

Research



Cite this article: Francis A, Jarvis PD. 2022 Brauer and partition diagram models for phylogenetic trees and forests. *Proc. R. Soc. A* **478**: 20220044.
<https://doi.org/10.1098/rspa.2022.0044>

Received: 18 January 2022

Accepted: 4 May 2022

Subject Areas:

applied mathematics, algebra, combinatorics

Keywords:

phylogenetics, Brauer diagram, sandwich semigroup, partition monoid

Author for correspondence:

Andrew Francis

e-mail: a.francis@westernsydney.edu.au

Brauer and partition diagram models for phylogenetic trees and forests

Andrew Francis¹ and Peter D. Jarvis²

¹Centre for Research in Mathematics and Data Science, Western Sydney University, Penrith South, New South Wales, Australia

²School of Mathematics and Physics, University of Tasmania, Hobart, Tasmania, Australia

AF, 0000-0002-9938-3499; PDJ, 0000-0002-5330-6789

We introduce a correspondence between phylogenetic trees and Brauer diagrams, inspired by links between binary trees and matchings described by Diaconis and Holmes (1998 *Proc. Natl Acad. Sci. USA* **95**, 14 600–14 602. (doi:10.1073/pnas.95.25.14600)). This correspondence gives rise to a range of semigroup structures on the set of phylogenetic trees, and opens the prospect of many applications. We furthermore extend the Diaconis–Holmes correspondence from binary trees to non-binary trees and to forests, showing for instance that the set of all forests is in bijection with the set of partitions of finite sets.

1. Introduction

Phylogenetic trees are a fundamental and persistent idea used to represent evolutionary relationships between species for nearly two centuries. Their use extends beyond biology to the representation of language evolution, and to decision processes in algorithms. Their appeal is that they provide an extra dimension to the ways to relate the elements of a set beyond a linear order. They have been studied directly in numerous ways, through stochastic processes, combinatorics and geometry (e.g. [1,2]).

An indirect but powerful way to study tree structures is to consider correspondences, or ways to represent trees, within other mathematical objects. This is a standard approach in much of mathematics of course (representation theory is defined by this idea). And there are several known bijections between rooted binary phylogenetic trees and other structures.

© 2022 The Authors. Published by the Royal Society under the terms of the Creative Commons Attribution License <http://creativecommons.org/licenses/by/4.0/>, which permits unrestricted use, provided the original author and source are credited.

For instance, trees correspond to certain polynomials [3,4], to perfect matchings [5], and to more general partitions of finite sets [6]. The latter has been extended to classes of forests relevant to phylogenetics via a correspondence between forests and trees, providing a way to enumerate such forests using Stirling numbers [7]. Other frameworks for capturing the combinatorics of tree counting and tree shapes have been through the analysis of binary sequences [8], via symmetric function theory [9], and numerous others (see for example the OEIS listing A001147 [10]). The particular connection with perfect matchings has prompted the suggestion by Diaconis & Holmes [5] that there may be a relation to certain diagrams that have independently arisen within several branches of algebra, notably those given by the Brauer algebra [11].

In this paper, we take up and develop the link to Brauer algebras, and develop correspondences between the related diagram structures and phylogenetic trees, that provide an algebraic framework for their study. Each labelled, rooted binary tree will correspond to a unique element in the Brauer category, whose elements can be represented by asymmetrical Brauer diagrams. More generally, a corresponding structure for non-binary trees can be defined using the associated partition category (for partition categories see for instance [12,13]), and we are able to extend this correspondence to the set of all forests. The extension to forests is associated with a new bijective correspondence between forests and partitions of finite sets that directly extends the ideas of Erdős & Székely [6] in a direction distinct from that in Erdős [7].

The Brauer diagram framework has the potential to reveal more structure within the space of tree shapes (sometimes called topologies), and to provide new methods to randomly move about that space (something important for many tree reconstruction algorithms). In particular, the inherited structure provides numerous ways to break down tree space: for instance, through the use of Green's relations from semigroup theory, we can break the set of phylogenetic trees into \mathcal{L} -classes, \mathcal{R} -classes, \mathcal{H} -classes and \mathcal{D} -classes, each of which can be interpreted in the light of standard properties of trees (see §4 for definitions of these semigroup classes). A concomitant of this is the possibility to craft new metrics on tree space. Finally, we are able to define an operation on tree-space, effectively a multiplication of trees, *relative to a fixed tree*, that turns the space into a semigroup, using the 'sandwich semigroup' product. It is then possible to consider the 'regular' elements of this semigroup, that provide a subsemigroup of the set of trees with respect to the chosen fixed tree.

The paper begins in §2 with some background to the algebraic structures we will be using to describe tree space, namely the Brauer algebra and related monoids. We also recall key concepts from semigroup theory, especially Green's relations and the corresponding equivalence classes into which a semigroup decomposes. We then describe (§3) the matching result of Diaconis & Holmes [5], and show how it extends to a correspondence with Brauer diagrams. Section 4 then explores the result of transferring the results on the analysis and properties known in the Brauer category, across to phylogenetic trees, and discusses the rich new structure for tree space that becomes available. Section 5 introduces a product on trees, relative to a given tree, derived from the definition of the sandwich semigroup on Brauer diagrams, and explores various algebraic consequences of this powerful concept, including for example the characterization of related entities such as the associated regular subsemigroup of trees. Finally, §6 sketches a further broad generalization of the correspondence, between non-binary trees, and associated partition diagrams and the partition category. A key result is a correspondence results for trees and forests with sets of partitions of finite sets (theorem 6.5). We end with a discussion of directions that this algebraic landscape for phylogenetic trees might open up for further research (§7).

2. Background on phylogenetic trees, matchings and semigroups

For a set X of cardinality $n \geq 2$, a rooted (binary) phylogenetic X -tree T is a graph with n labelled, valence 1 leaf vertices labelled by X , together with $n - 2$ unlabelled, valence three, internal vertices, and an additional unlabelled valence 2 root vertex. Note that we do not consider the root to be an internal vertex. We will usually assume without loss of generality that $X = [n] =$

$\{1, 2, \dots, n\}$. If v is a vertex of T we write $c(v)$ for the set of children of v : vertices that are the targets of edges whose source is v , and we say v is the *parent* of the vertices in $c(v)$. The cardinality of the set $\mathcal{RP}_n^{\text{bin}}$ of such leaf-labelled binary trees is $|\mathcal{RP}_n^{\text{bin}}| = (2n - 3)!!$. This and other tree-related counting problems are well studied [6,8,14] (see also [9,15,16]).

In the latter part of this paper, we will be working with trees that are not necessarily binary. These satisfy the same properties as the binary X -trees, except that the internal vertices are not necessarily of valence 3 (for which we use the term non-binary trees¹): instead they have in-degree 1, and out-degree *at least* 2. The set of all rooted phylogenetic X -trees is denoted \mathcal{RP}_n .

In some contexts, we also permit $|X| = 1$, in which case the ‘trivial’ tree is an isolated leaf labelled by the element of X , and which is also the root. Note, the trivial tree is not an element of \mathcal{RP}_n or $\mathcal{RP}_n^{\text{bin}}$, which are restricted to trees on $n \geq 2$ leaves.

An X -forest is a set of trees whose leaf sets partition X . The *components* of a forest are the individual trees that make it up. The *trivial* forest on X is the one in which all component trees are trivial; that is, a set of isolated leaves labelled by the elements of X .

A perfect matching on a set is a partition of the set into pairs, and the number of perfect matchings on n elements (n even) is $(n - 1)!!$. This gives rise to a natural correspondence between rooted phylogenetic trees, and perfect matchings, on $2(n - 1)$ elements (in this case, the leaf set $[n]$, augmented by an additional $(n - 2)$ elements, $[n + 1, 2n - 2] = \{n + 1, n + 2, \dots, 2n - 3, 2n - 2\}$). This correspondence was formalized as a bijective map by Diaconis & Holmes [5], crystallizing that in Erdős & Székely [6], and is described as follows.

Associating the leaves of T with the first n nodes of the perfect matching, the initial set of pairings among all elements $[2n - 2] = \{1, 2, \dots, 2n - 2\}$ is simply generated from the cherries of the tree (leaves which share an immediate common ancestor), which are at most $\lfloor \frac{1}{2}n \rfloor$ in number. The next unmatched node, $n + 1$, is assigned to the internal tree vertex, both of whose child vertices are already labelled, one of which has the numerically lowest label. The next lowest available node label for any unmatched internal tree vertex or vertices is in turn assigned, among those with the already labelled child vertices, to the one containing the numerically lowest child. The process repeats until the node $2n - 2$, corresponding to the last unlabelled internal tree vertex, is finally identified with its partner. An example, with a chord diagram representation (see below) of the node matchings, used in [5], is given in figure 4. This algorithm is formalized in pseudocode in algorithm 1.

Algorithm 1. Internal vertex labelling algorithm (Erdős- Székely, Diaconis- Holmes).

Require: $T \in \mathcal{RP}_n^{\text{bin}}$

Require: $n \geq 1$

$m \leftarrow 0$

▷ The number of labelled non-leaf vertices

$I \leftarrow$ set of unlabelled vertices in T

▷ Always non-empty (includes the root)

while $|I| > 1$ **do**

$C \leftarrow \{v \in I \mid c(v) \text{ are all labelled}\}$

▷ Always non-empty

for $v \in C$ **do**

if $c(v)$ have the lowest label among elements of C **then**

$v \leftarrow$ labelled $n+m+1$

$m \leftarrow m+1$

$I \leftarrow I \setminus \{v\}$

end if

end for

end while

The correspondence with perfect matchings gives rise naturally to a different family of diagrams, namely those in the Brauer algebra \mathfrak{B}_{n-1} , which (unsurprisingly) has the same

¹Note that the standard term ‘non-binary trees’ includes the set of binary trees.

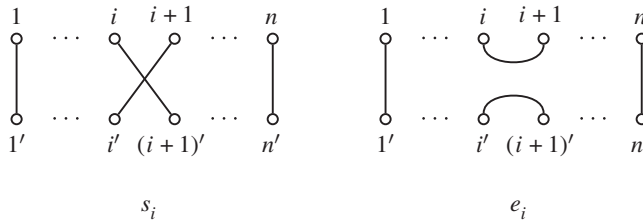


Figure 1. Diagrammatic representation of the generators s_i and $e_i \in \mathfrak{B}_n$, for $i = 1, 2, \dots, n - 1$.

dimension, and the partial Brauer monoid $\mathfrak{B}_{n,n-2}$. In the next section, we introduce these algebraic structures, and describe the correspondence.

A (*set*) *partition* of a finite non-empty set X is a set of pairwise disjoint subsets of X , whose union is X . If π is a set partition of X , write $\ell(\pi)$ for the number of subsets in π , and write $|\pi| := |X|$. An (*integer*) *partition* of a positive integer n is a multiset of positive integers whose sum is n . In this paper, we will use ‘partition’ to refer to *set* partition unless otherwise noted.

A *semigroup* is a set with an associative operation. A simple example is the set of positive integers $\mathbb{N}^{>0} = \{1, 2, 3, \dots\}$ with the operation of addition. If the semigroup has an identity element with respect to its operation, it is called a *monoid*. An example is the non-negative integers $\mathbb{N} := \mathbb{N}^{\geq 0}$, in which the identity with respect to addition is of course 0. Later in the paper, we will introduce the semigroup notions of Green’s relations and ‘eggbox diagrams’ that represent these relations. For an introduction to some of these notions from semigroup theory, we recommend [17,18].

3. Connections between trees, the Brauer algebra \mathfrak{B}_{n-1} and Brauer diagrams $\mathfrak{B}_{n,n-2}$

In the first part of this paper, we develop the correspondence between binary phylogenetic trees and perfect matchings via a further diagrammatic setting, that was already referred to in [5]: namely, by exploiting diagrams linked to the Brauer algebra. These will share a basic structure and fundamental elements (the generators of the algebra), and so we briefly introduce the Brauer algebra $\mathfrak{B}_n(\lambda)$ itself, before moving to the specific set of diagrams $\mathfrak{B}_{n,n-2}$ that will be our focus.

(a) Introduction to Brauer diagrams

For each $n = 1, 2, \dots$, the Brauer algebra $\mathfrak{B}_n(\lambda)$ is the complex algebra generated by the elements [11]

$$\{s_1, s_2, \dots, s_{n-1}, e_1, e_2, \dots, e_{n-1}\}$$

subject to the defining relations given in appendix A. For the purposes of links to phylogenetics, we will make two significant restrictions to this generality.

First, we will restrict to $\lambda = 1$, and focus on $\mathfrak{B}_n = \mathfrak{B}_n(1)$ (although the potential to extend the ideas we discuss here to the full Brauer algebra makes an interesting question that we leave to future work). Second, we will treat \mathfrak{B}_n and related objects $\mathfrak{B}_{m,n}$ (see below) as monoids (or partial monoids), working with just the basis elements of the algebra. These have a standard diagram transcription, in which elements are associated with graphs consisting of $2n$ nodes: ‘upper’ nodes $[n] := \{1, 2, \dots, n\}$ and ‘lower’ nodes $[n'] := \{1', 2', \dots, n'\}$ with edges specified as set partitions of $[n] \cup [n']$ consisting of pairs of nodes. In particular, s_i contains the pairs $\{i, (i+1)'\}, \{(i+1), i'\}$ with all remaining nodes $\{1, 1'\}, \{2, 2'\}, \dots$, sequentially paired, while e_i contains $\{i, (i+1)\}, \{i', (i+1)'\}$, with all remaining nodes sequentially paired (figure 1).

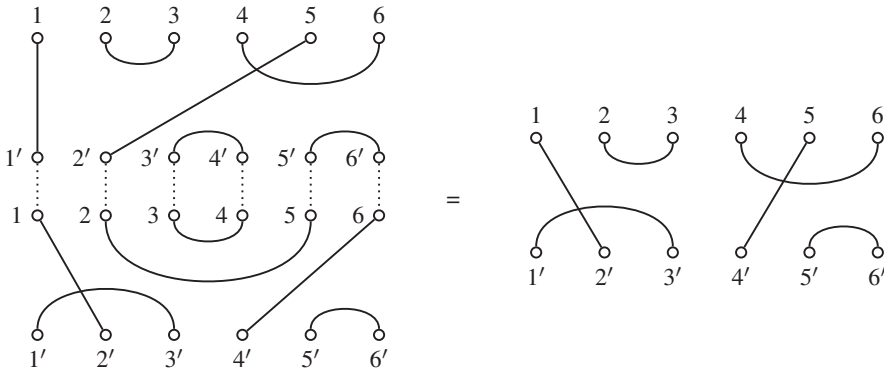


Figure 2. Multiplication of Brauer diagrams in \mathfrak{B}_6 .

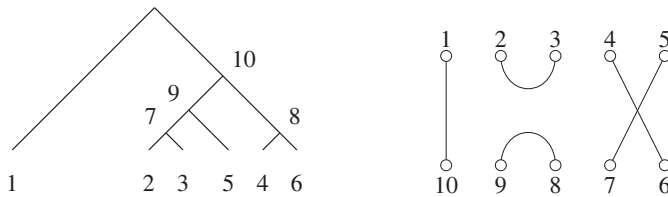


Figure 3. The six-leaf tree $T \in \mathcal{RP}_6$ on the left corresponds to the matching $\{\{1, 10\}, \{2, 3\}, \{4, 6\}, \{5, 7\}, \{8, 9\}\}$, giving the Brauer diagram $\delta(T) \in \mathfrak{B}_5$ shown on the right.

The (associative) product is formed by diagram concatenation, with edges joined via identification of the lower $[n']$ nodes of the first (upper) multiplicand, and the upper $[n]$ nodes of the second (lower) multiplicand. An example is shown in figure 2.

(b) Binary phylogenetic X -trees and their correspondence with Brauer diagrams

Given a binary tree on $n \geq 2$ leaves, once its internal vertices are labelled according to algorithm 1, we immediately have a matching on the set $\{1, \dots, 2n - 2\}$, obtained by pairing labels on sibling vertices [5,6]. The reverse direction—constructing a tree from a matching—begins with laying out the leaf vertices and then pairing matched vertices that are already there, in increasing order.

Informally, the relation to matchings can be seen as the result of the growth of a binary tree via successive bifurcations: starting from a simple tree with two leaves (a single ‘cherry’), with matching $\{1, 2\} = [2]$, one or the other edge suffers a bifurcation, giving for example $\{\{1, 2\}, \{21, 22\}\}$ (a matching of $[4]$), whereupon 2, the parent (still paired to its original partner) becomes an internal node, with the children $\{21, 22\}$ a new pair of nodes. In this way, by iteration, a correspondence is built with matchings, or indeed Brauer diagrams, provided a unique convention for internal node enumeration is given (tantamount simply to having ‘unlabelled’ internal nodes), as has been described above following [5]. The root location is inferred from the last matched pairing.

Through the bijection between binary trees and matchings described in §2, structural features of the Brauer monoid are induced on trees. Each binary tree on n leaves uniquely defines a matching on $[2n - 2]$, and each such matching uniquely defines a Brauer diagram in \mathfrak{B}_{n-1} by connecting nodes labelled by integers paired in the matching (see example in figure 3). Write $\delta(T) \in \mathfrak{B}_{n-1}$ for the Brauer diagram corresponding to $T \in \mathcal{RP}_n^{\text{bin}}$.

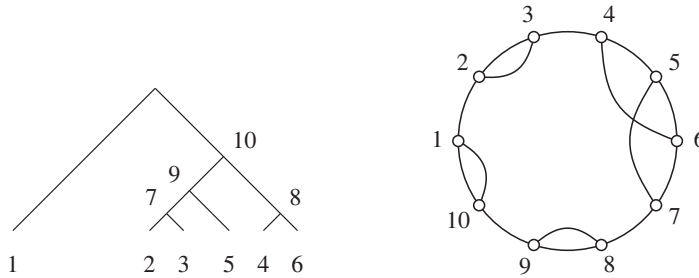


Figure 4. The six-leaf tree $T \in \mathcal{RP}_6^{\text{bin}}$ from figure 3, and its corresponding chord diagram representation.

If the product of two elements $\alpha, \alpha' \in \mathfrak{B}_{n-1}$ is written $\alpha\alpha'$ using juxtaposition, then there is an induced product on trees, defined by

$$T \cdot T' = \delta^{-1}(\delta(T)\delta(T')).$$

Similarly, the $*$ -involution in \mathfrak{B}_{n-1} , defined on pairings by interchanging $i \rightarrow i', i' \rightarrow i$ for all i , or on diagrams by reflection about the horizontal axis, induces an involution on trees,

$$T^* = \delta^{-1}(\delta(T)^*).$$

For example, the tree in figure 3 satisfies $T^* = T$, because its diagram is symmetric about the horizontal axis.

The bijection from trees of course extends to a variety of representations of the Brauer monoid or combinatorial objects tied thereto. From the perspective of matchings of an even set, for example, it is more natural to label elements as $[1, 2(n-1)]$ rather than marking them as $[n-1] \cup [(n-1)']$ (as already done in establishing bijection δ), and a diagram correspondence could simply be established via arcs between a linear arrangement of nodes $\{1, 2, \dots, 2(n-1)\}$. A less biased arrangement is a chord diagram, with arcs linking an even number $2(n-1)$ of nodes arranged in a circle, as in [5]. Relative to a fixed ordering, such a diagram represents a set of transpositions, an element of the symmetric group $\mathfrak{S}_{2(n-1)}$. Thus, the chord diagram in figure 4 corresponding to the six-leaf tree of figure 3, is obtained by bending up the ends of the lower rail of nodes $\{n, n+1, \dots, 2(n-1)\}$ of the corresponding Brauer diagram, so that they join on to the upper rail of nodes $\{1, 2, \dots, (n-1)\}$. Associating a labelled tree with an element of the symmetric group in this way, as a product of transpositions coming from the matching, confers yet another possible multiplicative structure (exploited in the transposition distance for trees [19]).

For manipulations on trees, in this paper, we will use a modified bijection, not between $\mathcal{RP}_n^{\text{bin}}$ and \mathfrak{B}_{n-1} , but between $\mathcal{RP}_n^{\text{bin}}$ and $\mathfrak{B}_{n,n-2}$, one of the equivalent diagrammatic presentations of matchings. In practice, the original δ and the modified bijection $\delta: \mathcal{RP}_n^{\text{bin}} \rightarrow \mathfrak{B}_{n,n-2}$ (for which we use the same symbol) are the same algorithmically, with the difference that the n nodes on the upper edge correspond to the leaves of the n -leaf trees, and the $n-2$ nodes on the lower edge are those reserved for the corresponding internal tree nodes (as mentioned, with the root location being inferred from the last pairing). It is also convenient to continue the numbering from top to bottom in clockwise fashion. See figure 5 for the Brauer transcription in $\mathfrak{B}_{6,4}$ of the six-leaf tree whose presentations in \mathfrak{B}_5 , and as an element of \mathfrak{S}_{10} (via a chord diagram), have been given in figures 3 and 4, respectively.

With these conventions, the setting of trees in $\mathcal{RP}_n^{\text{bin}}$ as a semigroup, afforded by the transcription to \mathfrak{B}_{n-1} , is supplanted by the categorical setting of partial monoids [20–22], where a multiplication on $\mathfrak{B}_{p,q} \times \mathfrak{B}_{r,s}$ exists only if $q = r$ (compare [23]). In practice, we analyse $\mathfrak{B}_{n,n-2}$, and hence $\mathcal{RP}_n^{\text{bin}}$, via left- and right-multiplication by $\mathfrak{B}_n, \mathfrak{B}_{n-2}$, respectively, and the associated equivalence classes. Moreover, via multiplication with the help of intermediate ‘sandwich’ elements belonging to $\mathfrak{B}_{n-2,n} \cong \mathfrak{B}_{n,n-2}^*$, a semigroup structure can indeed be reimposed (see §5).

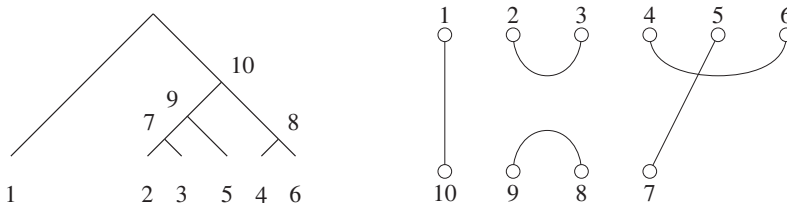


Figure 5. The six-leaf tree $T \in \mathcal{RP}_6^{\text{bin}}$ from figure 3, and its corresponding Brauer element $\alpha = \delta(T) \in \mathfrak{B}_{6,4}$. For this diagram, we have $\text{dom}(\alpha) = \{1, 5\}$, $\text{codom}(\alpha) = \{7, 10\}$, $\text{ker}(\alpha) = \{(2, 3), (3, 2), (4, 6), (6, 4)\}$ and $\text{coker}(\alpha) = \{(8, 9), (9, 8)\}$ and $\text{rank}(\alpha) = 2$.

This turns out to admit a rather universal description, allowing further structural aspects among the participating trees to be distinguished.

We now briefly introduce some language for describing features of a diagram $\alpha \in \mathfrak{B}_{n,n-2}$, that echoes that for functions, as follows. A *block* of a diagram α is a connected set of nodes in α . A *transversal* is an edge that passes between the top row and the bottom row. The *domain* of a diagram α , $\text{dom}(\alpha)$, is the set of points along the top of the transversals, a subset of $[n]$. The *rank* of α is $|\text{dom}(\alpha)|$. The *codomain*, $\text{codom}(\alpha)$, is the set of points along the bottom of the transversals. The kernel $\text{ker}(\alpha)$ and cokernel $\text{coker}(\alpha)$ are defined slightly differently, for technical reasons which will become apparent later

$$\text{ker}(\alpha) = \{(i, j) \in [n] \times [n] : i \text{ and } j \text{ belong to the same block}\}$$

and

$$\text{coker}(\alpha) = \{(i, j) \in [n+1, 2n-2] \times [n+1, 2n-2] : i \text{ and } j \text{ belong to the same block}\}.$$

Note that if $\{i, j\}$ is a block in α with $i, j \in [n]$, then both (i, j) and (j, i) are in $\text{ker}(\alpha)$. An example is given in figure 5. We will occasionally want to refer to the *underlying set* of the kernel or cokernel. By this we mean the set of all points in $[n]$ or $[n+1, 2n-2]$ respectively that appear in relations in the sets. That is, for a set of binary relations S , $U(S) := \{i : (i, j) \in S \text{ for some } j\} \cup \{i : (j, i) \in S \text{ for some } j\}$.

With these definitions, we note the following properties that hold for such a diagram $\alpha \in \mathfrak{B}_{n,n-2}$, that will be used in the sequel

$$\text{dom}(\alpha) \cup U(\text{ker}(\alpha)) = [n]$$

and

$$\text{codom}(\alpha) \cup U(\text{coker}(\alpha)) = [n+1, 2n-2],$$

noting that $\text{dom}(\alpha) \cap U(\text{ker}(\alpha)) = \emptyset = \text{codom}(\alpha) \cap U(\text{coker}(\alpha))$, and also the numerical relations

$$|\text{dom}(\alpha)| = |\text{codom}(\alpha)|$$

and

$$|U(\text{ker}(\alpha))| = |U(\text{coker}(\alpha))| + 2.$$

4. Structuring tree space via Green's relations

As mentioned above, with the identification with phylogenetic trees in $\mathcal{RP}_n^{\text{bin}}$ as Brauer diagrams of type $\mathfrak{B}_{n,n-2}$ via the bijection $T \mapsto \delta(T)$, the correspondence with the monoidal structure in $\mathfrak{B}_{n-1} \equiv \mathfrak{B}_{n-1,n-1}$, afforded by the bijection δ , is no longer direct. Rather, it is supplanted by the categorical setting of partial monoids [20–22], where a multiplication on $\mathfrak{B}_{p,q} \times \mathfrak{B}_{r,s}$ exists only if $q = r$.

In practice, we analyse $\mathfrak{B}_{n,n-2}$, and hence $\mathcal{RP}_n^{\text{bin}}$, via left- and right-multiplication by elements of \mathfrak{B}_n and \mathfrak{B}_{n-2} , respectively, and the associated equivalence classes. It is important to note here

that the labels on the nodes that are matched in the product are implicitly reassigned, along the lines of the description in §3a, so that instead of bottom nodes reading 8, 7, 6 from left to right in an element of $\mathfrak{B}_{5,3}$, we treat them as though labelled 1', 2', 3'. The multiplication on the bottom by an element of \mathfrak{B}_3 then matches 1', 2', 3' along the bottom of one diagram with 1, 2, 3 along the top of the other.

As in the previous usage, we pull the appropriate multiplications back to trees via the bijection

$$\sigma \cdot T = \delta^{-1}(\sigma \delta(T)), \quad \text{for } T \in \mathcal{RP}_n^{\text{bin}}, \sigma \in \mathfrak{B}_n$$

and

$$T \cdot \tau = \delta^{-1}(\delta(T)\tau), \quad \text{for } T \in \mathcal{RP}_n^{\text{bin}}, \tau \in \mathfrak{B}_{n-2}.$$

Examples of the results of multiplication of five-leaf trees by Brauer generators are shown in table 1.

In semigroup theory, elements that can be obtained from each other by left- and/or right-multiplication are classified according to Green's relations (see [18,24]). In the context of phylogenetic trees, we find a *restricted* version of these relations to be most useful, in which the actions are by elements of the symmetric group \mathfrak{S}_n on the left (top of the diagram) and \mathfrak{S}_{n-2} on the right (bottom), as opposed to the full Brauer monoid \mathfrak{B}_n or \mathfrak{B}_{n-2} . We define these restricted adaptations of Green's relations as follows:

Definition 4.1. Green's \mathcal{L} , \mathcal{R} , \mathcal{D} and \mathcal{H} relations are defined as follows, for $\mathfrak{B}_{n,n-2}$:

$$\begin{aligned} \mathcal{L} \text{ classes: } T' \mathcal{L} T &\Leftrightarrow \exists \alpha \in \mathfrak{S}_n : T' = \alpha \cdot T, & [T]_{\mathcal{L}} &= \{T' : \exists \alpha \in \mathfrak{S}_n : T' = \alpha \cdot T\} \\ \mathcal{R} \text{ classes: } T' \mathcal{R} T &\Leftrightarrow \exists \alpha \in \mathfrak{S}_{n-2} : T' = T \cdot \alpha, & [T]_{\mathcal{R}} &= \{T' : \exists \alpha \in \mathfrak{S}_{n-2} : T' = T \cdot \alpha\} \\ \mathcal{D} \text{ classes: } T' \mathcal{D} T &\Leftrightarrow T' \mathcal{L} T \text{ or } T' \mathcal{R} T, & [T]_{\mathcal{D}} &= [T]_{\mathcal{L}} \cup [T]_{\mathcal{R}} \\ \text{and } \mathcal{H} \text{ classes: } T' \mathcal{H} T &\Leftrightarrow T' \mathcal{L} T \text{ and } T' \mathcal{R} T, & [T]_{\mathcal{H}} &= [T]_{\mathcal{L}} \cap [T]_{\mathcal{R}}. \end{aligned}$$

In other words, the \mathcal{L} -class containing T in $\mathfrak{B}_{n,n-2}$ is the set of diagrams that can be reached from T by a left multiplication (by an element of \mathfrak{S}_n). Likewise, the \mathcal{R} -classes arise from right multiplication by elements of \mathfrak{S}_{n-2} . The \mathcal{H} -classes are sets of elements that are both \mathcal{L} and \mathcal{R} -related, and these form the smallest of this family of equivalence classes. By contrast, the \mathcal{D} classes are unions of intersecting \mathcal{L} and \mathcal{R} -classes.

These classes are displayed using arrays called *eggbox diagrams* (see ch. 2 of [18]). Each \mathcal{D} -class can be represented as a rectangular array whose entries are \mathcal{H} -classes. The \mathcal{L} -classes are then given by the rows of the \mathcal{D} -class, and the \mathcal{R} -classes by the columns.

Note that the actions by \mathfrak{S}_n and \mathfrak{S}_{n-2} are not faithful, because for instance the action of a transposition s_i on a cup between i and $i+1$ has the effect of the identity action (for example, in table 1, $s_3 T_2 = T_2$).

Because top actions cannot affect the bottom of a diagram, all diagrams in an \mathcal{L} -class have the same bottom half. And since the action of \mathfrak{S}_n on the top connects all arrangements of the top half of the diagram that have the same rank (\mathfrak{S}_n is the full symmetric group), each \mathcal{L} -class is the full set of diagrams with a particular bottom half.

Likewise, each \mathcal{R} equivalence class is indexed by the common *top* half of the diagrams within it. The \mathcal{H} -classes are then those diagrams that have the same top and bottom halves (the intersections of \mathcal{L} and \mathcal{R} -classes), and there will be $k!$ of these, where k is the rank of the diagrams in the class (this counts the number of ways to join into transversals the set of half-strings that come down from the top half, with those coming up from the bottom half). The \mathcal{D} -classes are unions of \mathcal{L} and \mathcal{R} -classes, namely all diagrams of the same rank.

Figure 6 shows a Green eggbox scheme displaying Brauer diagrams (and corresponding trees) for the \mathcal{D} -classes of the Brauer monoid $\mathfrak{B}_{6,4} \cong \mathcal{RP}_6^{\text{bin}}$ corresponding to six-leaf trees. The class \mathcal{D}_2 with two cherries (rank 2) is shown in detail, to illustrate the coordinatization of trees provided by the Brauer structure. This rank two \mathcal{D}_2 class [20–22] and eggbox represents a total of 540 trees, with six rows (\mathcal{L} -classes) and 45 columns (\mathcal{R} -classes), with 270 intersections (\mathcal{H} -classes, of cardinality

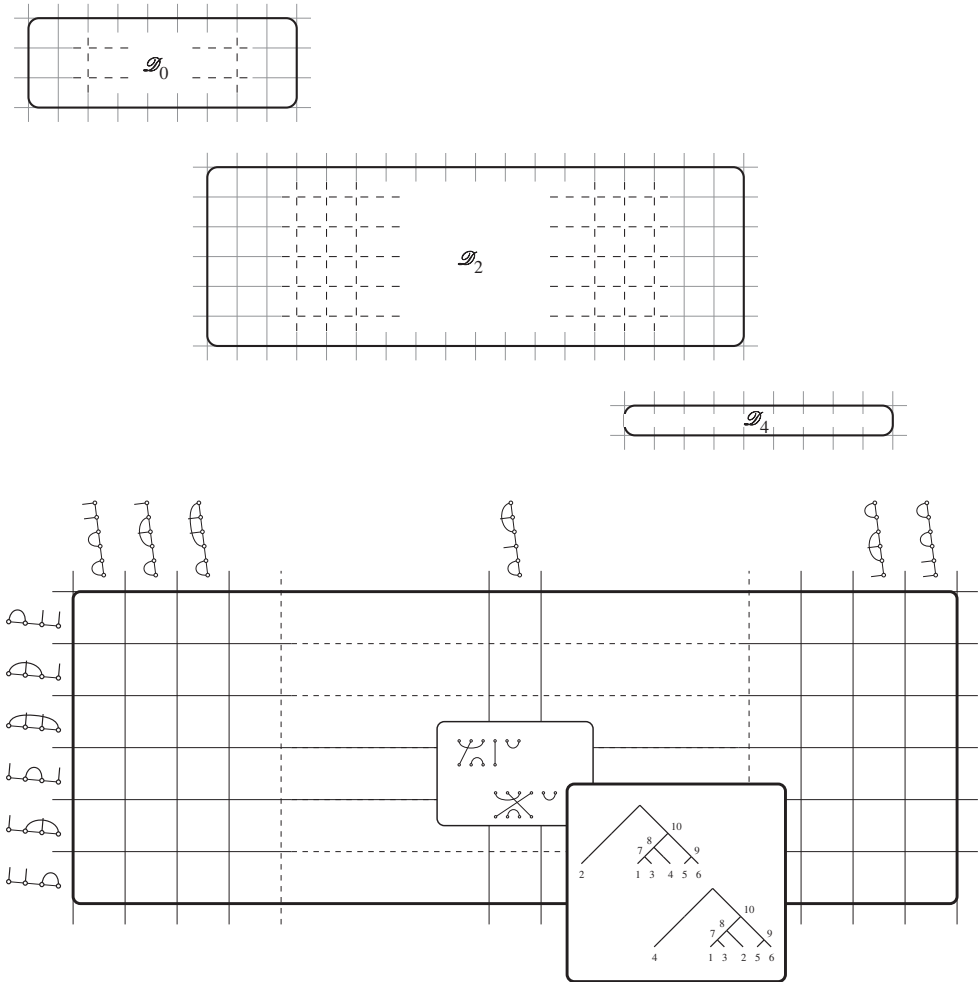


Figure 6. An eggbox diagram illustrating the enumeration of elements of the \mathcal{D} -classes of the Brauer monoid $\mathfrak{B}_{6,4} \cong \mathcal{RP}_6$, representing six-leaf trees. Upper diagram: schematic illustration of the separate eggboxes for the classes $\mathcal{D}_0, \mathcal{D}_2$ and \mathcal{D}_4 , of Brauer ranks $k = 0, 2, 4$, corresponding to trees with 3, 2 or 1 cherry, and comprising 45, 540 and 360 trees, respectively (see text for discussion). Lower diagram: a ‘close-up’ of the \mathcal{D}_2 eggbox for trees with two cherries (and Brauer rank 2). Row labels show bottom halves of Brauer diagrams, representing equivalence classes by left action (on the top of a diagram), the \mathcal{L} -classes, whereas column labels show top-halves of diagrams, representing equivalence classes by right action, the \mathcal{R} -classes. The inset shows two labelled trees, corresponding to the two diagrams belonging to the selected \mathcal{H} -class (the intersection of the selected row and column).

2!). The corresponding eggbox for rank 0 (three cherries), this eggbox (one cherry, rank 2) and the eggbox for rank 4 (caterpillar trees, with one cherry), together enumerate the totality of $|\mathcal{RP}_6^{\text{bin}}| = 9!! = 945 \equiv 45 + 540 + 360$ labelled six-leaf phylogenetic trees on six leaves, respectively.

Remark 4.2. These equivalence classes have natural interpretations in terms of trees. The \mathcal{R} -classes are those trees that are the same up to permutations of leaf labels: their diagrams can be reached by multiplication by an element of \mathfrak{S}_n along the top (which corresponds to the leaves). The \mathcal{L} -classes are those that have the same cherry structure at the leaves, but whose internal vertices have been permuted. The \mathcal{H} -classes represent trees with the same cherries but whose non-cherry leaves have been permuted. Finally, the \mathcal{D} -classes represent all trees with the same number of cherries (namely $\frac{1}{2}(n - k)$, where k is the rank).

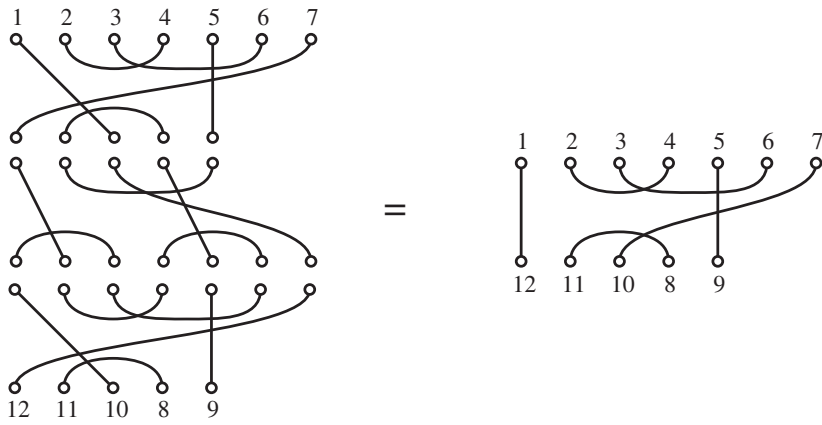


Figure 7. The sandwich product of two Brauer diagrams in $\mathfrak{B}_{7,5}$, relative to a third (in the middle on the left). The corresponding trees are shown in figure 8.

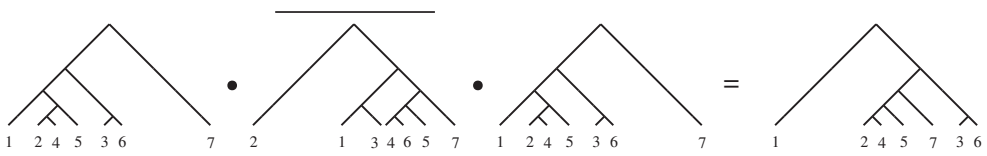


Figure 8. The sandwich product of two trees relative to a third (with the bar over it). This product is computed using the diagram product of the corresponding Brauer diagrams, as shown in figure 7.

5. Multiplicative structure: the set of phylogenetic trees as a semigroup

In using a bijection between trees and an algebraic object like a Brauer monoid, the pay-off is the algebraic structure that comes to the set of trees. In choosing to use the unbalanced diagrams of $\mathfrak{B}_{n,n-2}$, we preserve information about the tree structure (with the leaves all along the top axis of the diagram), but as noted above, we lose the capacity to multiply diagrams in the way that is possible if the top and bottom axes have the same number of nodes.

(a) The sandwich product

There is, nevertheless, still a multiplicative structure available to unbalanced diagrams such as $\mathfrak{B}_{n,n-2}$, namely the *sandwich product*. This requires a fixed diagram (tree) T , and then allows the product of two diagrams T_1 and T_2 to be composed by inverting T (flipping it in the horizontal axis) to obtain a diagram $\bar{T} \in \mathfrak{B}_{n-2,n}$, and sandwiching it between the two trees

$$T_1 *_T T_2 := T_1 \cdot \bar{T} \cdot T_2,$$

where \cdot is composition of diagrams. This product then allows us to define a semigroup relative to T , which we denote $\mathfrak{B}_{n,n-2}^T$. Figure 7 illustrates such a sandwich product in terms of Brauer diagrams in $\mathfrak{B}_{7,5}$ while the induced operation at the level of the corresponding trees is shown in figure 8 (as can be seen, in this case the examples show a sandwich-square, of the form $T *_T T := T \cdot \bar{T} \cdot T$).

Interestingly, the sandwich semigroups relative to different trees are isomorphic if the trees have the same *rank* (the number of transversals in the diagram) [20, Theorem 6.4]. This means that the choice of tree T as sandwich is only important (in terms of semigroup structure) up to its rank.

Note that if a diagram has rank k , then its composition with any other diagram must have rank at most k , because composing diagrams cannot generate additional transversals. In particular, the sandwich semigroup relative to tree T does not contain an identity in general, because no sandwich product with a tree T' of rank greater than $\text{rank}(T)$ can ever return a tree of the same rank as T' . As a consequence, the sandwich semigroup is not a monoid.

(b) The regular subsemigroup relative to a given tree

Let $T \in \mathfrak{B}_{n,n-2}$ be an n leaf tree. Using the sandwich product defined above, the set of trees with operation relative to T is called the sandwich semigroup, denoted $\mathfrak{B}_{n,n-2}^T$. The ‘regular’ elements of this semigroup form a subsemigroup $\text{Reg}(\mathfrak{B}_{n,n-2}^T)$. A semigroup element x is said to be *regular* if there exists an a such that $axa = x$. For the sandwich semigroup, this is equivalent to the property: α is regular if and only if $\bar{T} \cdot \alpha \cdot \bar{T}$ has the same rank as α [20].

The regular elements in this Brauer sandwich semigroup can be characterized as follows (we extract the case relevant to this context).

Proposition 5.1 ([20] proposition 6.13).

$$\text{Reg}(\mathfrak{B}_{n,n-2}^T) = \{\alpha \in \mathfrak{B}_{n,n-2} : \text{coker}(\alpha) \vee \ker(\bar{T}) \text{ separates } \text{codom}(\alpha) \\ \text{and } \ker(\alpha) \vee \text{coker}(\bar{T}) \text{ separates } \text{dom}(\alpha)\}.$$

Here, the *join* of two equivalence relations is their join in the lattice of equivalences, that is, the smallest equivalence relation that contains their union. The join $A \vee B$ *separates* C if each equivalence class in $A \vee B$ contains at most one element of C .

The elements of $\text{Reg}(\mathfrak{B}_{n,n-2}^T)$ can be enumerated according to their rank relative to that of T , as follows:

Theorem 5.2 ([20] corollary 6.17). *The cardinality of the regular subsemigroup of $\mathfrak{B}_{n,n-2}^T$ for any diagram T of rank k , is given by*

$$|\text{Reg}(\mathfrak{B}_{n,n-2}^T)| = \sum_{\substack{0 \leq j \leq k \\ j \equiv k \pmod{2}}} \binom{k}{j}^2 \frac{(k-j-1)!!^2 (n+j-1)!! (n+j-3)!! j!}{(k+j-1)!!^2}.$$

Example 5.3. The tree T whose diagram in $\mathfrak{B}_{6,4}$ is $\{\{1, 10\}, \{2, 9\}, \{3, 4\}, \{5, 8\}, \{6, 7\}\}$ has rank $k = 4$ (figure 9). The sum in theorem 5.2 is over $j = 0, 2, 4$, and can be computed as follows:

$$\begin{aligned} |\text{Reg}(\mathfrak{B}_{6,4}^T)| &= \sum_{j=0,2,4} \binom{4}{j}^2 \frac{(3-j)!!^2 (5+j)!! (3+j)!! j!}{(3+j)!!^2} \\ &= \binom{4}{0}^2 \frac{3!!^2 5!! 3!!}{3!!^2} + \binom{4}{2}^2 \frac{7!! 5!! 2!}{5!!^2} + \binom{4}{4}^2 \frac{9!! 7!! 4!}{7!!^2} \\ &= 45 + 504 + 216 \\ &= 765. \end{aligned}$$

It is interesting to note the number of elements of each rank: 45 of rank 0, 504 of rank 2 and 216 of rank 4. The total number of diagrams of these ranks is, respectively, 45, 540 and 360. In other words, this regular subsemigroup contains all trees of rank 0, 504/540 (93 $\frac{1}{3}$ %) of rank 2 and 216/360 (60%) of rank 4.

We can dig a little further into this counting using the conditions in proposition 5.1. Given that $\ker \bar{T} = \emptyset$ and $\text{coker} \bar{T} = \ker T = \{(3, 4), (4, 3)\}$, the condition for α to be in this subsemigroup are that both: $\text{coker}(\alpha) \vee \emptyset = \text{coker}(\alpha)$ separates $\text{codom}(\alpha)$; and $\ker(\alpha) \vee \{(3, 4), (4, 3)\}$ separates $\text{dom}(\alpha)$. The first of these is trivially satisfied since the underlying sets of codom and coker of α are disjoint (they partition the set $\{7, 8, 9, 10\}$). The second condition is more easily approached by considering when it will *not* hold, namely when the set underlying $\ker(\alpha) \vee \{(3, 4), (4, 3)\}$ has

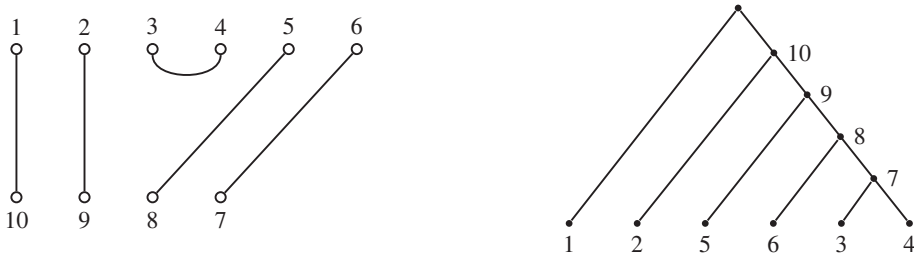


Figure 9. The Brauer diagram and the tree from the partition $\{\{1, 10\}, \{2, 9\}, \{3, 4\}, \{5, 8\}, \{6, 7\}\}$ in example 5.3.

two or more elements in common with $\text{dom}(\alpha)$. Since \ker and dom have disjoint underlying sets, this forces $\{3, 4\} \subseteq \text{dom}(\alpha)$.

These diagrams must have domain of size 2 or 4, and some simple counting gives the number with domain size 2 as 36, and the number of domain size 4 as 144, for a total of 180 diagrams not in the regular subsemigroup. This gives a total number of diagrams of $765 + 180 = 945$, which is indeed the number of rooted trees on six leaves (which is $(2n - 3)!!$, and here $9!! = 945$).

There are several interesting questions related to the regular subsemigroup with respect to a tree, that we will leave for further work. For instance, the regular subsemigroup of T constitutes a type of neighbourhood of T (noting that T can be easily checked using proposition 5.1 to be regular with respect to itself). There are many ways to define a neighbourhood of a tree, for instance using operations on trees like the nearest neighbour interchange (NNI) and subtree prune and regraft moves, or the transposition distance, also based on matchings [19,25]. It would be interesting to know the relationships among these neighbourhoods. Second, we can observe that in the case of example 5.3, all trees of rank 0 are in the regular subsemigroup. Is this a general property? Are there properties of a tree that make it regular with respect to a large proportion of other trees?

6. Non-binary trees and partition diagrams

In this section, we extend the results from binary trees $\mathcal{RP}_n^{\text{bin}}$ to all trees \mathcal{RP}_n , and to forests, taking advantage of a more general family of diagrams and an associated algebraic structure, called a *partition monoid* [20]. We begin with the generalization to trees where the binary constraint is lifted (so that internal vertices and the root may have out-degree greater than 2).

(a) Non-binary trees

Recall that the underlying correspondence for binary trees on $n \geq 2$ leaves is that a tree corresponds to a matching on the set $\{1, 2, \dots, 2n - 2\}$: a partition of the set of non-root vertices into components of size 2. The generalization to non-binary trees maps a tree to a partition of the set of non-root vertices into *subsets of size ≥ 2* (note that the number of non-root vertices will be less than $n - 2$ if the tree is not binary). For this reason, we will again exclude the trivial tree and require $n \geq 2$.

The generalization begins by observing that algorithm 1 applies without change when the input is a tree that is not necessarily binary (as in [6]). We then obtain a partition from a tree by taking a subset to be a set of sibling vertices (having the same parent), and the correspondence for non-binary trees immediately follows the same algorithm as for binary trees (figure 10).

Let Λ_t denote the set of partitions of a set of t objects, and $\Lambda_t^{(\geq 2)}$ the set of those partitions for which all constituent subsets have at least two elements. We will also refer to $\Lambda_t^{(2)}$, the set of those partitions for which all components have size exactly two. Note that $\Lambda_t^{(2)}$ is precisely the set of matchings on t elements (in this case t must be even). Write $\Lambda = \cup_{t > 0} \Lambda_t$ for the set of *all* partitions

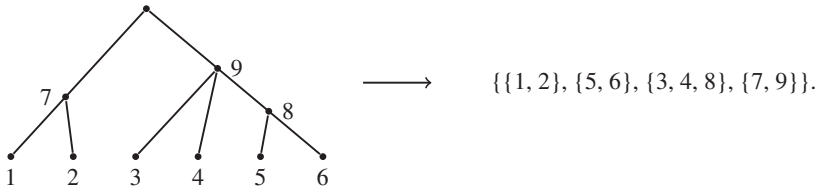


Figure 10. A partition corresponding to a non-binary tree. As with binary trees, non-leaf vertices in the tree are numbered in sequence using algorithm 1, choosing at each point the internal vertex whose children are all numbered and which has the lowest, numbered, child vertex.

of a finite set, and $\Lambda^{(\geq 2)}$ and $\Lambda^{(2)}$ for the corresponding sets when the sizes of components are at least 2 or exactly 2, respectively.

We now introduce partition diagrams, which are generalizations of the Brauer diagrams defined above from partitions that are matchings to more general partitions.

Recall that Brauer diagrams in $\mathfrak{B}_{n,n-2}$ have $2n - 2$ nodes in two rows with n nodes along the top numbered left to right 1 to n , and $n - 2$ nodes along the bottom, numbered right to left $n + 1$ to $2n - 2$. Nodes that are paired in the matching are connected by an edge. A *partition diagram* for a set partition of $[n + m]$ has n nodes along the top numbered from left to right 1 to n , and m nodes along the bottom numbered right to left $n + 1$ to $n + m$. Nodes are connected by edges if their labels are in the same constituent subset of the partition. If there are $k > 2$ nodes in the subset, we do not draw all $\binom{k}{2}$ edges, but instead draw the minimal number to show their common membership, which will be $k - 1$ edges. Examples are shown in figures 11 and 12.

Write $\mathfrak{D}_{n,m}$ for the set of partition diagrams with $n \geq 1$ nodes along the top and $m \geq 0$ along the bottom, and $\mathfrak{D}_{n,m}^{(\geq 2)}$ for the subset in which each partition is from $\Lambda_t^{(\geq 2)}$. The set $\mathfrak{B}_{n,n-2}$ is the subset of $\mathfrak{D}_{n,n-2}^{(\geq 2)}$ in which $m = n - 2$ and all blocks have size exactly 2. That is, $\mathfrak{B}_{n,n-2} = \mathfrak{D}_{n,n-2}^{(2)}$.

For $T \in \mathcal{RP}_n$ (that is, not necessarily binary), write $\delta(T)$ for the corresponding element of $\mathfrak{D}_{n,m}^{(\geq 2)}$, and $\pi(\delta(T))$ for the corresponding set partition of $[n + m]$. Recall that $\ell(\pi)$ is the number of subsets in the partition π of X , and $|\pi|$ is the cardinality of X .

The following lemma is the analogue of the property for diagrams from binary trees (that all satisfy $m = n - 2$).

Lemma 6.1. *If $T \in \mathcal{RP}_n$, then the diagram $\delta = \delta(T) \in \mathfrak{D}_{n,m}^{(\geq 2)}$ satisfies $m = \ell(\pi(\delta)) - 1$.*

Proof. The blocks in the diagram δ correspond to sets of vertices in T that have the same parent in T , therefore, they are in one-to-one correspondence with the set of non-leaf vertices in T . The non-leaf vertices in T are represented by the m numbered nodes along the bottom of the diagram, with the exception of the root of T . Therefore, $m = \ell(\pi(\delta)) - 1$. ■

Note, this result means that given a partition π of an integer $t = n + m$ with blocks of size at least 2, we can compute the values of n and m that give a tree corresponding to π .

Example 6.2. Consider the partition of a set of 12 elements given by

$$\{\{1, 3, 12\}, \{2, 9\}, \{4, 6, 8, 11\}, \{5, 7, 10\}\}.$$

Since there are four blocks, lemma 6.1 implies $m = 3$. The partition diagram and corresponding tree on nine leaves with three internal non-root vertices are shown in figure 11.

While it is clear that each non-binary tree may be expressed as a partition diagram, it is not the case that every partition diagram is obtained from a tree, because some will not satisfy lemma 6.1. The same observation holds, of course, for binary trees.

For example, the diagram in figure 12a does not represent a tree. However, the partition it displays, $\{\{1, 2, 6\}, \{3, 4, 5\}\}$, corresponds to a tree via a diagram that we can find using lemma 6.1 as follows (noting that $t = 6$ and there are two blocks of size 3): $n = 5$ and $m = 1$. The corresponding diagram and tree are shown in figure 12b,c.

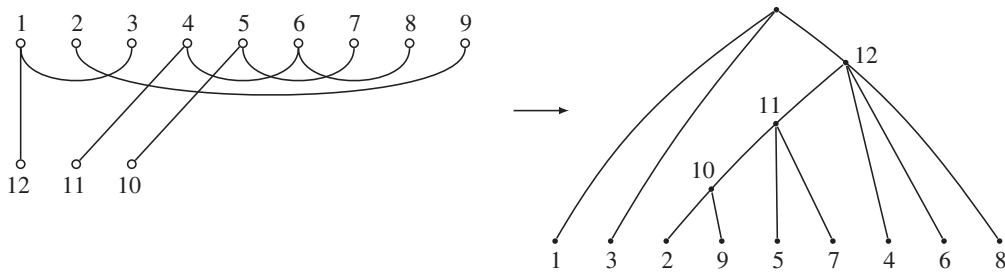


Figure 11. Obtaining a non-binary tree from a partition with components of size at least 2, via a partition diagram. Here, the partition is $\{\{1, 3, 12\}, \{2, 9\}, \{4, 6, 8, 11\}, \{5, 7, 10\}\}$ given in example 6.2.

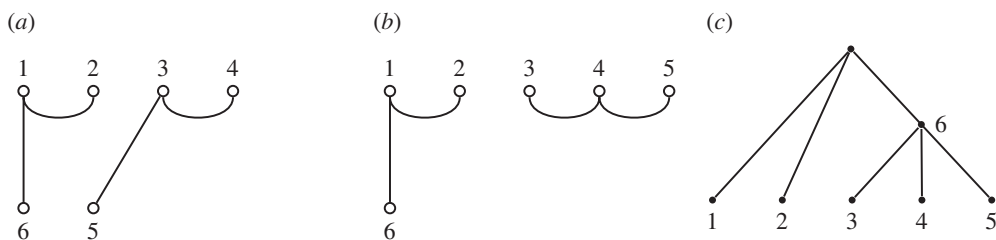


Figure 12. (a) A diagram that does not correspond to a tree, because the corresponding partition $\pi = \{\{1, 2, 6\}, \{3, 4, 5\}\}$ has $\ell(\pi) = 2$ but the diagram does not have $m = \ell(\pi) - 1 = 1$, as required by lemma 6.1. (b) The diagram of the same partition, but with the correct value of $m = 1$. (c) The corresponding tree.

We are now able to generalize the correspondence between binary trees and matchings, to non-binary trees and sets of partitions. Recalling that $\Lambda_t^{(\geq 2)}$ is the set of partitions of a set of t elements into subsets of size ≥ 2 , let

$$\Lambda_{[n]}^{(\geq 2)} = \bigcup_{m=0}^{n-2} \{\pi \in \Lambda_{n+m}^{(\geq 2)} \mid \ell(\pi) - 1 = m\}.$$

Note that by using the fact that $m = |\pi| - n$, this may also be written

$$\Lambda_{[n]}^{(\geq 2)} = \{\pi \in \Lambda^{(\geq 2)} \mid |\pi| - \ell(\pi) + 1 = n\}. \tag{6.1}$$

Theorem 6.3. *There is a 1-1 correspondence between the set of partitions of finite sets into components of size ≥ 2 , and the set of (non-trivial) rooted phylogenetic trees.*

Note, the ‘set of partitions of finite sets’ is not self-referential because the set of such partitions is infinite.

Proof. This result is a direct corollary to theorem 6.5 below. ■

In each direction, a partition diagram may be constructed using the partition and the values of m and n , so we also have as a consequence the following corollary. Let

$$\mathfrak{D}_{[n]}^{(\geq 2)} := \{\alpha \in \mathfrak{D}_{n,m} \mid \pi(\alpha) \in \Lambda_{[n]}^{(\geq 2)}\}. \tag{6.2}$$

Corollary 6.4. *The set \mathcal{RP}_n is in bijection with the set of partition diagrams $\mathfrak{D}_{[n]}^{(\geq 2)}$.*

The correspondence in theorem 6.3 provides the potential for new ways to enumerate the set of rooted phylogenetic trees, by decomposing the set of partitions.

For example, the set of all partitions of ordered sets into blocks of size ≥ 2 is naturally sliced up according to the size of the ordered set, t . In terms of trees on n leaves, this groups them according their number of non-root vertices. In light of the above bijections, trees with

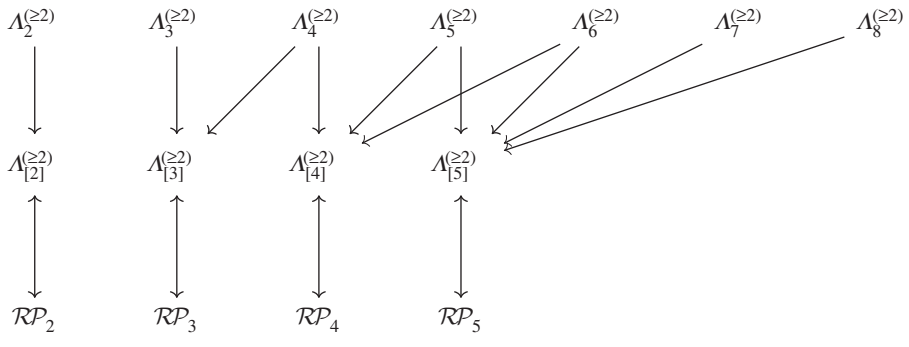


Figure 13. The correspondence between sets of phylogenetic trees and sets of partitions described in theorem 6.3, showing how the sets of partitions decompose the sets of trees. An example of this decomposition for $\Lambda_{[5]}^{(\ge 2)}$ is shown in figure 14.

$$\begin{aligned} \Lambda_{[5]}^{(2)} &:= \bigcup_{m=0}^3 \{ \pi \in \Lambda_{5+m}^{(2)} \mid \ell(\pi) = m + 1 \} \\ &= \{ \pi \in \Lambda_5^{(2)} \mid \ell(\pi) = 1 \} && \begin{array}{c} \circ \circ \circ \circ \circ \end{array} && \begin{array}{c} \diagup \diagdown \diagup \diagdown \diagup \diagdown \end{array} && (1 \text{ tree}) \\ &\cup \{ \pi \in \Lambda_6^{(2)} \mid \ell(\pi) = 2 \} && \begin{array}{c} \circ \circ \circ \circ \circ \\ | \end{array} && \begin{array}{c} \diagup \diagdown \diagup \diagdown \end{array} && (25) \\ &\cup \{ \pi \in \Lambda_7^{(2)} \mid \ell(\pi) = 3 \} && \begin{array}{c} \circ \circ \circ \circ \circ \\ | \quad | \end{array} && \begin{array}{c} \diagup \diagdown \diagup \diagdown \end{array} && (105) \\ &\cup \{ \pi \in \Lambda_8^{(2)} \mid \ell(\pi) = 4 \} && \begin{array}{c} \circ \circ \circ \circ \circ \\ | \quad | \quad | \end{array} && \begin{array}{c} \diagup \diagdown \diagup \diagdown \end{array} && (105) \end{aligned}$$

Figure 14. The decomposition of $\Lambda_{[5]}^{(\ge 2)}$ into sets of partitions. Examples of corresponding partition diagram shapes are in the centre column, with example corresponding tree shapes in the right-hand column. The number of trees in each category is shown on the right, for instance, in the second row, there are $\binom{6}{5} + \frac{1}{2}\binom{6}{3}$ diagrams, and of course the last row is $(2(5) - 3)!!$, giving the Ward numbers [26] [10, Seq. A269939]. Note, these are just examples and there are other possible diagram and tree structures with, for instance, two internal vertices (partitions of [7] into three blocks). This decomposes the set of all trees on five leaves according to the numbers of internal vertices, indicated by the number m of nodes along the bottoms of the diagrams.

particular characteristics such as this are able to be counted via the partial Bell polynomials [16] $B_{t,\ell}(x_1, x_2, \dots, x_{t-\ell+1})$, whose monomial coefficients count the number of set partitions π of $[t]$, with $\ell(\pi) = \ell$ blocks with specific frequencies. Note that $t - \ell + 1 = |\pi| - \ell + 1 = n$, so that these are polynomials in n indeterminates. Thus the above sequence of trees sliced by the size of the ordered set, is given by $\sum_{\ell=1}^t B_{t,\ell}(0, 1, \dots, 1)$, whose first few terms are 1, 1, 4, 11, 41, 162, ... (sequence A000296 of the On-Line Encyclopedia of Integer Sequences [10]), so that for instance there are $|\Lambda_5^{(\ge 2)}| = 11$ partitions of a set of size 5 into partitions without singletons: $\binom{5}{2}$ ways to split into subsets of size 3 and 2 (trees on four leaves with one internal vertex), and one way to have a subset of size 5 (the star tree on five leaves). Similarly, the total number of trees with bifurcations or trifurcations only is the sequence $\sum_{\ell=1}^t B_{t,\ell}(0, 1, 1, 0, \dots, 0)$, whose first few terms are 1, 1, 3, 10, 25, 105, 385, ... (sequence A227937 of [10]).

The former decomposition together with the correspondence in corollary 6.4 can be represented in the diagram in figure 13.

(b) Forests

The correspondence given in §6a between trees and partitions applies to partitions with non-trivial subsets. Recalling that subsets in a partition correspond to sibling vertices in a tree, a natural interpretation for a singleton (trivial) subset is that it corresponds to a vertex with no siblings. Given the definition of a phylogenetic tree used here (and elsewhere) excludes non-root vertices of degree 2, the natural interpretation for a singleton subset is that it corresponds to a root vertex.² And therefore, a diagram with singleton vertices ought to correspond to a forest. Indeed, as in theorem 6.5, forests of phylogenetic trees on n leaves provide a one-to-one correspondence with a set of partitions $\Lambda_{[n]}$ defined in equation (6.3) below.

Let \mathcal{F}_n denote the set of X -forests, that is the set of forests whose leaves are labelled by elements of the set X , with $|X| = n \geq 1$. X -forests are graphs whose connected components are rooted phylogenetic trees, whose leaves partition X . Note that unlike the families of rooted trees, for forests we are allowing $n = 1$.

Write $\ell_{\geq 2}(\pi)$ for the number of non-trivial blocks of the partition $\pi \in \mathfrak{D}_{n,m}$.

Following the definition in equation (6.1), define

$$\Lambda_{[n]} := \{\pi \mid |\pi| - \ell_{\geq 2}(\pi) + 1 = n\} \quad (6.3)$$

and

$$\mathfrak{D}_{[n]} := \{\alpha \in \mathfrak{D}_{n,m} \mid \pi(\alpha) \in \Lambda_{[n]}\}. \quad (6.4)$$

Note, in $\mathfrak{D}_{[n]}$, $m = \ell_{\geq 2}(\pi(\alpha)) - 1$. Let \mathcal{F}_n denote the set of all forests on n leaves, and $\mathcal{F} = \cup_{n \geq 1} \mathcal{F}_n$ the set of all forests.

Theorem 6.5.

- (i) \mathcal{F} is in bijection with the set of partitions Λ ;
- (ii) \mathcal{F}_n is in bijection with the set of partitions $\Lambda_{[n]}$; and
- (iii) The set of non-trivial forests on n leaves with τ components is in bijection with the set $\mathfrak{D}_{n,m}$ of partition diagrams with $\tau - 1$ singleton nodes that satisfy $m = \ell(\pi) - \tau$.

Proof. (1) If $F \in \mathcal{F}_n$ is trivial, so that all its trees are isolated leaves, it will map to the trivial partition of $[n]$, that is, $\{\{1\}, \dots, \{n\}\}$. So we need to prove the correspondence between non-trivial forests and non-trivial partitions.

Each (non-trivial) forest gives a partition, by first numbering non-leaf vertices according to algorithm 1, and then forming sets of sibling vertices, with labelled root vertices forming singletons. As with trees, this algorithm leaves a single root vertex un-labelled.

We now explore the properties of the partition arising from a forest, to help in constructing the map back from partitions to forests. Suppose $\pi = \pi(F)$ is the partition obtained from the forest F .

If F is non-trivial, we have a correspondence between non-leaf vertices in F , and non-trivial blocks, given by the children of each non-leaf vertex. It follows that the number of non-leaf vertices in F is precisely $\ell_{\geq 2}(\pi)$.

The number of all vertices in a non-trivial forest F is $|\pi| + 1$, because one vertex (one of the roots) is left unlabelled by algorithm 1. The vertices in F are also either leaves or non-leaves, and so this number is also equal to $n + \ell_{\geq 2}(\pi)$. Therefore,

$$|\pi| + 1 = n + \ell_{\geq 2}(\pi)$$

and so $n = |\pi| - \ell_{\geq 2}(\pi) + 1$.

²Note that Erdős & Székely [6] do not strictly consider phylogenetic trees as we define them here, in that they allow internal vertices of degree 2 in their trees. Hence, their correspondence gives trees rather than forests.

Now consider a non-trivial partition $\pi \in \Lambda$. We will show how a forest can be constructed from π . Set $n = |\pi| - \ell_{\geq 2}(\pi) + 1$, and create a starting forest F_0 consisting of n vertices as leaves, labelled $1, \dots, n$. We will successively add vertices and edges to the forest as follows.

First, set $\pi_0 = \pi$, and consider the non-trivial sets in π_0 . We have assumed that $\pi_0 = \pi$ is non-trivial, so there will be at least one. We claim that at least one of these is contained in $\{1, \dots, n\}$. Observe that there are $|\pi_0| - n = \ell_{\geq 2}(\pi_0) - 1$ integers outside $\{1, \dots, n\}$ and included in π_0 . But there are $\ell_{\geq 2}(\pi_0)$ non-trivial subsets, and so at least one cannot contain an element outside $\{1, \dots, n\}$, as required.

Let $\hat{\pi}_0$ be the set of non-trivial subsets in π_0 contained in $\{1, \dots, n\}$ (that label vertices in F_0), and let S_0 be the element of $\hat{\pi}_0$ containing the least integer. This is well-defined because, as argued, $\hat{\pi}_0$ is non-empty.

Add a vertex v_1 to F_0 , as a parent to the vertices labelled by the elements of S_0 , to create a new forest F_1 . Then remove S_0 from π_0 to define

$$\pi_1 := \pi_0 \setminus S_0.$$

If π_1 has no non-trivial subsets, then end the algorithm and output F_1 . (In this case, we will have had $\ell_{\geq 2}(\pi) = 1$, and so $|\pi| = n$ and all integers in π are labelling vertices in F_1 .)

Otherwise, label the vertex v_1 in F_1 by $n + 1$. F_1 has leaves labelled $1, \dots, n$ and one other vertex labelled $n + 1$, which is the parent of at least two of the leaves.

As before, we claim that π_1 contains a set that is a subset of $\{1, \dots, n, n + 1\}$, and the argument naturally extends as follows:

- π_1 has $\ell_{\geq 2}(\pi_1) = \ell_{\geq 2}(\pi) - 1 \neq 0$ non-trivial subsets;
- There are $|\pi| - (n + 1) = \ell_{\geq 2}(\pi) - 2$ integers in π outside of $\{1, \dots, n + 1\}$; and
- Therefore, it is not possible for all non-trivial subsets in π_1 to include an element outside $\{1, \dots, n + 1\}$.

Thus, the set $\hat{\pi}_1$ of non-trivial subsets in π_1 contained in $\{1, \dots, n + 1\}$ is non-empty. Choose the subset in $\hat{\pi}_1$ with the least integer, and call it S_1 .

This process can continue, as described in algorithm 2, until we reach a point where π_i has no non-trivial subsets and we output the resulting forest. The forest will have n leaves, and $i = \ell_{\geq 2}(\pi)$ additional vertices, of which $i - 1$ will have labels: one root vertex will remain unlabelled. Thus all elements of π will be labelling vertices, since $|\pi| = n + \ell_{\geq 2}(\pi) - 1$.

Note that singletons in the partition π are also labelling vertices in the forest. Any singleton in π that is $\leq n$ is already labelling an isolated leaf from the outset, and so represents a trivial tree. Any singleton greater than n will be labelling a root vertex in a tree in the forest, because it will be assigned as a parent of vertices in a block, and will not be assigned a parent because it has no other elements in its block. See example 6.6 for an illustration of these observations.

The process deterministically defines a forest on n leaves whose vertices (except one root) are labelled by the elements of the partition π , and completes the proof of (1).

(2) immediately follows from the construction described above, which gives a correspondence between a forest with n leaves and a partition satisfying the condition to be an element of $\Lambda_{[n]}$.

For (3), suppose F has n leaves and τ component trees, and is non-trivial. Labelling the vertices according to algorithm 1, all but one of these trees will have a labelled root, and so the corresponding partition will have $\tau - 1$ singletons. As noted above, the number of non-leaf vertices is $\ell_{\geq 2}(\pi)$, because they correspond to sets of siblings, which correspond to blocks of the partition. This includes the single non-labelled root vertex, and so the number of non-leaf labels in the partition for F is $\ell_{\geq 2}(\pi) - 1$, and this is the value of m in the partition diagram. But since the partition has $\tau - 1$ singleton sets, we have $\ell(\pi) = \ell_{\geq 2}(\pi) + (\tau - 1)$, and it follows that $m = \ell(\pi) - \tau$ as required. The reverse direction takes a diagram to a partition that then constructs a forest according to algorithm 2. The conditions in the statement follow. ■

Algorithm 2. Construction of a forest from a partition.**Require:** $\pi \in \Lambda$ $n \leftarrow |\pi| - \ell_{\geq 2}(\pi) - 1$ $F_0 \leftarrow$ the trivial forest on leaves $\{1, \dots, n\}$ $\pi_0 \leftarrow \pi$ $\hat{\pi}_0 \leftarrow$ the subsets of π_0 of size ≥ 2 whose elements are labels in F_0 $S_0 \leftarrow$ the subset in $\hat{\pi}_0$ containing the least integer**for** $i > 0$ **do** Create a vertex v_i and edges $E_i := \{(v_i, s)\}$ for vertices s in F_{i-1} labelled by S_{i-1} $F_i \leftarrow F_{i-1}$ with the additional vertex v_i and edges E_i $\pi_i \leftarrow \pi_{i-1} \setminus S_{i-1}$ **if** π_i contains a non-trivial subset **then** Label v_i by $n+i$. **else** $F \leftarrow F_i$ Break ▷ Go to Output statement **end if** $\hat{\pi}_i \leftarrow$ subsets in π_i of size ≥ 2 labelling vertices in F_i . $S_i \leftarrow$ the subset in $\hat{\pi}_i$ with the lowest integer**end for**Output F .

Example 6.6 (Construction of a forest from a partition via algorithm 2). Take the partition π given by

$$\pi = \{\{1, 10\}, \{3, 4, 6\}, \{5\}, \{7, 8\}, \{9, 13\}, \{2, 12\}, \{11\}\}.$$

Here, we have $|\pi| = 13$, $\ell_{\geq 2}(\pi) = 5$, and $n = |\pi| - \ell_{\geq 2}(\pi) + 1 = 9$. Set F_0 to be the forest with nine isolated leaves labelled $1, \dots, 9$, and set $\pi_0 = \pi$. We have $\hat{\pi}_0 = \{\{3, 4, 6\}, \{7, 8\}\}$, and so $S_0 = \{3, 4, 6\}$.

Let F_1 be the forest obtained from F_0 by adding a vertex v_1 that is a parent of the vertices labelled 3, 4, 6. Set $\pi_1 = \pi_0 \setminus S_0 = \{\{1, 10\}, \{5\}, \{7, 8\}, \{9, 13\}, \{2, 12\}, \{11\}\}$. Since π_1 has non-trivial subsets we label v_1 by 10, and $\hat{\pi}_1 = \{\{1, 10\}, \{7, 8\}\}$, with $S_1 = \{1, 10\}$.

Let F_2 be the forest obtained from F_1 by adding a vertex v_2 that is a parent of the vertices labelled 1, 10. Set $\pi_2 = \pi_1 \setminus S_1 = \{\{5\}, \{7, 8\}, \{9, 13\}, \{2, 12\}, \{11\}\}$. Since π_2 has non-trivial subsets we label v_2 by 11, and $\hat{\pi}_2 = \{\{7, 8\}\}$, with $S_2 = \{7, 8\}$.

Let F_3 be the forest obtained from F_2 by adding a vertex v_3 that is a parent of the vertices labelled 7, 8. Set $\pi_3 = \pi_2 \setminus S_2 = \{\{5\}, \{9, 13\}, \{2, 12\}, \{11\}\}$. Since π_3 has non-trivial subsets we label v_3 by 12, and $\hat{\pi}_3 = \{\{2, 12\}\}$, with $S_3 = \{2, 12\}$.

Let F_4 be the forest obtained from F_3 by adding a vertex v_4 that is a parent of the vertices labelled 2, 12. Set $\pi_4 = \pi_3 \setminus S_3 = \{\{5\}, \{9, 13\}, \{11\}\}$. Since π_4 has non-trivial subsets we label v_4 by 13, and $\hat{\pi}_4 = \{\{9, 13\}\}$, with $S_4 = \{9, 13\}$.

Let F_5 be the forest obtained from F_4 by adding a vertex v_5 that is a parent of the vertices labelled 9, 13. Set $\pi_5 = \pi_4 \setminus S_4 = \{\{5\}, \{11\}\}$. Since π_5 has no non-trivial subsets we end the algorithm and output $F = F_5$.

The forests generated in this example are shown in figure 15.

Note, theorem 6.3, relating to non-binary trees, is a special case of this theorem for those forests that consist of a single tree, and partitions without singletons.

The correspondence with forests given in theorem 6.5 creates a broad set of correspondences between sets of partitions and sets of phylogenetic objects, as shown in figure 16. Recall that we have defined the following sets:

$$\Lambda_n^* = \text{Set of partitions of a set of cardinality } n,$$

whose components satisfy the size condition \star .

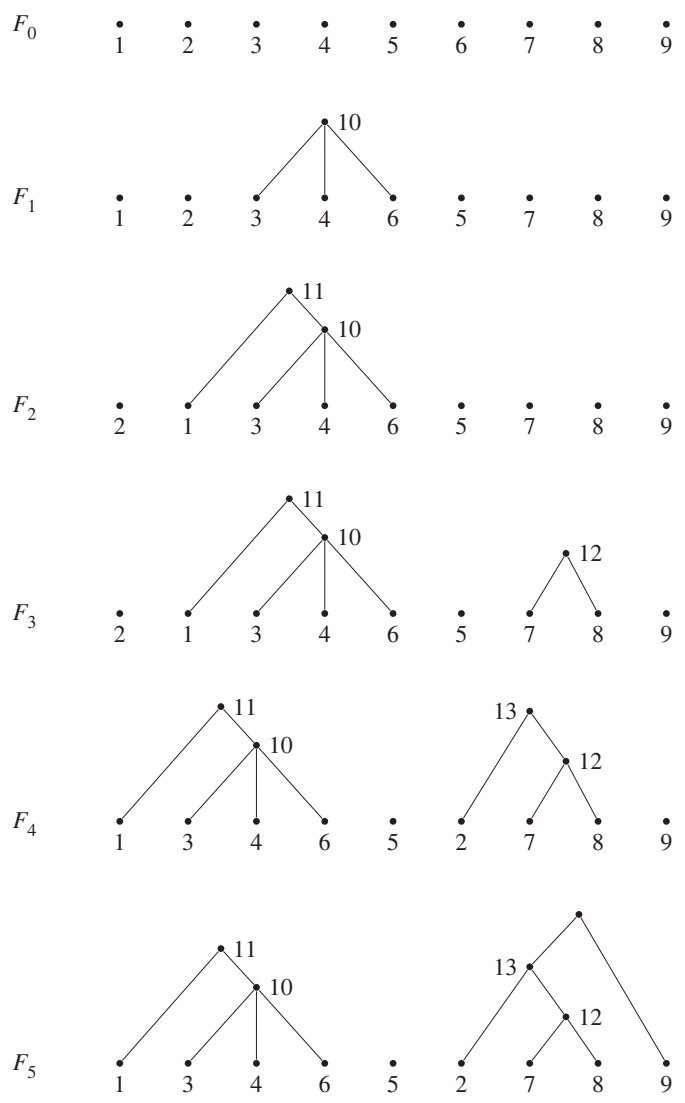


Figure 15. The forests generated using algorithm 2, to obtain a forest from the partition $\pi = \{\{1, 10\}, \{3, 4, 6\}, \{5\}, \{7, 8\}, \{9, 13\}, \{2, 12\}, \{11\}\}$, as described in example 6.6, outputting $F = F_5$.

$$\mathfrak{D}_{n,m}^* = (n, m) \text{ partition diagrams from partitions in } \Lambda_{n+m}^*.$$

$$\Lambda_{[n]}^* = \{\pi \in \Lambda^* \mid |\pi| - \ell(\pi) + 1 = n\}.$$

and

$$\mathfrak{D}_{[n]}^* = \{\alpha \in \mathfrak{D}_{n,|\pi|-n}^* \mid \pi(\alpha) \in \Lambda_{[n]}^*\}.$$

Here \star might be empty (no restrictions), or (2) or (≥ 2) , meaning components must be of size 2 or at least 2. But in the next two corollaries, we see correspondences for when components have size at most 2.

Corollary 6.7. *The set of binary forests is in bijection with the set of partitions $\Lambda^{(\leq 2)}$: those whose subsets have size at most 2.*

Corollary 6.8. *The set of binary forests on n leaves is in bijection with the set of partitions $\Lambda_{[n]}^{(\leq 2)}$.*

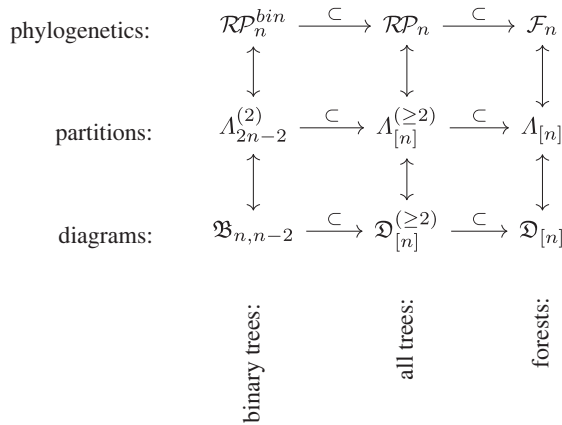


Figure 16. Correspondences between sets of trees or forests, all on n leaves, sets of partitions and partition diagrams. In the left column, the components of the partition are all size exactly 2, and so $m = n - 2$.

(c) Semigroup structure

The semigroup structures that we have described for binary trees in §5, also extend to non-binary trees or forests, with some caveats. For instance, immediately we see that an action by the Temperley–Lieb generators e_i will not be able to be defined for non-binary trees in general: the product of e_i by a partition diagram that includes a block $\{i, i + 1, j\}$, where j is on the opposite side to i and $i + 1$, will result in a block of size one, namely $\{j\}$. So the action of a Temperley–Lieb generator on a non-binary tree may result in a forest.

However, action by the symmetric group generators s_i does preserve the restriction on the partition diagram. In fact, action by s_i preserves the number and size of the blocks in the partition. Therefore, we are still able to construct the eggbox diagram decomposition of the set of non-binary trees, and indeed for forests. This decomposition will have a richer structure, however, as the \mathcal{D} -classes are not determined simply by rank, but by other factors (the number and size of the blocks in the partition). This is a topic we leave for further investigation.

The sandwich semigroup construction allows multiplication of non-binary trees, as it does in the binary case, although in general the product will not be closed. That is, if one of the trees involved in the sandwich product is not binary, it is possible that the product of diagrams results in an isolated node, which means the diagram may correspond to a forest of more than one tree (figure 17). Thus, the set of diagrams for non-binary trees is not closed under the sandwich product. Whether there are subfamilies of non-binary trees for which the product is defined is an interesting further question.

Likewise, the sandwich product allows the multiplication of two forests relative to a third. As with non-binary trees, there will be some products that are not defined, because the numbers of nodes in the diagrams do not match (if the numbers of non-leaf vertices in the forest are not equal). But additionally, even if this is satisfied so that the diagram product is defined, the product may not result in a forest. And this also applies to sandwich products of non-binary trees: the product of two non-binary trees, or forests, relative to a third, may result in a diagram that does not even represent a forest, in that it violates the condition in theorem 6.5(iii). An example is shown in figure 18.

7. Discussion

The link between phylogenetic trees and algebraic structures such as Brauer monoids and partition monoids that we have described gives rise to a wealth of questions that need further exploration.

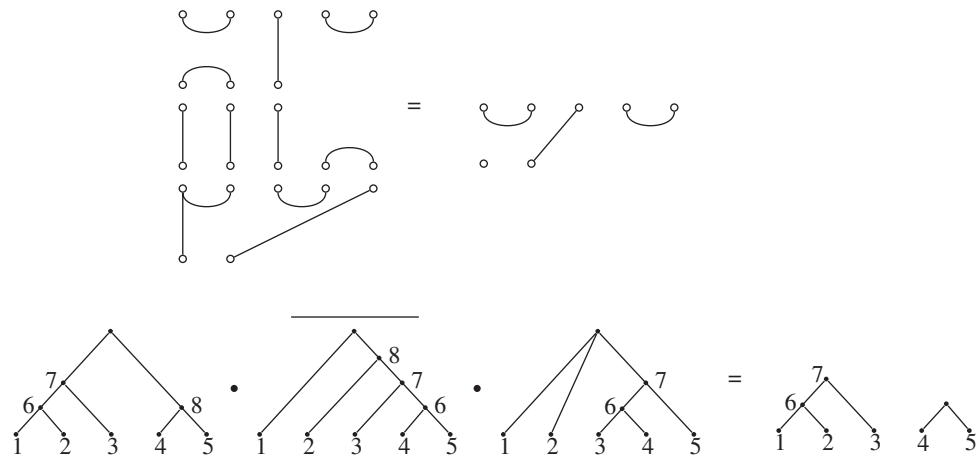


Figure 17. Top: a sandwich product involving a diagram from a non-binary tree that results in a diagram that does not correspond to a tree, because it has an isolated node, but to a forest with two components. Bottom: the same sandwich product showing the trees and forest involved.

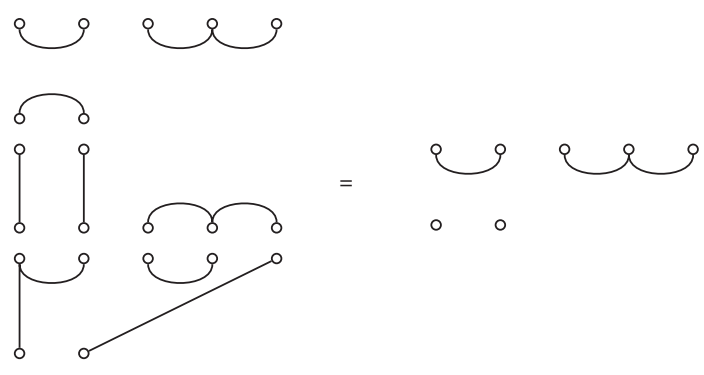


Figure 18. A sandwich product involving diagrams from two non-binary trees relative to a third, and which results in a diagram that does not define a tree or a forest, because it violates the condition in theorem 6.5(iii). This product diagram has $m = 2$, $\tau = 3$ (one more than the number of singleton nodes) and $\ell(\pi) = 4$, so does not satisfy $m = \ell(\pi) - \tau$.

To begin, some opportunities for further development were raised by Holmes and Diaconis in 1998 [5]. For instance, they suggest the representation of a tree as a matching that they introduced (and treating the matched pairs as 2-cycles) allows for the use of multiplication in the symmetric group to create a random walk in tree space. They also suggested the use of multiplication of matchings through use of the Brauer algebra, that we have developed further in this paper in a direction they perhaps did not anticipate (by preserving the leaves along the top of a diagram to create an unbalanced but biologically interpretable model).

The developments here, representing the matching (or partition more generally) as a Brauer or partition diagram, allow a more targeted random walk to be defined that preserves certain key structures of the trees. That is, a random step can be performed by acting on the top or bottom of a diagram by an element of \mathfrak{S}_n or \mathfrak{S}_m , and allow movement along an \mathcal{R} -class or \mathcal{L} -class, which preserves certain structural properties of the tree, as described in §4.

There are many further questions that warrant exploration, some of which we list here

- (i) Questions about products of trees in the sandwich semigroup

- (a) We have defined a way to multiply two binary trees relative to a third, via the sandwich product. How do features of trees (such as their balance) behave when they are multiplied together?
- (b) Are some properties of trees ‘closed’ under multiplication? If two trees share the same property, when does their product share the same property?
- (c) A feature of the sandwich semigroup construction is that it creates a product *relative* to a fixed tree. Is there some way to exploit this feature in phylogenetics? For instance, is it possible that gene trees may be constrained to be inside some neighbourhood defined by the product relative to the species tree (see [27] for discussion of the challenges understanding this relationship)?
- (d) The sandwich semigroup product works naturally for binary trees, whose Brauer diagrams have predictable numbers of nodes. Given the comments in §6b and figure 18, are there subclasses of forests (other than binary trees) for which the product is defined and closed?
- (ii) Questions about the regular subsemigroup
- (a) To what extent are features of trees preserved within the regular subsemigroup corresponding to the tree? Using balance as an example again, are the trees in the subsemigroup all close to being balanced under one of the standard balance measures?
- (b) How do topological operations on a tree, such as the NNI, interact with the structures described here? For instance, is the NNI neighbourhood of a tree contained in the regular subsemigroup of the sandwich semigroup at T ? Or as mentioned at the end of §5, are trees of rank 0 (with maximal cherries) always in the regular subsemigroup?
- (iii) Other links to tree space
- (a) As noted in the Introduction, Holmes and Diaconis raised the prospect of using the product of trees to randomly move around tree space or search for an optimal solution to a problem. The properties of this random walk, or one extended to non-binary trees or forests, are unknown.
- (b) If we restrict attention only to Brauer diagrams that are *planar*, that is, for which there are no crossings of lines, then we obtain another closed algebraic structure. That is, in the sandwich semigroup corresponding to a tree with planar diagram, the product of two planar diagrams will remain planar. Any diagram can be made planar by relabelling the leaves, so would this semigroup describe some sort of canonical representatives of tree space?
- (c) Can operations on trees, such as edge-cutting that take a tree and produce a forest (but also even tree rearrangement operations such as NNI), be implemented by a sandwich product, or an action by an element of the partition monoid? For instance, can sandwich products such as that seen in figure 17 be controlled systematically?
- (d) There is an interesting correspondence between ‘augmented perfect matchings of $[2n - 2]$ containing ℓ wiggly lines’ and phylogenetic trees on n leaves and $n - \ell - 1$ internal and root vertices, recently described in [28]. It would be interesting to investigate how such augmented matchings link to diagrams, and the set of partitions $\Lambda_{[n]}^{(\geq 2)}$ described here for such trees in theorem 6.3.
- (iv) Other links to semigroup theory and the Brauer algebra
- (a) Within semigroup theory, and indeed broadly within other algebraic areas, the idempotents (elements x for which $x^2 = x$) play a very important role. What are the idempotents within the sandwich semigroup, and what relationships do the corresponding trees share? A formula for the number of idempotents in the sandwich semigroup is known [20, Theorem 6.18]: for instance, for a six-leaf tree with three cherries (rank 0), there are 45 idempotents in the corresponding

sandwich semigroup (which is also the number of rank 0 trees on six leaves). A characterization could be illuminating. Narrowing it down further, the mid-identities (elements that are idempotents when regular but satisfying additional conditions) could have a concrete phylogenetic relationship to the original tree.

- (b) It would be interesting to study further the restricted Green's relations in which the action on left and right is from a subgroup or subsemigroup (like our \mathfrak{S}_n action in §5).
- (c) The general relations in the Brauer algebra (see appendix A) contain a central generator λ that we ignore by setting it to be 1. In the algebra, it tracks loops that can occasionally be generated by diagram concatenation, as occurs in the example in figure 2. Is there phylogenetically relevant information that can be captured by loops arising in products, and that might benefit from use of the full $\mathfrak{B}_n(\lambda)$?

There are several opportunities to extend the ideas in this paper in different directions. For instance, is it possible to represent phylogenetic *networks* within a diagram semigroup framework? And coming from the algebraic point of view, is there a role for other monoidal and categorical systems, such as those described in [20] (compare also Loday [29]) to play within phylogenetics, or other scientific and combinatorial problems?

Clearly, there is opportunity for exploration and development of this approach at an algebraic, combinatorial, phylogenetic and computational level. Diagram semigroups, monoids, algebras and categories have found numerous diverse and powerful applications within mathematics and physics, and it is exciting to think that they may open new doors to phylogeneticists.

Data accessibility. This article has no additional data.

Conflict of interest declaration. We declare we have no competing interests.

Authors' contributions. A.F.: conceptualization, formal analysis, investigation, methodology, writing—original draft, writing—review and editing; P.D.J.: conceptualization, formal analysis, investigation, methodology, writing—original draft, writing—review and editing.

Both authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Funding. The authors were both supported in this research by Australian Research Council Discovery Grant DP180102215.

Acknowledgements. When the authors began exploring the potential for modelling phylogenetic trees with Brauer diagrams, they made the fortunate decision to ask James East for advice. They are tremendously grateful for his patient explanations of ideas such as the sandwich semigroup, and for pointing them to important references. They are also grateful to the referees for their careful reading, suggestions of important additional references and instructive comments that have greatly improved the presentation of the results.

Appendix A. Presentation for the Brauer algebra

For each $n = 1, 2, \dots$, the Brauer algebra $\mathfrak{B}_n(\lambda)$ is the complex algebra generated by the elements $\{s_1, s_2, \dots, s_{n-1}, e_1, e_2, \dots, e_{n-1}\}$ subject to the defining relations [11]

$$\begin{aligned} s_i^2 &= 1, & s_i s_{i+1} s_i &= s_{i+1} s_i s_{i+1}, \\ s_i s_j &= s_j s_i & \text{for } |i - j| > 1; \\ e_i^2 &= \lambda e_i, & e_i e_{i\pm 1} e_i &= e_i, \\ s_i s_{i\pm 1} e_i &= e_{i\pm 1} e_i, & e_i s_{i\pm 1} s_i &= e_i e_{i\pm 1}, \\ s_i e_j &= e_j e_i, & s_i s_j &= s_j s_i, & s_i e_j &= e_j s_i, & \text{for } |i - j| > 1; \\ s_i e_i &= e_i s_i = e_i, & e_i s_{i\pm 1} e_i &= e_i. \end{aligned}$$

For the generalization of these relations arising from the connection with binary trees, to the full partition algebra underlying the extension to non-binary trees and forests, see [30,31].

1. Semple C, Steel M. 2003 *Phylogenetics*, vol. 24. Oxford, UK: Oxford University Press on Demand.
2. Steel M. 2016 *Phylogeny: discrete and random processes in evolution*. Philadelphia, PA: SIAM.
3. Liu P. 2021 A tree distinguishing polynomial. *Discrete Appl. Math.* **288**, 1–8. (doi:10.1016/j.dam.2020.08.019)
4. Liu P, Biller P, Gould M, Colijn C. 2020 Polynomial phylogenetic analysis of tree shapes. *bioRxiv*. (doi:10.1101/2020.02.10.942367)
5. Diaconis PW, Holmes SP. 1998 Matchings and phylogenetic trees. *Proc. Natl Acad. Sci. USA* **95**, 14 600–14 602. (doi:10.1073/pnas.95.25.14600)
6. Erdős PL, Székely LA. 1989 Applications of antilexicographic order. I. An enumerative theory of trees. *Adv. Appl. Math.* **10**, 488–496. (doi:10.1016/0196-8858(89)90026-2)
7. Erdős PL. 1993 A new bijection on rooted forests. *Discrete Math.* **111**, 179–188. (doi:10.1016/0012-365X(93)90154-L)
8. Klarner DA. 1970 Correspondences between plane trees and binary sequences. *J. Comb. Theory* **9**, 401–411. (doi:10.1016/S0021-9800(70)80093-X)
9. Read RC. 1968 The use of S-functions in combinatorial analysis. *Can. J. Math.* **20**, 808–841. (doi:10.4153/CJM-1968-080-x)
10. OEIS Foundation Inc. 2022 The on-line encyclopedia of integer sequences. Published electronically at <https://oeis.org>.
11. Brauer R. 1937 On algebras which are connected with the semisimple continuous groups. *Ann. Math. (2)* **38**, 857–872. (doi:10.2307/1968843)
12. Jones VFR. 1994 The Potts model and the symmetric group. In *Subfactors (Kyuzeso, 1993)* (eds H Araki, H Kosaki, Y Kawahigashi), pp. 259–267. World Scientific.
13. Martin P. 1994 Temperley-Lieb algebras for non-planar statistical mechanics—the partition algebra construction. *J. Knot Theory Ramif.* **3**, 51–82. (doi:10.1142/S0218216594000071)
14. Felsenstein J. 1978 The number of evolutionary trees. *Syst. Biol.* **27**, 27–33. (doi:10.1093/sysbio/27.4.401)
15. Felsenstein J. 2004 *Inferring phylogenies*. Sunderland, MA: Sinauer Associates.
16. Stanley. RP. 1986 *Enumerative combinatorics*, vol. I. The Wadsworth & Brooks/Cole Mathematics Series. Monterey, CA: Cole Advanced Books & Software.
17. Everitt B. 2021 The sympathetic sceptic’s guide to semigroup representations. *Expo. Math.* **39**, 197–237. (doi:10.1016/j.exmath.2021.03.002)
18. Howie JM. 1995 *Fundamentals of semigroup theory*, vol. 12. Oxford, UK: Oxford University Press.
19. Valiente. G. 2005 A fast algorithmic technique for comparing large phylogenetic trees. In *SPIRE’05: Proc. 12th Int. Conf. on String Processing and Information Retrieval, November*, pp. 370–375. Lecture Notes in Computer Science, vol. 3772. Berlin, Heidelberg, Germany: Springer.
20. Dolinka I, Đurđev I, East J. 2021 Sandwich semigroups in diagram categories. *Int. J. Algebra Comput.* **31**, 1339–1404. (doi:10.1142/S021819672150048X)
21. Dolinka I, Đurđev I, East J, Honyam P, Sangkhanan K, Sanwong J, Sommanee W. 2018 Sandwich semigroups in locally small categories I: foundations. *Algebra Univers.* **79**, 75. (doi:10.1007/s00012-018-0537-5)
22. Dolinka I, Đurđev I, East J, Honyam P, Sangkhanan K, Sanwong J, Sommanee W. 2018 Sandwich semigroups in locally small categories II: transformations. *Algebra Univers.* **79**, 76. (doi:10.1007/s00012-018-0539-3)
23. Morton H, Traczyk P. 1990 Knots and algebras. In *Contribuciones matemáticas en homenaje al profesor D. Antonio Plans Sanz de Bremond* (eds E Martin-Peinador, A Rodez Usan), pp. 201–220. Zaragoza, Spain: University of Zaragoza.
24. Green JA. 1951 On the structure of semigroups. *Ann. Math.* **54**, 163–172. (doi:10.2307/1969317)
25. Alberich R, Cardona G, Rosselló F, Valiente G. 2009 An algebraic metric for phylogenetic trees. *Appl. Math. Lett.* **22**, 1320–1324. (doi:10.1016/j.aml.2009.03.003)
26. Ward M. 1934 The representation of Stirling’s numbers and Stirling’s polynomials as sums of factorials. *Am. J. Math.* **56**, 87–95. (doi:10.2307/2370916)
27. Szöllösi GJ, Tannier E, Daubin V, Boussau B. 2015 The inference of gene trees with species trees. *Syst. Biol.* **64**, e42–e62. (doi:10.1093/sysbio/syu048)

28. Price AE, Sokal AD. 2020 Phylogenetic trees, augmented perfect matchings, and a Thron-type continued fraction (T -fraction) for the Ward polynomials. *Electron. J. Comb.* **27**, P4.6. (doi:10.37236/9571)
29. Loday J-L. 2002 Arithmetree. *J. Algebra* **258**, 275–309. Special issue in celebration of Claudio Procesi's 60th birthday. (doi:10.1016/S0021-8693(02)00510-0)
30. Halverson T, Ram A. 2005 Partition algebras. *Eur. J. Comb.* **26**, 869–921. Combinatorics and Representation Theory. (doi:10.1016/j.ejc.2004.06.005)
31. Martin P. 1996 The structure of the partition algebras. *J. Algebra* **183**, 319–358. (doi:10.1006/jabr.1996.0223)