706 paper. Indeed, as we have demonstrated, sometimes expressing indices in terms of clusters or  
707 splits can lead to more concise proofs for showing that they have certain properties. For example,  
708 it would be interesting to consider some of the questions asked in Section 6 of Wicke and Steel  
709 (2020) within our new framework.

710 Finally, concerning the generalization of the Pauplin index presented in Section 7, we saw in  
711 the examples in Figure 10 and Figure 11 that even when a circular split system  $\mathcal{S}$  is not full  
712 there may exist right inverse matrices  $R$  for  $M(\mathcal{S})$  such that for the complete linear phylogenetic  
713 diversity index  $\Psi_R$  the matrix  $\Gamma_{\Psi_R}$  has non-negative entries. Can we characterize when this is  
714 the case and, more specifically, give the number of parameters in the matrix  $\Gamma_{\Psi_R}$ ? As a potential  
715 direction for further generalization, one could consider split systems  $\mathcal{S}$  for which the matrix  
716  $M(\mathcal{S})$  has full rank, which are known as *linearly independent* split systems. There exist such  
717 split systems that are not circular (Bryant and Dress, 2007). Can Theorem 7.2 be generalized in  
718 some way to all maximum sized linearly independent split systems?

719 **Data availability.** No data was generated.

720 **Acknowledgment.** All authors thank Schloss Dagstuhl – Leibniz Centre for Informatics – for  
721 hosting the Seminar 19443 *Algorithms and Complexity in Phylogenetics* in October 2019, where  
722 this work was initiated. The authors also thank the two anonymous reviewers for their helpful  
723 comments on a previous version of this paper.

724

## REFERENCES

725 Abhari, N., Colijn, C., Mooers, A., and Tupper, P. (2024). Capturing diversity: Split systems and  
726 circular approximations for conservation. *Journal of Theoretical Biology*, 578:111689.

- 727 Aumann, R. J. (1994). The Shapley value. In *Game-Theoretic Methods in General Equilibrium*  
728 *Analysis*, pages 61–66. Springer.
- 729 Bandelt, H.-J. and Dress, A. (1992). A canonical decomposition theory for metrics on a finite  
730 set. *Advances in Mathematics*, 92:47–105.
- 731 Bandelt, H.-J. and Dress, A. W. (1989). Weak hierarchies associated with similarity mea-  
732 sures—an additive clustering technique. *Bulletin of Mathematical Biology*, 51(1):133–166.
- 733 Bordewich, M. and Semple, C. (2024). Quantifying the difference between phylogenetic diver-  
734 sity and diversity indices. *Journal of Mathematical Biology*, 88(4):1–25.
- 735 Branzei, R., Dimitrov, D., and Tijs, S. (2008). *Models in cooperative game theory*, volume 556.  
736 Springer Science & Business Media.
- 737 Bryant, D. and Dress, A. (2007). Linearly independent split systems. *European Journal of*  
738 *Combinatorics*, 28(6):1814–1831.
- 739 Bryant, D. and Moulton, V. (2004). Neighbor-net: an agglomerative method for the construction  
740 of phylogenetic networks. *Molecular Biology and Evolution*, 21(2):255–265.
- 741 Chepoi, V. and Fichet, B. (1998). A note on circular decomposable metrics. *Geometriae Dedi-*  
742 *cata*, 69:237–240.
- 743 Coronado, T. M., Riera, G., and Rosselló, F. (2018). The fair proportion is a Shapley value on  
744 phylogenetic networks too. In *Enjoying Natural Computing*, pages 77–87. Springer.
- 745 Dress, A. (1997). Towards a theory of holistic clustering. In *Mathematical Hierarchies and*  
746 *Biology*, pages 271–289. American Mathematical Society.
- 747 Dress, A. (2012). *Basic phylogenetic combinatorics*. Cambridge University Press.
- 748 Dubey, P. (1975). On the uniqueness of the Shapley value. *International Journal of Game*  
749 *Theory*, 4:131–139.
- 750 Faith, D. P. (1992). Conservation evaluation and phylogenetic diversity. *Biological Conservation*,  
751 61(1):1–10.
- 752 Fuchs, M. and Jin, E. Y. (2015). Equality of Shapley value and fair proportion index in phyloge-  
753 netic trees. *Journal of Mathematical Biology*, 71(5):1133–1147.
- 754 Gumbs, R., Gray, C. L., Böhm, M., Burfield, I. J., Couchman, O. R., Faith, D. P., Forest, F.,  
755 Hoffmann, M., Isaac, N. J. B., Jetz, W., Mace, G. M., Mooers, A. O., Safi, K., Scott, O., Steel,  
756 M., Tucker, C. M., Pearse, W. D., Owen, N. R., and Rosindell, J. (2023). The EDGE2 protocol:  
757 Advancing the prioritisation of evolutionarily distinct and globally endangered species for  
758 practical conservation action. *PLOS Biology*, 21(2):e3001991.
- 759 Haake, C.-J., Kashiwada, A., and Su, F. E. (2008). The Shapley value of phylogenetic trees.  
760 *Journal of Mathematical Biology*, 56(4):479–497.
- 761 Huson, D. H. and Bryant, D. (2005). Application of phylogenetic networks in evolutionary  
762 studies. *Molecular Biology and Evolution*, 23(2):254–267.
- 763 Isaac, N. J., Turvey, S. T., Collen, B., Waterman, C., and Baillie, J. E. (2007). Mammals on the  
764 EDGE: Conservation priorities based on threat and phylogeny. *PLoS ONE*, 2(3):e296.
- 765 Jukna, S. (2011). *Extremal combinatorics: with applications in computer science*. Springer.
- 766 Kleinman, A., Harel, M., and Pachter, L. (2013). Affine and projective tree metric theorems.  
767 *Annals of Combinatorics*, 17:205–228.
- 768 Manson, K. (2024). The robustness of phylogenetic diversity indices to extinctions. *Journal of*  
769 *Mathematical Biology*, 89(1).

- 770 Manson, K. and Steel, M. (2023). Spaces of phylogenetic diversity indices: combinatorial and  
771 geometric properties. *Bulletin of Mathematical Biology*, 85(78).
- 772 Pardi, F. and Gascuel, O. (2012). Combinatorics of distance-based tree inference. *Proceedings*  
773 *of the National Academy of Sciences*, 109(41):16443–16448.
- 774 Pauplin, Y. (2000). Direct calculation of a tree length using a distance matrix. *Journal of Molec-*  
775 *ular Evolution*, 51:41–47.
- 776 Redding, D. W. (2003). Incorporating genetic distinctness and reserve occupancy into a conser-  
777 vation prioritisation approach. Master’s thesis, University Of East Anglia, Norwich, UK.
- 778 Redding, D. W., Hartmann, K., Mimoto, A., Bokal, D., DeVos, M., and Mooers, A. (2008).  
779 Evolutionarily distinctive species often capture more phylogenetic diversity than expected.  
780 *Journal of Theoretical Biology*, 251(4):606–615.
- 781 Redding, D. W., Mazel, F., and Mooers, A. Ø. (2014). Measuring evolutionary isolation for  
782 conservation. *PLoS ONE*, 9(12):e113490.
- 783 Redding, D. W. and Mooers, A. Ø. (2006). Incorporating evolutionary measures into conserva-  
784 tion prioritization. *Conservation Biology*, 20(6):1670–1678.
- 785 Semple, C. and Steel, M. (2003). *Phylogenetics*. Oxford University Press.
- 786 Shapley, L. S. (1953). A value for  $n$ -person games. *Contributions to the Theory of Games*,  
787 2(28):307–317.
- 788 Spillner, A., Nguyen, B. T., and Moulton, V. (2008). Computing phylogenetic diversity for split  
789 systems. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 5(2):235–  
790 244.
- 791 Steel, M. (2016). *Phylogeny: Discrete and random processes in evolution*. SIAM, Philadelphia  
792 PA.
- 793 Tucker, C. M., Cadotte, M. W., Carvalho, S. B., Davies, T. J., Ferrier, S., Fritz, S. A., Grenyer, R.,  
794 Helmus, M. R., Jin, L. S., Mooers, A. O., Pavoine, S., Purschke, O., Redding, D. W., Rosauer,  
795 D. F., Winter, M., and Mazel, F. (2016). A guide to phylogenetic metrics for conservation,  
796 community ecology and macroecology. *Biological Reviews*, 92(2):698–715.
- 797 Vane-Wright, R., Humphries, C., and Williams, P. (1991). What to protect?—Systematics and  
798 the agony of choice. *Biological Conservation*, 55(3):235–254.
- 799 Volkmann, L., Martyn, I., Moulton, V., Spillner, A., and Mooers, A. O. (2014). Prioritizing  
800 populations for conservation using phylogenetic networks. *PLoS One*, 9(2).
- 801 Wicke, K. (2020). *Novel Aspects of Mathematical Phylogenetics*. PhD thesis, Universität Greif-  
802 swald.
- 803 Wicke, K. and Fischer, M. (2018). Phylogenetic diversity and biodiversity indices on phyloge-  
804 netic networks. *Mathematical Biosciences*, 298:80–90.
- 805 Wicke, K. and Steel, M. (2020). Combinatorial properties of phylogenetic diversity indices.  
806 *Journal of Mathematical Biology*, 80(3):687–715.

## APPENDIX

	$\lambda(S_i)$
$S_1 = \{a\} \{f, h, m, r, s\}$	0.182
$S_2 = \{h\} \{a, f, m, r, s\}$	0.348
$S_3 = \{m\} \{a, f, h, r, s\}$	0.418
$S_4 = \{r\} \{a, f, h, m, s\}$	0.166
$S_5 = \{s\} \{a, f, h, m, r\}$	0.495
$S_6 = \{a, f\} \{h, m, r, s\}$	0.075
$S_7 = \{a, f, r\} \{h, m, s\}$	0.133
$S_8 = \{a, f, r, h\} \{m, s\}$	0.071
$S_9 = \{a, m\} \{f, h, r, s\}$	0.044
$S_{10} = \{a, m, s\} \{f, h, r\}$	0.050

FIGURE 12. The split system  $\mathcal{S} = \{S_1, \dots, S_{10}\}$  on  $X = \{a, f, h, m, r, s\}$  and its weighting  $\lambda$  represented by the network in Figure 2(a). The elements in  $X$  represent sampling locations of spotted owls in western North America ( $a$  = Aguascalientes, Mexico;  $f$  = San Fransisco Peaks, AZ;  $h$  = Huachuca Mountains, AZ;  $m$  = Marin County, CA;  $r$  = Capitol Reef National Park, UT;  $s$  = San Bernardino Mountains, CA). The weighted split system is computed from the pairwise genetic distances between these populations given in Table S1 of (Volkman et al., 2014) using the implementation of NeighborNet (Bryant and Moulton, 2004) in Splitstree (Huson and Bryant, 2005). The same methodology was applied by Volkman et al. 2014 to all 32 populations in the data set. Here six populations are selected for illustration purposes. The total weight of all splits in  $\mathcal{S}$  is 1.980 (all weights rounded to three decimal places).

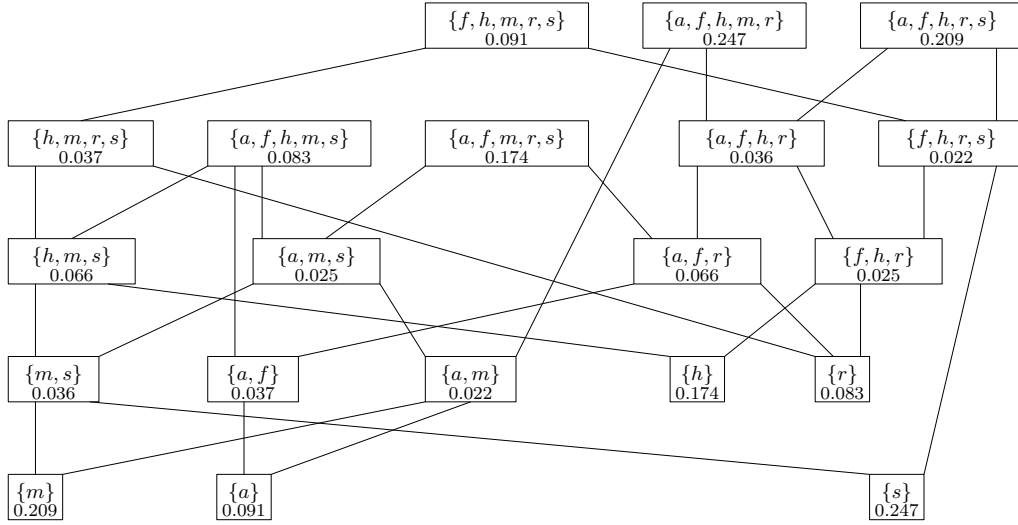


FIGURE 13. The Hasse diagram for the clusters in the cluster system  $\mathcal{C}(\mathcal{S})$  on  $X = \{a, f, h, m, r, s\}$  computed from the split system  $\mathcal{S}$  on  $X$  in Figure 12. The number below each cluster is the weight of the cluster obtained by the map  $\tau$  defined in (17) from the weights of the splits in  $\mathcal{S}$ .

	$a$	$f$	$h$	$m$	$r$	$s$
$\{a\}$	1	0	0	0	0	0
$\{h\}$	0	0	1	0	0	0
$\{m\}$	0	0	0	1	0	0
$\{r\}$	0	0	0	0	1	0
$\{s\}$	0	0	0	0	0	1
$\{a, f\}$	$\frac{1}{2}$	$\frac{1}{2}$	0	0	0	0
$\{a, m\}$	$\frac{1}{2}$	0	0	$\frac{1}{2}$	0	0
$\{m, s\}$	0	0	0	$\frac{1}{2}$	0	$\frac{1}{2}$
$\{a, f, r\}$	$\frac{1}{4}$	$\frac{1}{4}$	0	0	$\frac{1}{2}$	0
$\{a, m, s\}$	$\frac{1}{4}$	0	0	$\frac{1}{2}$	0	$\frac{1}{4}$
$\{f, h, r\}$	0	$\frac{1}{3}$	$\frac{1}{3}$	0	$\frac{1}{3}$	0
$\{h, m, s\}$	0	0	$\frac{1}{2}$	$\frac{1}{4}$	0	$\frac{1}{4}$
$\{a, f, h, r\}$	$\frac{1}{8}$	$\frac{7}{24}$	$\frac{1}{6}$	0	$\frac{5}{12}$	0
$\{f, h, r, s\}$	0	$\frac{1}{6}$	$\frac{1}{6}$	0	$\frac{1}{6}$	$\frac{1}{2}$
$\{h, m, r, s\}$	0	0	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{8}$
$\{a, f, h, m, r\}$	$\frac{5}{16}$	$\frac{7}{48}$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{5}{24}$	0
$\{a, f, h, m, s\}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{4}$	0	$\frac{1}{6}$
$\{a, f, h, r, s\}$	$\frac{1}{16}$	$\frac{11}{48}$	$\frac{1}{6}$	0	$\frac{7}{24}$	$\frac{1}{4}$
$\{a, f, m, r, s\}$	$\frac{1}{4}$	$\frac{1}{8}$	0	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{8}$
$\{f, h, m, r, s\}$	0	$\frac{1}{12}$	$\frac{5}{24}$	$\frac{1}{16}$	$\frac{1}{3}$	$\frac{5}{16}$

FIGURE 14. The matrix  $\Gamma$  for the equal splits index on  $\mathbb{L}(\mathcal{C})$  for the cluster system  $\mathcal{C}$  in Figure 13.

	$a$	$f$	$h$	$m$	$r$	$s$
$\{a\}$	1	0	0	0	0	0
$\{h\}$	0	0	1	0	0	0
$\{m\}$	0	0	0	1	0	0
$\{r\}$	0	0	0	0	1	0
$\{s\}$	0	0	0	0	0	1
$\{a, f\}$	$\frac{1}{2}$	$\frac{1}{2}$	0	0	0	0
$\{m, s\}$	0	0	0	$\frac{1}{2}$	0	$\frac{1}{2}$
$\{a, f, r\}$	$\frac{1}{4}$	$\frac{1}{4}$	0	0	$\frac{1}{2}$	0
$\{h, m, s\}$	0	0	$\frac{1}{2}$	$\frac{1}{4}$	0	$\frac{1}{4}$
$\{a, f, h, r\}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{2}$	0	$\frac{1}{4}$	0
$\{h, m, r, s\}$	0	0	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{8}$
$\{a, f, h, m, r\}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{8}$	0
$\{a, f, h, m, s\}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{8}$	0	$\frac{1}{8}$
$\{a, f, h, r, s\}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{4}$	0	$\frac{1}{8}$	$\frac{1}{2}$
$\{a, f, m, r, s\}$	$\frac{1}{8}$	$\frac{1}{8}$	0	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
$\{f, h, m, r, s\}$	0	$\frac{1}{2}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{4}$	$\frac{1}{16}$

FIGURE 15. The matrix  $\Gamma$  for the equal splits index on  $\mathbb{L}(\mathcal{C}_1)$  for the cluster system  $\mathcal{C}_1$  obtained from the subset  $\mathcal{S}'_1 = \{S_1, S_2, S_3, S_4, S_5, S_6, S_7, S_8\}$  of the split system  $\mathcal{S}$  in Figure 12. Using the map  $\tau$  defined in (17) to compute the weights of the clusters in  $\mathcal{C}_1$  from the weights of the splits in  $\mathcal{S}'_1$ , we obtain  $\Psi_\tau(s) = 0.450$ ,  $\Psi_\tau(m) = 0.431$ ,  $\Psi_\tau(h) = 0.380$ ,  $\Psi_\tau(r) = 0.267$ ,  $\Psi_\tau(a) = 0.202$ , and  $\Psi_\tau(f) = 0.156$  (all values rounded to three decimal places).

	$a$	$f$	$h$	$m$	$r$	$s$
$\{a\}$	1	0	0	0	0	0
$\{h\}$	0	0	1	0	0	0
$\{m\}$	0	0	0	1	0	0
$\{r\}$	0	0	0	0	1	0
$\{s\}$	0	0	0	0	0	1
$\{a,m\}$	$\frac{1}{2}$	0	0	$\frac{1}{2}$	0	0
$\{a,m,s\}$	$\frac{1}{4}$	0	0	$\frac{1}{4}$	0	$\frac{1}{2}$
$\{f,h,r\}$	0	$\frac{1}{3}$	$\frac{1}{3}$	0	$\frac{1}{3}$	0
$\{f,h,r,s\}$	0	$\frac{1}{6}$	$\frac{1}{6}$	0	$\frac{1}{6}$	$\frac{1}{2}$
$\{a,f,h,m,r\}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{6}$	0
$\{a,f,h,m,s\}$	$\frac{1}{12}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{12}$	0	$\frac{1}{6}$
$\{a,f,h,r,s\}$	$\frac{1}{2}$	$\frac{1}{12}$	$\frac{1}{12}$	0	$\frac{1}{12}$	$\frac{1}{4}$
$\{a,f,m,r,s\}$	$\frac{1}{12}$	$\frac{1}{3}$	0	$\frac{1}{12}$	$\frac{1}{3}$	$\frac{1}{6}$
$\{f,h,m,r,s\}$	0	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{2}$	$\frac{1}{12}$	$\frac{1}{4}$

FIGURE 16. The matrix  $\Gamma$  for the equal splits index on  $\mathbb{L}(\mathcal{C}_2)$  for the cluster system  $\mathcal{C}_2$  obtained from the subset  $\mathcal{S}'_2 = \{S_1, S_2, S_3, S_4, S_5, S_9, S_{10}\}$  of the split system  $\mathcal{S}$  in Figure 12. Using the map  $\tau$  defined in (17) to compute the weights of the clusters in  $\mathcal{C}_2$  from the weights of the splits in  $\mathcal{S}_2$ , we obtain  $\Psi_\tau(s) = 0.388$ ,  $\Psi_\tau(m) = 0.355$ ,  $\Psi_\tau(a) = 0.296$ ,  $\Psi_\tau(h) = 0.280$ ,  $\Psi_\tau(r) = 0.220$ , and  $\Psi_\tau(f) = 0.164$  (all values rounded to three decimal places).