

# Linear Arboreal Categories

Samson Abramsky<sup>a,1</sup> Yoàv Montacute<sup>b,2</sup> Nihil Shah<sup>c,3</sup>

<sup>a</sup> *University College London, London, UK*

<sup>b</sup> *University of Cambridge, Cambridge, UK*

<sup>c</sup> *University of Oxford, Oxford, UK*

---

## Abstract

Arboreal categories, introduced by Abramsky and Reggio, axiomatise categories with tree-shaped objects. These categories provide a categorical language for formalising behavioural notions such as simulation, bisimulation, and resource-indexing. In this paper, we strengthen the axioms of an arboreal category to exclude ‘branching’ behaviour, obtaining a notion of ‘linear arboreal category’. We then demonstrate that every arboreal category satisfying a linearisability condition has an associated linear arboreal subcategory related via an adjunction. This identifies the relationship between the pebble-relation comonad, of Montacute and Shah, and the pebbling comonad, of Abramsky, Dawar, and Wang, and generalises it further. As another outcome of this new framework, we obtain a linear variant of the arboreal category for modal logic. By doing so we recover different linear-time equivalences between transition systems as instances of their categorical definitions. We conclude with new preservation and characterisation theorems relating trace inclusion and trace equivalence with different linear fragments of modal logic.

*Keywords:* coalgebra, categories, comonad, trace equivalence, transition system, factorisation system,

---

## 1 Introduction

In this paper, we bring several lines of research together: the categorical perspective on model comparison games and the associated logical equivalences introduced in [1,6]; the notion of arboreal categories introduced in [3], which provide an axiomatic framework for these constructions; the linear variant of one of these constructions introduced in [11] to capture pathwidth, motivated by the work of Dalmau on constraint satisfaction [7]; and the linear-time branching-time spectrum of behavioural equivalences studied by Van Glabbeek [16,17] and motivated by concurrency theory. We develop a general axiomatic framework for linear-time logics and equivalences within finite model theory, and in this setting prove new preservation and characterisation theorems.

In recent work by Montacute and Shah [11], a ‘linear’ variant of the pebbling comonad  $\mathbb{P}_k$ —the pebble-relation comonad  $\mathbb{P}\mathbb{R}_k$ —was introduced. Coalgebras over  $\mathbb{P}\mathbb{R}_k$  correspond to path decompositions of width  $< k$  and hence provide a new definition for pathwidth. Interestingly, the logic captured by  $\mathbb{P}\mathbb{R}_k$  corresponds to the restricted conjunction fragment of the  $k$ -variable logic captured by  $\mathbb{P}_k$ . Dalmau [7] showed

---

<sup>1</sup> Email: [s.abramsky@ucl.ac.uk](mailto:s.abramsky@ucl.ac.uk)

<sup>2</sup> Email: [yoav.montacute@cl.cam.ac.uk](mailto:yoav.montacute@cl.cam.ac.uk)

<sup>3</sup> Email: [nihil.shah@cs.ox.ac.uk](mailto:nihil.shah@cs.ox.ac.uk)

that definability of constraint satisfaction problems in this logic corresponds to definability in linear Datalog and restricted Krom SNP. Abramsky and Reggio [3] crystallised the intuition that the coalgebras of such *game comonads* are ‘tree-shaped’ covers of some structures. That is, they defined the notion of an *arboreal category* and showed that each game comonad discovered so far arises from an arboreal cover, i.e. an adjunction between a category of extensional objects and an arboreal category. Moreover, a key part of how game comonads capture equivalence in their associated logic is by utilising the notion of open map bisimulation, adapted from [10], which arises naturally from the arboreal structure of their category of coalgebras. In the case of modal logic, logical equivalence is ordinary bisimulation and is recovered as open map bisimulations in the category of coalgebras for the modal comonad [6]. Bisimulation is at the top of the linear-time branching-time spectrum [16,17] on transitions systems, which gives an account of the behavioural relations between trace inclusion and bisimulation.

In this paper, we bring these strands together, linking logic, behavioural relations and arboreal categories. We do this by generalising and lifting the relationship between the pebble-relation comonad  $\mathbb{P}\mathbb{R}_k$  and the pebbling comonad  $\mathbb{P}_k$  in order to obtain linear variants of comonads arising from arboreal covers. We define a new structure called a *linear arboreal category* by strengthening the definition of an arboreal category. Afterwards, we discuss the relationship between arboreal categories and linear arboreal categories. We also demonstrate sufficient conditions for when an arboreal category has an associated linear arboreal subcategory. We then provide a categorical definition for several behavioural relations including labelled trace equivalence native to a given linear arboreal category. This allows us to define a linear arboreal cover for every arboreal cover.

The linear variants of previously examined arboreal categories generalise the relationships between branching equivalences and linear equivalences (e.g. bisimulation and trace equivalence), and between tree-like parameters and their linear variants (e.g. treewidth and pathwidth). Further, we show that linear equivalences on Kripke frames and transition systems correspond to truth preservation in linear fragments of logics captured by the branching equivalences. As an application of this framework, we obtain new preservation theorems [14] and a new characterisation theorem [15,13] relating labelled trace equivalence with linear fragments of modal logic. The latter is based on the work of Otto [12] who introduced a proof combining the finite and infinite cases. This work stands in contrast to other recent approaches for capturing behavioural relations in the linear-time branching-time spectrum using category theory. In particular, there are two well-known identifiable styles for the categorification of behavioural equivalences; first in the work on graded monads (e.g. [8]), and the second in the work involving fibrations (e.g. [18]).

**Outline.** Section 2 introduces the required preliminaries for the understanding of the paper. Section 3 summarises the necessary definitions and results on arboreal categories. Section 4 discusses linear arboreal categories and their relationship with arboreal categories. Section 5 recovers the pebble-relation comonad as an linear arboreal cover and features a modified all-in-one  $k$ -pebble game capturing equivalence in existential  $k$ -variable logic with restricted conjunction. Section 6 introduces the linear variant of the modal arboreal cover and establishes the relationship between linear arboreal categories and behavioural relations. Section 7 proves new Rossman homomorphism preservation theorems for linear variants of modal logic. Section 8 proves a new Van Benthem-Rosen characterisation theorem for a linear variants of modal logic. Section 9 concludes with directions for future work.

## 2 Preliminaries

### 2.1 Set notation

Given a partially ordered set  $(X, \leq)$  and  $x \in X$ , we denote by  $\downarrow x = \{y \in X \mid y \leq x\}$  the *down-set* of  $x$ . The notion of an *up-set*  $\uparrow x$  can be defined analogously. Given a partially ordered set  $(X, \leq)$ , we define a *covering relation*  $\prec$ , such that  $x \prec y$  iff  $x < y$  and there does not exist  $z \in X$  with  $x < z < y$ . A partially ordered set is *linearly ordered* if every two elements are comparable.

If  $(T, \leq)$  is a partially ordered set such that for all  $x \in T$ ,  $\downarrow x$  is linearly ordered by  $\leq$  and finite, then we say that  $\leq$  *forest orders*  $T$  and  $(T, \leq)$  is called a *forest*. If  $(T, \leq)$  is a forest and there exists  $\perp \in T$  such that for all  $x \in T$ ,  $\perp \leq x$ , then we say that  $\leq$  *tree orders*  $T$ ,  $(T, \leq)$  is called a *tree*, and  $\perp$  is called the *root* of  $(T, \leq)$ . If  $(T, \leq)$  is a forest and every up-set  $\uparrow x$  is linearly-ordered and finite, then  $(T, \leq)$  is

called a *linear forest*. If  $(T, \leq)$  is a linear forest and a tree, then it follows that  $\leq$  is linearly-ordered and  $(T, \leq)$  is a *chain*. If  $(T, \leq)$  is a tree with root  $\perp$ , and  $(T \setminus \{\perp\}, \leq)$  is a linear forest, then  $(T, \leq)$  is called a *linear tree*.

For the sake of brevity, we write  $[n]$  to denote the set  $\{1, \dots, n\}$ . Given a sequence  $s = [x_1, \dots, x_n]$ , we write  $s[i, j]$ , where  $i < j$ , for the subsequence  $[x_i, \dots, x_j]$  and  $s(i, j)$  for the subsequence  $[x_{i+1}, \dots, x_j]$ . In addition, if  $s = [(p_1, a_1), \dots, (p_n, a_n)]$  is a sequence of pairs from a set  $[k] \times A$ , then we write  $\text{last}_p(s)$  to denote an element  $a_i$ , whenever  $p = p_i \in [k]$  is the last occurrence of  $p$  in the sequence  $s$ .

## 2.2 Category Theory

We assume familiarity with the standard category-theoretic notions of category, functor, natural transformation, and adjunction.

The categories that we are primarily interested in are categories of relational structures. A relational signature  $\sigma$  is a finite set of relational symbols  $R$ ; each with a specified positive arity. A  $\sigma$ -structure  $\mathcal{A}$  is given by a set  $A$ , the universe of the structure, and interpretations  $R^{\mathcal{A}} \subseteq A^n$  for every relational symbol  $R$  of arity  $n$ . A *homomorphism* of  $\sigma$ -structures  $h: \mathcal{A} \rightarrow \mathcal{B}$  is a set function  $h: A \rightarrow B$  such that for every relational symbol  $R \in \sigma$  of arity  $n$ ,  $R^{\mathcal{A}}(a_1, \dots, a_n)$  implies  $R^{\mathcal{B}}(h(a_1), \dots, h(a_n))$ , for all  $a_1, \dots, a_n \in A$ .

We use  $\mathbf{Struct}(\sigma)$  to denote the category of  $\sigma$ -structures and homomorphisms. For every  $\sigma$ -structure  $\mathcal{A}$ , the *Gaifman Graph* of  $\mathcal{A}$ , denoted by  $\mathcal{G}(\mathcal{A})$ , is an undirected graph with vertex set  $A$ , where elements  $a, a'$  are adjacent if  $a = a'$  or  $a, a'$  appear in the same tuple of  $R^{\mathcal{A}}$  for some  $R \in \sigma$ . We also consider the category of pointed  $\sigma$ -structures and homomorphisms  $\mathbf{Struct}_*(\sigma)$ . In  $\mathbf{Struct}_*(\sigma)$ , objects are pairs  $(\mathcal{A}, a_0)$  where  $\mathcal{A}$  is a  $\sigma$ -structure paired with a distinguished point  $a_0 \in A$ . A morphism  $h: (\mathcal{A}, a_0) \rightarrow (\mathcal{B}, b_0)$  in  $\mathbf{Struct}_*(\sigma)$  is a homomorphism  $h: \mathcal{A} \rightarrow \mathcal{B}$  of  $\sigma$ -structures such that  $h(a_0) = b_0$ . We denote by  $\mathbf{Struct}_*^f(\sigma)$  the category of *finite* pointed  $\sigma$ -structures.

**Definition 2.1** (lifting property). Given a category  $\mathcal{C}$  and morphisms  $e$  and  $m$  in  $\mathcal{C}$ , we say that  $e$  has the *left lifting property* with respect to  $m$  (or that  $m$  has the *right lifting property* with respect to  $e$ ), if for every commutative diagram as the left diagram below, there exists a *diagonal filler*  $d$  such that the right diagram below commutes.

$$\begin{array}{ccc} \bullet & \xrightarrow{e} & \bullet \\ \downarrow & & \downarrow \\ \bullet & \xrightarrow{m} & \bullet \end{array} \qquad \begin{array}{ccc} \bullet & \xrightarrow{e} & \bullet \\ \downarrow & \swarrow d & \downarrow \\ \bullet & \xrightarrow{m} & \bullet \end{array}$$

We denote this property by  $e \pitchfork m$ . For any class  $\mathcal{M}$  of morphisms in  $\mathcal{C}$ , let  ${}^{\pitchfork}\mathcal{M}$  denote the class of morphisms with the left lifting property with respect to every morphism in  $\mathcal{M}$ . The class  $\mathcal{M}^{\pitchfork}$  is defined analogously.

For any class  $\mathcal{M}$  of morphisms in  $\mathcal{C}$ , let  ${}^{\pitchfork}\mathcal{M}$  denote the class of morphisms with the left lifting property with respect to every morphism in  $\mathcal{M}$ . The class  $\mathcal{M}^{\pitchfork}$  is defined analogously.

**Definition 2.2** (weak factorisation system). Given two classes of morphisms  $\mathcal{Q}$  and  $\mathcal{M}$  in a category  $\mathcal{C}$ , the pair  $(\mathcal{Q}, \mathcal{M})$  is a *weak factorisation system* if the following conditions are satisfied:

- (i) For every morphism  $f$  in  $\mathcal{C}$ ,  $f = m \circ e$ , where  $e \in \mathcal{Q}$  and  $m \in \mathcal{M}$ ;
- (ii)  $\mathcal{Q} = {}^{\pitchfork}\mathcal{M}$  and  $\mathcal{M} = \mathcal{Q}^{\pitchfork}$ .

A weak factorisation system is *proper* if every  $e \in \mathcal{Q}$  is an epimorphism and every  $m \in \mathcal{M}$  is a monomorphism. A proper weak factorisation system is *stable* if for every  $e \in \mathcal{Q}$  and  $m \in \mathcal{M}$ , with the same codomain, there exists a pullback of  $e$  along  $m$  in  $\mathcal{Q}$ .

We refer to members of  $\mathcal{M}$  as *embeddings* (denoted by  $\mapsto$ ) and to members of  $\mathcal{Q}$  as *quotients* (denoted by  $\twoheadrightarrow$ ). Given two *embeddings*  $m: S \mapsto X$  and  $n: T \mapsto X$ , we write  $m \trianglelefteq n$  to denote that there is a morphism  $i: S \rightarrow T$  such that  $m = n \circ i$ . Note that  $\trianglelefteq$  induces a preorder on embeddings with the same codomain. The symmetrisation of  $\trianglelefteq$  induces an equivalence relation  $\sim$ . The relation  $\sim$  can be

characterised as  $m \sim n$  if there exists an isomorphism  $i : S \rightarrow T$  such that  $m = n \circ i$ . Let  $\mathbf{SX}$  denote the class of  $\sim$ -equivalence classes of embeddings with codomain  $X$  and partial order  $\leq$  induced by  $\trianglelefteq$ . The  $\sim$ -equivalence class with a representative  $m : S \rightarrow X$  is denoted  $[m]$ .

### 2.3 Kripke frames and transition systems

Consider the category  $\mathbf{Struct}_*(\sigma)$ , where the signature  $\sigma$  has only relation symbols of arity  $\leq 2$ . In this case, we will index all binary relations  $\{R_\alpha \mid \alpha \in \mathbf{Act}\}$  by an action alphabet  $\mathbf{Act}$  and all the unary relations  $\{P_l \mid l \in \mathbf{PV}\}$  by a set of propositional variables  $\mathbf{PV}$ . Objects in  $\mathbf{Struct}_*(\sigma)$  are then exactly pointed Kripke models. In a typical presentation, a *pointed Kripke model*  $\mathcal{A} = (A, a_0, \{R_\alpha\}, V)$  is specified by set of states  $A$  with distinguished initial state  $a_0 \in A$ , finitely many binary transition relations  $R_\alpha$ , and valuation map  $V : A \rightarrow \wp(\mathbf{PV})$  from states to sets of propositional variables. A *classical pointed Kripke model* is a pointed Kripke model with one binary relation. We recover this encoding of a pointed Kripke model from an object  $(\mathcal{A}, a_0) = (A, \{R_\alpha \mid \alpha \in \mathbf{Act}\}, \{P_l \mid l \in \mathbf{PV}\}, a_0)$  in  $\mathbf{Struct}_*(\sigma)$  by defining the valuation map  $V : A \rightarrow \wp(\mathbf{PV})$  as  $V(a) = \{l \in \mathbf{PV} \mid a \in P_l^A\}$ .

The objects in  $\mathbf{Struct}_*(\sigma)$  are general enough to define two slightly different notions of trace inclusion (resp. equivalence) on transitions systems. For every object  $(\mathcal{A}, a_0) \in \mathbf{Struct}_*(\sigma)$ , we define the set of *labelled traces*

$$\mathbf{ltraces}(a_0) = \{V(a_0)\alpha_1 V(a_1)\alpha_2 \dots \alpha_n V(a_n) \mid n \in \mathbb{N}, \forall i \in [n](R_{\alpha_i}(a_{i-1}, a_i))\}.$$

We will occasionally write  $a_0 \xrightarrow{\alpha_1} a_1 \dots \xrightarrow{\alpha_n} a_n$  instead of  $\forall i \in [n](R_{\alpha_i}(a_{i-1}, a_i))$ .

**Definition 2.3.** Given two objects  $(\mathcal{A}, a_0)$  and  $(\mathcal{B}, b_0)$  in  $\mathbf{Struct}_*(\sigma)$ , we define

- $a_0 \subseteq^{\text{tr}} b_0$  if for all labelled traces  $V(a_0)\alpha_1 V(a_1)\alpha_2 \dots \alpha_n V(a_n) \in \mathbf{ltraces}(a_0)$ , there exists a labelled trace  $V(b_0)\alpha_1 V(b_1)\alpha_2 \dots \alpha_n V(b_n) \in \mathbf{ltraces}(b_0)$  such that  $V(a_i) \subseteq V(b_i)$  for all  $i \in [n]$ .
- $a_0 \subseteq^{\text{ltr}} b_0$  if  $\mathbf{ltraces}(a_0) \subseteq \mathbf{ltraces}(b_0)$ . This equivalent is to  $a_0 \subseteq^{\text{tr}} b_0$  but with the stronger condition that  $V(a_i) = V(b_i)$ .

We say that there exists a *trace inclusion* from  $a_0$  to  $b_0$  if  $a_0 \subseteq^{\text{tr}} b_0$ . The object  $a_0$  is then said to be *trace included* in  $b_0$ . We write  $a_0 \sim^{\text{tr}} b_0$  to denote that there exists a *trace equivalence* between  $a_0$  and  $b_0$ , and say that  $a_0$  is *trace equivalent* to  $b_0$ , i.e. whenever both  $a_0 \subseteq^{\text{tr}} b_0$  and  $b_0 \subseteq^{\text{tr}} a_0$ . The relation  $\sim^{\text{ltr}}$  is defined analogously. Each of these relations can be graded by a resource parameter  $k > 0$ , e.g.  $\subseteq_k^{\text{tr}}$  for a grading of  $\subseteq^{\text{tr}}$ , where the definitions are restricted to traces of length  $\leq k$ .

A *transition system* is a tuple  $\mathcal{A} = (A, \{R_\alpha \mid \alpha \in \mathbf{Act}\}, a_0)$ , where  $A$  is considered as the set of states,  $\mathbf{Act}$  is the set of *actions* and  $a_0 \in A$  is an initial state. Note that a transition system is a special case of a pointed Kripke model where the map  $V$  is unspecified. This is also referred to as a *Kripke frame* in the literature. For transition systems, the relations  $\subseteq^{\text{tr}}$  and  $\subseteq^{\text{ltr}}$  are equivalent. Two structures  $(\mathcal{A}, a_0)$  and  $(\mathcal{B}, b_0)$  are (labelled) trace equivalent if they admit a (labelled) trace equivalence.

**Example 2.4.** The two transition systems  $(\mathcal{A}, a_0)$  and  $(\mathcal{B}, b_0)$  in Figure 1 are trace equivalent but not labelled trace equivalent. To see why, observe that there is a labelled trace  $\emptyset\alpha\emptyset\alpha\{p\}$  in  $\mathbf{ltraces}(b_0)$ , but not in  $\mathbf{ltraces}(a_0)$ .

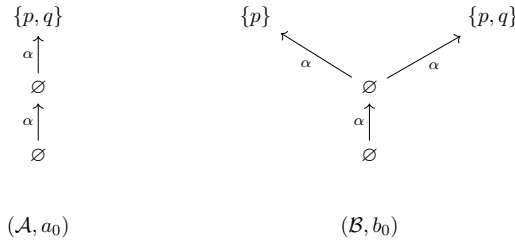


Fig. 1. Trace equivalent transition systems that are not labelled trace equivalent

### 3 Arboreal categories

In this section we cover some of the required preliminaries on arboreal categories from the papers by Abramsky and Reggio [3,5] introducing the subject.

Intuitively, arboreal categories  $\mathcal{C}$  are defined to formalise the notion of a category whose objects are ‘tree-shaped’. Similar to how objects in a locally finitely presentable category are generated from a subcategory  $\mathcal{C}_{fp}$  of finitely presentable objects, the tree-shaped objects of an arboreal category are generated from a subcategory  $\mathcal{C}_p$  of path objects. For this reason, we first define a notion of *path category*  $\mathcal{C}$  (Definition 3.2) that has enough structure to define a full subcategory  $\mathcal{C}_p$  of path objects.

We assume throughout, that we are dealing with categories  $\mathcal{C}$  equipped with a stable proper weak factorisation system.

**Definition 3.1** (path). An object  $P$  of  $\mathcal{C}$  is called a *path* if  $\mathbf{S}P$  is a finite chain.

A *path embedding* is an embedding with a path as its domain. Given an object  $X$  in  $\mathcal{C}$ , let  $\mathbf{P}X$  denote the sub-poset of  $\mathbf{S}X$  of path embeddings. A morphism  $f : X \rightarrow Y$  in  $\mathcal{C}$  is a *pathwise embedding* if  $f \circ m$  is a path embedding, for all path embeddings  $m : P \rightarrow X$ . Pathwise embeddings induce a mapping  $\mathbf{P}f : \mathbf{P}X \rightarrow \mathbf{P}Y$ , where  $\mathbf{P}f(m) = f \circ m$  for all  $m \in \mathbf{P}X$ . A morphism  $f : X \rightarrow Y$  in  $\mathcal{C}$  is *open* if, given a commutative square

$$\begin{array}{ccc} P & \rightarrow & Q \\ \downarrow & & \downarrow \\ X & \xrightarrow{f} & Y \end{array}, \quad (1)$$

where  $P$  and  $Q$  are paths, there exists a diagonal filler embedding  $Q \rightarrow X$ . A span of type  $X \xleftarrow{f} R \xrightarrow{g} Y$  such that  $f$  and  $g$  are open pathwise embeddings is called an *arboreal bisimulation from  $X$  to  $Y$* . Intuitively, an open map  $f : X \rightarrow Y$  states that if a path  $P \rightarrow X$  in the domain  $X$  can be extended to a path  $Q \rightarrow Y$  in the codomain  $Y$ , then  $P$  can be extended to  $Q$  in the domain  $X$  as witnessed by the diagonal  $Q \rightarrow X$ . This definition of open map differs slightly from the original notion [10] as we require all but the arrow  $f : X \rightarrow Y$  in diagram (1) to be embeddings. For objects in an arboreal category, we say that  $X$  and  $Y$  are *bisimilar* if there exists an arboreal bisimulation from  $X$  to  $Y$ .

Given a category  $\mathcal{C}$ , there is a full subcategory  $\mathcal{C}_p$  of  $\mathcal{C}$  containing only the path objects which satisfy Definition 3.1. Suppose  $\mathcal{C}$  is a category which has all coproducts of small families of path objects. Then we say that an object  $X$  of  $\mathcal{C}$  is *connected* if, for all non-empty small families of paths  $\{P_i\}$ , a morphism  $X \rightarrow \coprod_{i \in I} P_i$  factors through some coproduct injection  $P_j \rightarrow \coprod_{i \in I} P_i$ . In the case this coproduct injection is unique, we say that  $X$  of  $\mathcal{C}$  is *strongly connected*.

**Definition 3.2** (path category [5]). A category  $\mathcal{C}$  equipped with a stable weak proper factorisation system is a *path category* if it satisfies the following conditions:

- (i)  $\mathcal{C}$  has all coproducts  $\coprod_{j \in J} X_j$ , where  $\{X_j\}$  is a small family of paths.
- (ii) For all paths  $P, Q$  and  $R$ , if  $P \rightarrow Q \rightarrow R$  is a quotient, then  $P \rightarrow Q$  is a quotient.
- (iii) Every path in  $\mathcal{C}$  is connected.

The full subcategory  $\mathcal{C}_p$  of a path category  $\mathcal{C}$  is central to many of the semantic constructions which generalise notions of property-preserving bisimulation, simulation, and trace equivalence that originate in the theory of transition systems, Kripke structures, and automata.

An object  $X$  of a path category  $\mathcal{C}$  is *path-generated* if  $X$  is the colimit of the cocone consisting of all commuting triangles of the form

$$\begin{array}{ccc} P & \xrightarrow{\quad} & Q \\ & \searrow & \swarrow \\ & X & \end{array} .$$

**Definition 3.3** (arboreal category). An *arboreal category* is a path category in which all objects are

path-generated.

**Example 3.4.** The arboreal category  $\mathcal{R}(\sigma)$  [3, Example 3.4] consists of objects  $(\mathcal{A}, \leq)$ , where  $\mathcal{A}$  is a  $\sigma$ -structure and  $\leq$  is a forest order on  $A$ . Morphisms  $f: (\mathcal{A}, \leq) \rightarrow (\mathcal{B}, \leq')$  in  $\mathcal{R}(\sigma)$  are  $\sigma$ -homomorphisms that preserve roots, i.e. minimal elements of  $\leq$ , and the induced covering relation  $\prec$  of the forest order  $\leq$ . As discussed in [3], the categories of coalgebras for the Ehrenfeucht-Fraïssé, pebbling, and modal comonads are isomorphic to the arboreal categories  $\mathcal{R}^E(\sigma)$ ,  $\mathcal{R}^P(\sigma)$  and  $\mathcal{R}^M(\sigma)$  based on  $\mathcal{R}(\sigma)$ . They are described as follows:

- For the Ehrenfeucht-Fraïssé comonad, the category  $\mathcal{R}^E(\sigma)$  is the full subcategory of  $\mathcal{R}(\sigma)$  of objects  $(\mathcal{A}, \leq)$  which satisfy the condition  
(E) If  $a, a' \in \mathcal{A}$  are adjacent in the Gaifman graph of  $\mathcal{A}$ , then  $a \leq a'$  or  $a' \leq a$ .  
The category  $\mathcal{R}_k^E(\sigma)$  is the subcategory of  $\mathcal{R}^E(\sigma)$  with forest order of height  $\leq k$ . Unpacking the definition, we have that  $(\mathcal{A}, \leq) \in \mathcal{R}_k^E(\sigma)$  is a  $k$ -height forest cover of the Gaifman graph of  $\mathcal{A}$ .
- For the pebbling comonad, the category  $\mathcal{R}_k^P(\sigma)$  has objects  $(\mathcal{A}, \leq, p: A \rightarrow [k])$ , where  $(\mathcal{A}, \leq) \in \mathcal{R}^E(\sigma)$ , and the pebbling function  $p: A \rightarrow [k]$  satisfies the condition  
(P) If  $a, a' \in \mathcal{A}$  are adjacent in the Gaifman graph of  $\mathcal{A}$  and  $a \leq a'$ , then for all  $b \in (a, a']$ ,  $p(b) \neq p(a)$ .  
The morphisms of  $\mathcal{R}_k^P(\sigma)$  are morphisms of  $\mathcal{R}^E(\sigma)$  that also preserve the pebbling function. The object  $(\mathcal{A}, \leq, p) \in \mathcal{R}_k^P(\sigma)$  is a  $k$ -pebble forest cover of  $\mathcal{A}$ .
- For the modal comonad, the category  $\mathcal{R}^M(\sigma)$  is the category of tree-ordered  $\sigma$ -structures  $(\mathcal{A}, a_0, \leq)$ , where  $a_0$  is the root of  $(\mathcal{A}, \leq)$  and  $\sigma$  contains only relations of arity  $\leq 2$  satisfying the condition  
(M)  $a \prec a'$  if and only if there exists a unique binary relation  $R_\alpha \in \sigma$  such that  $R_\alpha^A(a, a')$ .  
The category  $\mathcal{R}_k^M(\sigma)$  is the subcategory of  $\mathcal{R}^M(\sigma)$  with tree order of height  $\leq k$ .

Arboreal categories have a process structure that allows for ‘dynamic’ notions such as property-preserving simulations, bisimulations, and back-and-forth systems. In particular, arboreal bisimulation is equivalent to a general back-and-forth game [6]. Utilising an adjunction, we can relate the process structure of an arboreal category  $\mathcal{C}$  to an ‘extensional’ category  $\mathcal{E}$  of ‘static’ objects.

**Definition 3.5** (arboreal cover). An *arboreal cover* of  $\mathcal{E}$  by  $\mathcal{C}$  is a comonadic adjunction  $(L, R, \varepsilon, \eta)$  such that

$$\mathcal{C} \begin{array}{c} \xrightarrow{L} \\ \perp \\ \xleftarrow{R} \end{array} \mathcal{E} .$$

Every adjunction yields a comonad  $(\mathbb{C}, \varepsilon, \delta)$  over  $\mathcal{E}$ , where  $\mathbb{C} = LR$ , and the component  $\delta_a: LRa \rightarrow LRLRa$  for  $a \in \mathcal{E}$  is defined as  $L(\eta_{Ra})$ . The comonadicity condition states that the arboreal category  $\mathcal{C}$  is isomorphic to the Eilenberg-Moore category of coalgebras for the comonad  $\mathbb{C}$ . Intuitively, this means that we can view the tree-shaped objects of  $\mathcal{C}$  as covers, or unravelings, of the objects in  $\mathcal{E}$ . Moreover, this adjunction allows us to study the objects of  $\mathcal{E}$  using the process structure of  $\mathcal{C}$ .

One of the purposes of using the arboreal setting is to analyse resources associated with the process structure of  $\mathcal{C}$ . To formalise this resource structure, an arboreal category  $\mathcal{C}$  may be graded by a resource parameter  $k > 0$ .

**Definition 3.6** (resource indexing). Let  $\mathcal{C}$  be an arboreal category with full subcategory of paths  $\mathcal{C}_p$ . The arboreal category  $\mathcal{C}$  is *resource-indexed* by a parameter  $k$  if for all  $k > 0$ , there is full subcategory  $\mathcal{C}_p^k$  of  $\mathcal{C}_p$  closed under embeddings, i.e. if  $Q \in \mathcal{C}_p^k$  and  $P \hookrightarrow Q \in \mathcal{C}_p$ , then  $P \in \mathcal{C}_p^k$ , with inclusions

$$\mathcal{C}_p^1 \hookrightarrow \mathcal{C}_p^2 \hookrightarrow \mathcal{C}_p^3 \hookrightarrow \dots$$

This induces a corresponding tower of full subcategories  $\mathcal{C}_k$  of  $\mathcal{C}$  with the objects of  $\mathcal{C}_k$  being those whose cocone of path embeddings with domain in  $\mathcal{C}_p^k$  is a colimit cocone in  $\mathcal{C}$ .

To be explicit that this resource-indexing induces a family of categories, we will use the notation  $\{\mathcal{C}_k\}$  to denote an arboreal category which is resource-indexed by a parameter  $k > 0$ . Note that the  $k$ -height



resource indexing of  $\{\mathcal{R}_k^E(\sigma)\}$  and  $\{\mathcal{R}_k^M(\sigma)\}$ , and the  $k$ -pebble indexing of  $\{\mathcal{R}_k^P(\sigma)\}$  in Example 3.4, are all instances of resource-indexed arboreal categories.

By Proposition 7.6 of [3], if  $\{\mathcal{C}_k\}$  is a resource-indexed arboreal category, then each subcategory  $\mathcal{C}_k$  is itself an arboreal category. Consequently, we can resource-index arboreal covers of extensional categories. Given a resource-indexed arboreal category  $\{\mathcal{C}_k\}$ , a *resource-indexed arboreal cover* of  $\mathcal{E}$  by  $\{\mathcal{C}_k\}$  is an indexed family of comonadic adjunctions

$$\mathcal{C}_k \begin{array}{c} \xrightarrow{L_k} \\ \perp \\ \xleftarrow{R_k} \end{array} \mathcal{E} ,$$

yielding corresponding comonads  $\mathbb{C}_k = L_k R_k$  on  $\mathcal{E}$ .

**Definition 3.7.** Consider a resource-indexed arboreal cover of  $\mathcal{E}$  by  $\mathcal{C}$ , and two objects  $a, b$  of  $\mathcal{E}$ . For all  $k > 0$ , we define:

- $a \rightarrow_k^{\mathcal{C}} b$  if there is a morphism  $R_k(a) \rightarrow R_k(b)$  in  $\mathcal{C}_k$ .
- $a \dashrightarrow_k^{\mathcal{C}} b$  if there is a pathwise embedding  $R_k(a) \rightarrow R_k(b)$  in  $\mathcal{C}_k$ .
- $a \leftrightarrow_k^{\mathcal{C}} b$  if there is an arboreal bisimulation between  $R_k(a)$  and  $R_k(b)$  in  $\mathcal{C}_k$ .
- $a \cong_k^{\mathcal{C}} b$  if there is an isomorphism  $R_k(a) \cong R_k(b)$  in  $\mathcal{C}_k$ .

## 4 Linear arboreal categories

In an arboreal category, since every object  $X$  is path-generated,  $X$  is a colimit cocone of its branches. For the notion of linear arboreal category, we would like to impose additional conditions to exclude ‘non-trivial branching’. Branching appears in two forms for an arboreal category  $\mathcal{C}$ . The first form is evident in the objects of  $\mathcal{C}$ . Namely, objects in  $\mathcal{C}$  are colimits of their branch subobjects. Thus to exclude branching in the objects of  $\mathcal{C}$ , we additionally require that every object  $X \in \mathcal{C}$  is a coproduct of paths. The second form is evident in the morphisms of  $\mathcal{C}$ . In particular, path embeddings  $p: P \rightarrow X$  in  $\mathcal{C}$  isolate a partial branch of  $X$  that can be extended in possibly multiple ways, via extensions  $j: P \rightarrow Q$ , to longer partial branches  $q: Q \rightarrow X$  where  $p = q \circ j$ . Thus to exclude branching in the morphisms of  $\mathcal{C}$ , we need a path embedding  $P \rightarrow X$  to isolate a ‘full path’ or ‘trace’, rather than partial path. To accomplish this, we add an axiom to ensure that each path in  $\mathcal{C}$  has only trivial extensions. We formalise these exclusions in the following axiomatic definition.

**Definition 4.1** (linear arboreal category). A *linear arboreal category* is an arboreal category  $\mathcal{C}$  such that the following two axioms are satisfied:

- (L1) Every object  $X \in \mathcal{C}$  is a coproduct  $\bigsqcup_{P_i \in \mathcal{C}_p} P_i$  of path objects  $P_i$ .
- (L2) For every two non-initial paths  $P$  and  $Q$ , if  $j: P \rightarrow Q$  is an embedding, then  $j: P \cong Q$  is an isomorphism.

We say that an arboreal category  $\mathcal{C}$  is a *quasi-linear arboreal category* if (L1) is satisfied.

One counterintuitive consequence of condition (L2) is that the path objects in a linear arboreal category appear ‘externally’ rather than internally. This is captured in the following proposition:

**Proposition 4.2.** *If  $\mathcal{C}$  is a linear arboreal category and  $P \in \mathcal{C}$  is a path, then  $\mathbf{ht}(P) \leq 1$ .*

Proposition 4.2 hints at the idea that ‘paths’ in the linear arboreal setting can be seen as types for full behaviours, i.e. traces and, in terms of model comparison games, produce ‘all-in-one’ variants.

Another counter-intuitive consequence of these axioms is that bisimulations in  $\mathcal{C}$  trivialise to ‘bidirectional’ pathwise embeddings.

**Corollary 4.3.** *Let  $\mathcal{C}$  be a linear arboreal category, and suppose the categorical product  $X \times Y$  exists in  $\mathcal{C}$ . Then the following are equivalent:*

- (1) *There exists a bisimulation  $X \xleftarrow{f} Z \xrightarrow{g} Y$  in  $\mathcal{C}$ .*

(2) There exist pathwise embeddings  $h: X \rightarrow Y$  and  $h': Y \rightarrow X$  in  $\mathcal{C}$ .

#### 4.1 Linear adjunction

In this section, we construct a linear arboreal subcategory  $\mathcal{C}^L$  from any arboreal category  $\mathcal{C}$  which satisfies the linearisability condition. The linear arboreal subcategory  $\mathcal{C}^L$  is related to  $\mathcal{C}$  via a right adjoint to the inclusion functor  $I: \mathcal{C}^L \rightarrow \mathcal{C}$ . Using this adjunction, we can derive linear variants of many of the corresponding constructions in  $\mathcal{C}$ .

To define the linearisability condition, we simply strengthen to notion of connectedness in Axiom (iii) of Definition 3.2 to the usual categorical notion.

**Definition 4.4.** An object  $X \in \mathcal{C}$  is *strongly connected* if every morphism  $X \rightarrow \bigsqcup_{i \in I} P_i$ , where  $\{P_i\}$  is a family of paths, factors through a unique coproduct injection  $P_j \rightarrow \bigsqcup_{i \in I} P_i$ .

The following condition, which we can define for any path category, is sufficient to guarantee that an arboreal category  $\mathcal{C}$  contains a linear arboreal subcategory  $\mathcal{C}^L$ .

**Definition 4.5.** A path category is *linearisable* if every non-initial path  $P \in \mathcal{C}$  is strongly connected.

The linearisability condition is independent from the axioms of a path category.

For every linearisable arboreal category  $\mathcal{C}$ , we construct a subcategory  $\mathcal{C}^L$  which is a linear arboreal category. To construct  $\mathcal{C}^L$ , we first construct the quasi-linear arboreal category  $\mathcal{C}^{qL}$  from  $\mathcal{C}$  by restricting the objects of  $\mathcal{C}$  to those that are generated by their maximal paths.

**Definition 4.6.** A path embedding  $m: P \rightarrow X$  is *maximal* if  $P$  is not initial and for all  $n: P' \rightarrow X$  such that  $m \leq n$ ,  $m \sim n$ .

Equivalently,  $m: P \rightarrow X$  is maximal if the  $\sim$ -equivalence class  $[m]$  is a maximal element in the poset  $\mathbf{P}(X)$ . For every object  $X$  in a path category, let  $\mathbf{P}^\top(X)$  be the subset of maximal elements in  $\mathbf{P}(X)$ . With the definition of maximality in place, we obtain  $\mathcal{C}^{qL}$  as the full subcategory whose objects are generated by their maximal paths. Formally, such objects are defined as follows:

**Definition 4.7.** An object  $X \in \mathcal{C}$  of a path category  $\mathcal{C}$  is *linearly path-generated* if it is the coproduct of its maximal elements, i.e.

$$X \cong \bigsqcup_{[m] \in \mathbf{P}^\top(X)} \text{dom}(m).$$

The next step in the construction of  $\mathcal{C}^L$  is to obtain  $\mathcal{C}^L$  from  $\mathcal{C}^{qL}$  by restricting to those morphisms of  $\mathcal{C}^{qL}$  that preserve maximal elements. Formally, such morphisms are defined as follows:

**Definition 4.8.** A morphism  $f: X \rightarrow Y \in \mathcal{C}$  is a *leaf morphism* in  $\mathcal{C}$  if for every  $[m] \in \mathbf{P}^\top(X)$ ,

$$\mathbf{P}(f)([m]) = [\exists_f m] \in \mathbf{P}^\top(Y).$$

Thus,  $\mathcal{C}^L$  is the subcategory of  $\mathcal{C}$  where the objects are the linearly-path generated objects of  $\mathcal{C}$  and the morphisms are the leaf morphisms of  $\mathcal{C}$ .

**Proposition 4.9.** *If  $\mathcal{C}$  is a linearisable arboreal category, then  $\mathcal{C}^L$  is a linear arboreal category.*

For the linear arboreal subcategory  $\mathcal{C}^L$  of an arboreal category  $\mathcal{C}$ , there is an inclusion  $I: \mathcal{C}^L \hookrightarrow \mathcal{C}$ . Paths of an object in  $X \in \mathcal{C}^L$  are essentially the maximal paths of  $I(X) \in \mathcal{C}$ . The inclusion  $I: \mathcal{C}^L \rightarrow \mathcal{C}$  has a right adjoint  $T: \mathcal{C} \rightarrow \mathcal{C}^L$ . The right adjoint  $T$  will be central for computing from a game comonad its linear variant. To construct the object mapping of  $T$ , suppose  $X$  is an object of arboreal category  $\mathcal{C}$ . We define  $T(X)$  as the coproduct of all paths of  $X$ . In notation,

$$T(X) = \bigsqcup_{[p] \in \mathbf{P}(X)} \text{dom}(p).$$



For every  $X \in \mathcal{C}$ , by the universal property of the coproduct  $I(T(X))$ , there exists a unique morphism  $\varepsilon_X^L: I(T(X)) \rightarrow X$  such that for all  $[p] \in \mathbf{P}(X)$ ,  $p = \varepsilon_X^L \circ i_{[p]}$ , where  $i_{[p]}: \text{dom}(p) \rightarrow I(T(X))$  is the coproduct injection. The following theorem demonstrates that for every object  $X \in \mathcal{C}$ ,  $T(X)$  and  $\varepsilon_X^L$  is a universal morphism from  $I$  to  $X$ . Thus,  $T$  extends to a functor  $T: \mathcal{C} \rightarrow \mathcal{C}^L$  and is right adjoint to  $I$ .

**Theorem 4.10.** *For every linearisable arboreal category  $\mathcal{C}$ ,  $I \dashv T$  is an adjunction from  $\mathcal{C}^L$  to  $\mathcal{C}$ .*

**Example 4.11.** The linear arboreal subcategory  $\mathcal{R}^L(\sigma)$  of  $\mathcal{R}(\sigma)$ , described in Example 3.4, consists of objects that are forest-ordered  $\sigma$ -structures  $(\mathcal{A}, \leq)$ , where  $(\mathcal{A}, \leq)$  is a linear forest, and of forest-ordered  $\sigma$ -morphisms that preserve maximal elements. Using  $\mathcal{R}^L(\sigma)$ , we can define the linear subcategories  $\mathcal{R}^{EL}(\sigma)$ ,  $\mathcal{R}^{PL}(\sigma)$ , and  $\mathcal{R}^{ML}(\sigma)$  of  $\mathcal{R}^E(\sigma)$ ,  $\mathcal{R}^P(\sigma)$ , and  $\mathcal{R}^M(\sigma)$ , respectively.

- For the Ehrenfeucht-Fraïssé comonad, the linear arboreal subcategory  $\mathcal{R}^{EL}(\sigma)$  of  $\mathcal{R}^E(\sigma)$  has objects  $(\mathcal{A}, \leq)$ , where  $\leq$  linear orders each connected component of the  $\sigma$ -structure  $\mathcal{A}$  and morphisms which, as in  $\mathcal{R}^L(\sigma)$ , preserve roots, the covering relation, and maximal elements. Intuitively, these morphisms merge connected components of equal height in the order  $\leq$ .
- For the pebbling comonad, the linear arboreal subcategory  $\mathcal{R}_k^{PL}(\sigma)$  of  $\mathcal{R}_k^P(\sigma)$  is isomorphic to the category of coalgebras of the pebble-relation comonad [11, Theorem 4.9]. The category  $\mathcal{R}_k^{PL}(\sigma)$  is the subcategory of  $\mathcal{R}_k^P(\sigma)$  consisting of objects  $(\mathcal{A}, \leq, p)$  such that  $(\mathcal{A}, \leq) \in \mathcal{R}^L(\sigma)$ , and morphisms  $f: (\mathcal{A}, \leq, p) \rightarrow (\mathcal{B}, \leq', p')$  such that  $f: (\mathcal{A}, \leq) \rightarrow (\mathcal{B}, \leq') \in \mathcal{R}^L(\sigma)$  and  $p'(f(a)) = p(a)$ , for all  $a \in A$ . We explore this category in Section 5.
- For the modal comonad, the linear arboreal subcategory  $\mathcal{R}^{ML}(\sigma)$  of  $\mathcal{R}^M(\sigma)$  is a category of linear synchronisation trees. The category  $\mathcal{R}^{ML}(\sigma)$  is the subcategory of  $\mathcal{R}^M(\sigma)$  consisting of objects  $(\mathcal{A}, a_0, \leq)$  such that  $(\mathcal{A}, \leq)$  is a linear tree with root  $a_0$ . We discuss this category in Section 6.

## 4.2 Linear behavioural relations

For every arboreal cover of  $\mathcal{E}$  by  $\mathcal{C}$ , we can obtain a linear arboreal cover of  $\mathcal{E}$  by  $\mathcal{C}^L$  by observing that the adjunction  $L \circ I \dashv T \circ R$  is comonadic.

**Proposition 4.12.** *If  $L \dashv R$  is an arboreal cover of  $\mathcal{E}$  by  $\mathcal{C}$ , then  $L \circ I \dashv T \circ R$  is a linear arboreal cover of  $\mathcal{E}$  by  $\mathcal{C}^L$ .*

Moreover, if an arboreal cover of  $\mathcal{E}$  by  $\mathcal{C}$  is resource-indexed by a parameter  $k > 0$  as in Definition 3.6, then this induces a resource indexing by the same parameter  $k > 0$  on the linear arboreal cover of  $\mathcal{E}$  by  $\mathcal{C}^L$ . If an arboreal category  $\mathcal{C}$  is indexed by a parameter  $k$ , then  $\mathcal{C}$  is equipped, for all  $k > 0$ , with a full subcategory  $\mathcal{C}_p^k$  of  $\mathcal{C}_p$  closed under embeddings with inclusions

$$\mathcal{C}_p^1 \hookrightarrow \mathcal{C}_p^2 \hookrightarrow \mathcal{C}_p^3 \hookrightarrow \dots$$

This induces a corresponding resource indexing  $\mathcal{C}_k^L$  of  $\mathcal{C}^L$  with objects of  $\mathcal{C}^L$  being those that are coproducts of objects in  $\mathcal{C}_p^k$ . Equivalently,  $\mathcal{C}_k^L$  is the full subcategory of objects in  $\mathcal{C}_k$  that are also in  $\mathcal{C}^L$ . As with the case of arboreal categories, each of the subcategories  $\mathcal{C}_k^L$  is a linear arboreal category. Consequently, the adjunctions  $I_k \dashv T_k$ , where  $I_k: \mathcal{C}_k^L \hookrightarrow \mathcal{C}_k$  and  $T_k: \mathcal{C}_k \rightarrow \mathcal{C}_k^L$  witness a linear arboreal cover of  $\mathcal{C}$  by  $\mathcal{C}^L$ . These adjunctions can be composed with the arboreal covers  $L_k \dashv R_k$  of  $\mathcal{E}$  by  $\{\mathcal{C}_k\}$ , obtaining a resource-indexed linear arboreal cover  $L_k \circ I_k \dashv T_k \circ R_k$  of  $\mathcal{E}$  by  $\{\mathcal{C}_k^L\}$ .

From a resource-indexed arboreal cover  $L_k \dashv R_k$ , we obtain relations on objects in  $\mathcal{E}$  which are categorical versions of ‘branching’ behavioural relations like simulation  $(\rightarrow_k^{\mathcal{C}})$ , property-preserving simulation  $(\rightarrow_k^{\mathcal{C}})$ , bisimulation  $(\leftrightarrow_k^{\mathcal{C}})$ , and graded bisimulation  $(\cong_k^{\mathcal{C}})$ , defined in terms of constructions in  $\mathcal{C}$  (Definition 3.7). On the other hand, Definition 3.7 applied to the resource-indexed linear arboreal cover  $L_k \circ I_k \dashv T_k \circ R_k$  provides a language for applying a categorical version of ‘linear’ behavioural relations like trace inclusion  $(\rightarrow_k^{\mathcal{C}^L})$ , labelled trace equivalence  $(\leftrightarrow_k^{\mathcal{C}^L})$ , and bijective labelled trace equivalence  $(\cong_k^{\mathcal{C}^L})$  (Definition 6.5) to objects in  $\mathcal{E}$  via the same constructions in the subcategory  $\mathcal{C}^L$  of  $\mathcal{C}$ . This intuition is crystallised in Section

6 by investigating these relations on the linear variant of  $\mathcal{R}_k^M(\sigma)$ . In Kripke structures and transition systems, it is well known that the ‘branching’ relations imply ‘linear’ relations. These results generalise to the setting of arboreal categories and is stated in Theorem 4.15. Given that these relations are defined in terms of pathwise embeddings and open maps, we first show that  $T$  preserves both pathwise embeddings and open maps.

**Proposition 4.13.** *If  $g: X \rightarrow Y$  is a pathwise embedding, then  $T(g): T(X) \rightarrow T(Y)$  is a pathwise embedding.*

**Proposition 4.14.** *If  $g: X \rightarrow Y$  is an open map, then  $T(g): T(X) \rightarrow T(Y)$  is an open map.*

**Theorem 4.15.** *Given a resource-indexed arboreal cover of  $\mathcal{E}$  by  $\{\mathcal{C}_k\}$  and two objects  $a, b$  of  $\mathcal{E}$ , for all  $k > 0$ ,*

- (i)  $a \rightarrow_k^{\mathcal{C}} b$  implies  $a \rightarrow_k^{\mathcal{C}^L} b$ .
- (ii)  $a \dashrightarrow_k^{\mathcal{C}} b$  implies  $a \dashrightarrow_k^{\mathcal{C}^L} b$ .
- (iii)  $a \leftrightarrow_k^{\mathcal{C}} b$  implies  $a \leftrightarrow_k^{\mathcal{C}^L} b$ .
- (iv)  $a \cong_k^{\mathcal{C}} b$  implies  $a \cong_k^{\mathcal{C}^L} b$ .

Viewing  $I_k \dashv T_k$  as a linear arboreal cover of  $\mathcal{C}_k$  by  $\mathcal{C}_k^L$ , the proofs of the statements in Theorem 4.15 can be reformulated as generalisations of stronger statements about transition systems. For instance, the proof of statement (iii) of Theorem 4.15 demonstrates that two pointed transition systems  $(\mathcal{A}, a), (\mathcal{B}, b)$  are labelled trace equivalent for traces of length  $\leq k$  if and only if  $T_k(R_k(\mathcal{A}, a))$  and  $T_k(R_k(\mathcal{B}, b))$  are bisimilar up to depth  $\leq k$ .

## 5 The pebbling linear arboreal cover

Recall the definition of  $\mathcal{R}^{PL}(\sigma)$  from Example 4.11. The following proposition follows from Proposition 4.9 and the proof that  $\mathcal{R}_k^P(\sigma)$  is a linearisable arboreal category [6,5].

**Proposition 5.1.**  *$\mathcal{R}_k^{PL}(\sigma)$  is a linear arboreal category.*

There is a forgetful functor  $L_k: \mathcal{R}_k^{PL}(\sigma) \rightarrow \mathbf{Struct}(\sigma)$  mapping  $(\mathcal{A}, \leq, p)$  to  $\mathcal{A}$ . For each  $\sigma$ -structure  $\mathcal{A}$ , let  $LR(\mathcal{A})$  be the structure with universe

$$LR(\mathcal{A}) = \{(s, i) \mid s = [(p_1, a_1), \dots, (p_n, a_n)] \text{ and } i \in [n], \forall n \in \mathbb{N}\}.$$

We define a counit map  $\varepsilon: LR(\mathcal{A}) \rightarrow \mathcal{A}$  as  $(s, i) \mapsto a_i$ . Let  $s(i, j)$  be the subsequence of  $s$  from index  $i + 1$  to  $j$ . Suppose  $R \in \sigma$  is an  $m$ -ary relation. Then  $R^{LR(\mathcal{A})}((s_1, i_1), \dots, (s_m, i_m))$  iff

- (i)  $\forall j \in [m], s_j = s$ ;
- (ii)  $p_{i_j}$  does not appear in  $s(i_j, \max\{i_1, \dots, i_m\})$ ;
- (iii)  $R^{\mathcal{A}}(\varepsilon_{\mathcal{A}}(s, i_1), \dots, \varepsilon_{\mathcal{A}}(s, i_m))$ .

There is a natural ordering  $\leq$  on  $LR(\mathcal{A})$  such that  $x \leq y$  iff  $x = (s, i), y = (s, j)$  and  $i \leq j$ . There is a pebbling function  $p: LR(\mathcal{A}) \rightarrow [k]$ , where  $p(s, i) = p_i$  for  $s = [(p_1, a_1), \dots, (p_n, a_n)]$ . The triple  $(LR(\mathcal{A}), \leq, p)$  is a linear forest satisfying condition (P), hence an object of  $\mathcal{R}_k^{PL}(\sigma)$ . This extends to a functor  $R_k: \mathbf{Struct}(\sigma) \rightarrow \mathcal{R}_k^{PL}(\sigma)$ , where  $R_k(\mathcal{A}) = (LR(\mathcal{A}), \leq, p)$ . The following proposition follows from the fact that  $L_k \dashv R_k$  is a comonadic adjunction [11, Theorem 4.9] and from Proposition 5.1.

**Proposition 5.2.**  *$L_k \dashv R_k$  is a resource-indexed linear arboreal cover of  $\mathbf{Struct}(\sigma)$  by  $\mathcal{R}_k^{PL}(\sigma)$ .*

The comonad  $(\mathbb{P}_k^L, \varepsilon, \delta)$  induced by this resource-indexed arboreal cover is a linear variant of  $\mathbb{P}_k$  and was the original motivation for the notion of linear arboreal category. This comonad was called the pebble-relation comonad  $\mathbb{P}\mathbb{R}_k$  by Montacute and Shah in [11]. There it was demonstrated that the existence of a morphism between two relational structures  $\mathcal{A}$  and  $\mathcal{B}$  in the coKleisli category over  $\mathbb{P}_k^L$  is equivalent to a

winning strategy for Duplicator in the all-in-one positive  $k$ -pebble game  $\exists^+ \mathbf{PPeb}_k(\mathcal{A}, \mathcal{B})$  between  $\mathcal{A}$  and  $\mathcal{B}$ . This game characterises preservation of sentences in the restricted conjunction fragment of existential positive  $k$ -variable logic  $\exists^+ \wedge \mathcal{L}^k$ . It was also shown that an isomorphism in the coKleisli category over  $\mathbb{P}_k^L$  is equivalent to a winning strategy for Duplicator in the all-in-one bijective  $k$ -pebble game between  $\mathcal{A}$  and  $\mathcal{B}$ . This game characterises equivalence of sentences in the fragment of  $k$ -variable logic with restricted conjunction and ‘walk counting’  $\# \wedge \mathcal{L}^k$ .

Applying the machinery we developed in Section 4, we address the open question in the conclusion of [11] by showing that pathwise embeddings, and thus bisimulation in the category of coalgebras of  $\mathbb{P}_k^L$ , isomorphic to  $\mathcal{R}_k^{PL}(\sigma)$  correspond to Duplicator winning strategies in the all-in-one  $k$ -pebble game with partial isomorphism winning condition  $\exists \mathbf{PPeb}_k(\mathcal{A}, \mathcal{B})$ . This game characterises preservation of sentences in the restricted conjunction fragment of existential, i.e. negations are only allowed on atomic formulas,  $k$ -variable logic  $\exists \wedge \mathcal{L}^k$ . This addition to [11] is summarised in the following theorem:

**Theorem 5.3.** *The following are equivalent for all  $\sigma$ -structures  $\mathcal{A}, \mathcal{B}$ :*

- (i) *There exists a bisimulation  $R_k(\mathcal{A}) \leftarrow X \rightarrow R_k(\mathcal{B})$ .*
- (ii) *There exist pathwise embeddings  $R_k(\mathcal{A}) \rightarrow R_k(\mathcal{B})$  and  $R_k(\mathcal{B}) \rightarrow R_k(\mathcal{A})$ .*
- (iii) *Duplicator has a winning strategy in  $\exists \mathbf{PPeb}_k(\mathcal{A}, \mathcal{B})$  and  $\exists \mathbf{PPeb}_k(\mathcal{B}, \mathcal{A})$ .*
- (iv)  $\mathcal{A} \equiv^{\exists \wedge \mathcal{L}^k} \mathcal{B}$ .

## 6 The modal linear arboreal cover

We will now turn to analysing a linear fragment of modal logic utilising the abstraction of linear arboreal categories. To this end, we assume that  $\sigma$  is a *modal signature*. That is, the relation symbols in  $\sigma$  have arity  $\leq 2$ . Recall the definition of the category of linear synchronisation trees  $\mathcal{R}^{ML}(\sigma)$  defined in Example 4.11. For all  $k > 0$ , we consider  $\mathcal{R}_k^{ML}(\sigma)$  to be the full subcategory of  $\mathcal{R}^{ML}(\sigma)$  consisting of linear synchronisation trees with height  $\leq k$ . Observe that  $\mathcal{R}^{ML}(\sigma)$  and  $\mathcal{R}_k^{ML}(\sigma)$  are the linear subcategories of  $\mathcal{R}^M(\sigma)$  and  $\mathcal{R}_k^M(\sigma)$ , respectively. The following proposition follows from Proposition 4.9 and the proof that  $\mathcal{R}^M(\sigma)$  and  $\mathcal{R}_k^M(\sigma)$  are linearisable arboreal categories [6,5].

**Proposition 6.1.**  *$\mathcal{R}^{ML}(\sigma)$  is a linear arboreal category and, for every  $k > 0$ ,  $\mathcal{R}_k^{ML}(\sigma)$  is a linear arboreal category.*

There is an obvious forgetful functor  $L: \mathcal{R}^{ML}(\sigma) \rightarrow \mathbf{Struct}_*(\sigma)$  mapping  $(\mathcal{A}, a_0, \leq)$  to  $(\mathcal{A}, a_0)$ . For each pointed relational structure  $(\mathcal{A}, a_0)$  with modal signature  $\sigma$ , we define a new structure  $LR(\mathcal{A}, a_0)$  with a universe

$$LR(\mathcal{A}, a_0) = \{a_0\} \cup \bigcup_{n \in \mathbb{N}} \{(s, i) \mid s \in \text{runs}_n(\mathcal{A}, a_0), i \in [n]\}, \text{ where}$$

$$\text{runs}_n(\mathcal{A}, a_0) = \{a_0 \xrightarrow{\alpha_1} a_1 \xrightarrow{\alpha_2} \dots \xrightarrow{\alpha_n} a_n \mid \alpha_i \in \text{Act}, a_i \in A, i \in [n]\}.$$

We define a counit map  $\varepsilon: LR(\mathcal{A}, a_0) \rightarrow (\mathcal{A}, a_0)$ , which preserves the distinguished point  $a_0$ , and the pair  $(s, i)$  is mapped to the  $i$ -th element in the run  $s$ . Unary relations  $P \in \sigma$  are interpreted as  $P^{LR(\mathcal{A}, a_0)}(s)$  iff  $P^{\mathcal{A}}(\varepsilon_{\mathcal{A}}(s))$ . For binary relations  $R_\alpha \in \sigma$ , the interpretation is  $R_\alpha^{LR(\mathcal{A}, a_0)}(x, y)$  iff either

- (i)  $x = a_0, y = (s, 1)$  and the first transition appearing in  $s$  is  $\alpha$ ; or
- (ii)  $x = (s, i), y = (s, i + 1)$  and the  $(i + 1)$ -th transition appearing in  $s$  is  $\alpha$ .

There is a natural ordering  $\leq$  on the universe  $LR(\mathcal{A}, a_0)$ , where  $x \leq y$  if either  $x = a_0$  and  $y = (s, 1)$ , or  $x = (s, i), y = (s, j)$  and  $i \leq j$ . The pair  $(LR(\mathcal{A}, a_0), \leq)$  is a linear tree that satisfies condition (M) as stated in Example 3.4. Consequently, we have that  $(LR(\mathcal{A}, a_0), a_0, \leq)$  is an object of  $\mathcal{R}^{ML}(\sigma)$ . This construction extends to a functor  $R: \mathbf{Struct}_*(\sigma) \rightarrow \mathcal{R}^{ML}(\sigma)$ , where  $R(\mathcal{A}, a_0) = (LR(\mathcal{A}, a_0), a_0, \leq)$ . The following proposition follows from Proposition 4.12 and [6, Theorem 9.6].

**Proposition 6.2.**  *$L \dashv R$  is a linear arboreal cover of  $\mathbf{Struct}_*(\sigma)$  by  $\mathcal{R}^{ML}(\sigma)$ .*

This arboreal cover can be resource-indexed: For every  $k > 0$ , the forgetful functor  $L_k: \mathcal{R}_k^{ML}(\sigma) \rightarrow \mathbf{Struct}_*(\sigma)$  has right adjoint  $R_k: \mathbf{Struct}_*(\sigma) \rightarrow \mathcal{R}_k^{ML}(\sigma)$ , where the universe of the underlying  $\sigma$ -structure and order of  $R_k(\mathcal{A}, a_0)$  is the induced  $\sigma$ -structure and induced order of  $R(\mathcal{A}, a_0)$  on the subset

$$\{a_0\} \cup \bigcup_{n \leq k} \{(s, i) \mid s \in \text{runs}_n(\mathcal{A}, a_0), i \in [n]\}.$$

The proof of Proposition 6.2 restricts to the height  $\leq k$  linear synchronisation trees in the subcategory  $\mathcal{R}_k^{ML}(\sigma)$  of  $\mathcal{R}^{ML}(\sigma)$ .

**Proposition 6.3.**  $L_k \dashv R_k$  is a resource-indexed linear arboreal cover of  $\mathbf{Struct}_*(\sigma)$  by  $\{\mathcal{R}_k^{ML}(\sigma)\}$ .

For all  $k > 0$ , the comonad  $(\mathbb{M}_k^L, \varepsilon, \delta)$  induced by the above resource-indexed arboreal cover is a linear variant of the modal comonad defined in [6]. In fact, as with the modal comonad  $\mathbb{M}_k$ , the linear variant  $\mathbb{M}_k^L$  is idempotent: The unit  $\eta$  of the adjunction in Proposition 6.3 is an isomorphism; i.e.  $a_i \in (\mathcal{A}, a_0, \leq)$  is mapped to  $(s, i)$ , where  $s$  is the unique linearly-ordered run containing  $a_i$  as its  $i$ -th element.

**Proposition 6.4.**  $\mathbb{M}_k^L$  is an idempotent comonad. Equivalently,  $\mathcal{R}_k^{ML}(\sigma)$  is a coreflective subcategory of  $\mathbf{Struct}_*(\sigma)$ .

Let  $ML_k$  denote the fragment of modal logic with modal-depth  $k$ . In [6] and [3], it was shown that the resource-indexed arboreal cover of  $\mathbf{Struct}_*(\sigma)$  by  $\{\mathcal{R}_k^M(\sigma)\}$  captures logical equivalence in  $ML_k$ , as well as in the following variants of  $ML_k$ :

- The fragment  $\exists^+ ML_k$ , which excludes negation and the operator  $\square$ ;
- The fragment  $\exists ML_k$ , which includes negation over propositional variables but excludes  $\square$ ;
- The extension  $\#ML_k$  which includes graded modalities, i.e. modalities of the form  $\diamond^{\leq k}$  defined for a Kripke model  $\mathcal{M}$  and a point  $w$  as  $(\mathcal{M}, w) \models \diamond^{\leq k} \varphi$  iff  $w$  has at most  $k$  successors satisfying  $\varphi$ . The modality  $\diamond^{\geq k}$  is defined similarly but with the interpretation of ‘at least’ instead.

We can show that the relations resulting from the resource-indexed arboreal cover of Proposition 6.3 recover behavioural relations, e.g. trace inclusion and bijective labelled trace equivalence, and equivalence in their corresponding linear fragments of modal logic. We begin by defining the linear fragment  $\exists \wedge ML_k$ .

A formula  $\varphi \in ML_k$  is called *linear* if each conjunction in every proper subformula of  $\varphi$  contains at most one formula with modal operations. Explicitly, the language  $\exists \wedge ML_k$  can be defined recursively as

$$\varphi ::= \varphi \wedge \varphi \mid \diamond_\alpha \psi, \quad \psi ::= \top \mid \perp \mid \psi \vee \psi \mid p \wedge \psi \mid \neg p \wedge \psi \mid \diamond_\alpha \psi, \quad p \in \mathbf{Var}, \alpha \in \mathbf{Act}. \quad (2)$$

This entails for example that  $\diamond(\diamond p \wedge \diamond q)$  is not a linear modal formula. Accordingly, let  $\wedge ML_k$  denote the fragment of  $ML_k$  in which every formula is linear.

Below, given a list  $\alpha = [\alpha_1, \dots, \alpha_n] \in \mathbf{Act}^*$  and pointed  $\sigma$ -structure  $(\mathcal{A}, a_0)$ , we define the set of processes

$$\text{proc}_\alpha(\mathcal{A}, a_0) := \{a_0 \xrightarrow{\alpha_1} a_1 \xrightarrow{\alpha_2} \dots \xrightarrow{\alpha_n} a_n \mid a_i \in A, i \in [n]\}.$$

Note that  $\bigcup_{\alpha \in \mathbf{Act}^n} \text{proc}_\alpha(\mathcal{A}, a_0) = \text{runs}_n(\mathcal{A}, a_0)$ .

Let  $\# \wedge ML_k$  denote the extension of  $\exists \wedge ML_k$  with walk counting modalities  $\diamond_\alpha^m$ , where  $\alpha = [\alpha_1, \dots, \alpha_n] \in \mathbf{Act}^*$  and  $m$  is a positive integer. Given a pointed Kripke structure  $(\mathcal{A}, a)$ , the semantics of these walk counting modalities  $\diamond_\alpha^m$  is defined by induction on the length of  $\alpha$ :

- $\mathcal{A}, a \models \diamond_\epsilon^1 \varphi$  if and only if  $\mathcal{A}, a \models \varphi$
- $\mathcal{A}, a \models \diamond_{[\beta]\alpha}^m \varphi$  if for every  $m_\alpha, m_\beta \in \mathbb{N}$  where  $m = m_\alpha m_\beta$ , there exists  $m_\beta$ -many  $\beta$ -successors  $a'$  of  $a$  such that  $\mathcal{A}, a' \models \diamond_\alpha^{m_\alpha} \varphi$ .

Recall that  $[\beta]\alpha$  denotes the concatenation of  $[\beta]$  for  $\beta \in \mathbf{Act}$  and  $\alpha = [\alpha_1, \dots, \alpha_n] \in \mathbf{Act}^*$ .

The syntax of  $\# \wedge ML_k$  is the same as of  $\exists \wedge ML_k$  but with the addition of formulas of the form  $\diamond_\alpha^m \psi$  to the  $\varphi$  grammar of Equation (2).

In this paper, we consider these linear fragments  $\exists^+ \wedge ML_k$ ,  $\exists \wedge ML_k$  and  $\# \wedge ML_k$ .

**Definition 6.5** (bijective labelled trace equivalence). Given two pointed Kripke structures  $(\mathcal{A}, a_0)$  and  $(\mathcal{B}, b_0)$  in  $\mathbf{Struct}_*(\sigma)$ , we write  $(\mathcal{A}, a_0) \sim^{\text{bltr}} (\mathcal{B}, b_0)$  if for every  $\alpha = [\alpha_1, \dots, \alpha_n]$ , there exists a bijection

$$f_\alpha : \text{proc}_\alpha(\mathcal{A}, a_0) \rightarrow \text{proc}_\alpha(\mathcal{B}, b_0),$$

such that for each  $s = a_0 \xrightarrow{\alpha_1} a_1 \dots \xrightarrow{\alpha_n} a_n$  and  $t = b_0 \xrightarrow{\alpha_1} b_1 \dots \xrightarrow{\alpha_n} b_n$  with  $f_\alpha(s) = t$ , we have  $V(a_i) = V(b_i)$ , for all  $i \geq 0$ .

Whenever  $(\mathcal{A}, a_0) \sim^{\text{bltr}} (\mathcal{B}, b_0)$ , we say that  $(\mathcal{A}, a_0)$  and  $(\mathcal{B}, b_0)$  are *bijective labelled trace equivalent*. As with  $\subseteq^{\text{tr}}$  and  $\sim^{\text{ltr}}$ , we can also obtain the relation  $\sim_k^{\text{bltr}}$  by restricting the definition of  $\sim^{\text{bltr}}$  to traces of length  $\leq k$ .

Bijective labelled trace equivalence is incomparable with bisimulation. To see why, consider the structures in Figure 2. These are clearly non-bisimilar yet there is a bijection between their labelled traces, hence they are bijective labelled trace equivalent. On the other hand, by taking a point with a single  $\alpha$ -transition and comparing it to a point with a pair of  $\alpha$ -transitions, we obtain two transition systems which are bisimilar but not bijective labelled trace equivalent.

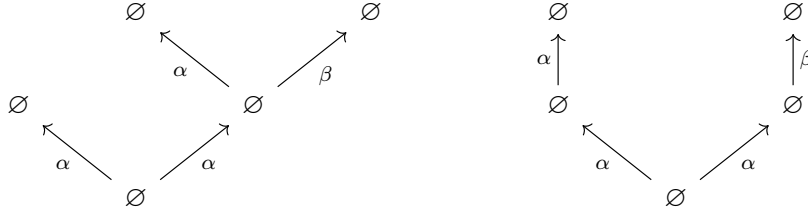


Fig. 2. bijective labelled trace equivalent but non-bisimilar transition systems

Below, let  $\Rightarrow_{\mathcal{L}}$  and  $\equiv_{\mathcal{L}}$  denote truth preservation and equivalence in  $\mathcal{L}$ , respectively.

**Theorem 6.6.** Let  $\mathcal{C} = \mathcal{R}_k^{ML}(\sigma)$ . For all  $(\mathcal{A}, a_0), (\mathcal{B}, b_0) \in \mathbf{Struct}_*(\sigma)$ , the following statements hold:

- (i)  $(\mathcal{A}, a_0) \rightarrow_k^{\mathcal{C}} (\mathcal{B}, b_0)$  iff  $(\mathcal{A}, a_0) \subseteq_k^{\text{tr}} (\mathcal{B}, b_0)$  iff  $(\mathcal{A}, a_0) \Rightarrow_{\exists^+ \wedge ML_k} (\mathcal{B}, b_0)$ .
- (ii)  $(\mathcal{A}, a_0) \leftrightarrow_k^{\mathcal{C}} (\mathcal{B}, b_0)$  iff  $(\mathcal{A}, a_0) \sim_k^{\text{ltr}} (\mathcal{B}, b_0)$  iff  $(\mathcal{A}, a_0) \equiv_{\exists \wedge ML_k} (\mathcal{B}, b_0)$ .
- (iii)  $(\mathcal{A}, a_0) \cong_k^{\mathcal{C}} (\mathcal{B}, b_0)$  iff  $(\mathcal{A}, a_0) \sim_k^{\text{bltr}} (\mathcal{B}, b_0)$  iff  $(\mathcal{A}, a_0) \equiv_{\# \wedge ML_k} (\mathcal{B}, b_0)$ , whenever  $(\mathcal{A}, a_0)$  and  $(\mathcal{B}, b_0)$  are image-finite.

Surprisingly, unlike in previous cases of arboreal covers [3], the back-and-forth relation  $\leftrightarrow_k^{\mathcal{C}}$  does not capture the full linear fragment  $\wedge ML_k$ . Instead, it captures the fragment  $\exists \wedge ML_k$ .

## 7 Preservation theorems

In his seminal work, Rossman [14] proved the equirank preservation theorem showing that a first order sentence  $\varphi$  with quantifier rank  $r$  is preserved under homomorphisms if, and only if,  $\varphi$  is equivalent to an existential positive formula with quantifier rank  $r$ . This refined the classical homomorphism preservation theorem, due to Łoś, Lyndon and Tarski, and was crucial towards the development of Rossman's celebrated finite homomorphism preservation theorem.

The perspective of arboreal categories provides an axiomatic framework for proving Rossman-style theorems by determining sufficient conditions an arboreal cover must satisfy in order for a similar characterisation to hold for the corresponding logic [4]. Arboreal covers for weaker logics satisfy stronger sufficient conditions, enabling a uniform proof of Rossman-style theorems over finite and all structures. In this section, we apply this perspective to prove a novel Rossman preservation theorem for the linear fragment characterising labelled trace equivalence.

Given a full subcategory  $\mathcal{D}$  of a category  $\mathcal{C}$ , we say that  $\mathcal{D}$  is *saturated* under some equivalence relation  $\sim$ , if  $a \in \mathcal{D}$  and  $a \sim b$  imply  $b \in \mathcal{D}$ , for all  $a, b \in \mathcal{C}$ . The saturated full subcategories  $\mathcal{D}$  allow us to talk about structures which are models of sentences.

For every modal formula  $\varphi$  in  $\wedge ML_k$ ,  $\exists \wedge ML_k$ , or  $\# \wedge ML_k$ , we denote by  $\mathbf{Mod}(\varphi)$  the full subcategory  $\mathbf{Struct}_*(\sigma)$  of models of  $\varphi$ . Let  $\rightarrow_k$ ,  $\leftrightarrow_k$  and  $\cong_k$  be the relations from Definition 3.7 induced by the resource-indexed arboreal cover of  $\mathcal{E}$  by  $\{\mathcal{R}_k^{ML}(\sigma)\}$ .

**Proposition 7.1.** *Let  $\mathcal{D}$  be a full subcategory of  $\mathbf{Struct}_*(\sigma)$ . Then*

- (a)  $\mathcal{D}$  is upward closed under  $\rightarrow_k$  iff  $\mathcal{D} = \mathbf{Mod}(\varphi)$  for some  $\varphi \in \exists^+ \wedge ML_k$ .
- (b)  $\mathcal{D}$  is saturated under  $\leftrightarrow_k$  iff  $\mathcal{D} = \mathbf{Mod}(\varphi)$  for some  $\varphi \in \exists \wedge ML_k$ .
- (c) If  $\mathcal{D} = \mathbf{Mod}(\varphi)$  for some  $\varphi \in \# \wedge ML_k$ , then  $\mathcal{D}$  is saturated under  $\cong_k$ .

Let  $\mathcal{C}$  be a resource-indexed arboreal category such that there is a resource-indexed arboreal cover of a category  $\mathcal{E}$  by  $\mathcal{C}$ . Consider the statement

(HP) If  $\mathcal{D}$  is a full subcategory of  $\mathcal{C}$  saturated under  $\leftrightarrow_k$ , then  $\mathcal{D}$  is closed under morphisms, i.e. if  $X \in \mathcal{D}$  and  $X \rightarrow Y \in \mathcal{C}$ , then  $Y \in \mathcal{D}$ , iff it is closed under  $\rightarrow_k$ .

We replace  $\leftrightarrow_k$  with  $\cong_k$  to obtain the stronger version

(HP#) If  $\mathcal{D}$  is a full subcategory of  $\mathcal{C}$  saturated under  $\cong_k$ , then  $\mathcal{D}$  is closed under morphisms iff it is closed under  $\rightarrow_k$ .

Consider a resource-indexed arboreal cover  $L_k \dashv R_k$  of  $\mathcal{E}$  by  $\{\mathcal{C}_k\}$ . The following is Proposition 4.5 and 4.7 of [4].

**Lemma 7.2.** *If  $a \leftrightarrow_k \mathbb{C}_k a$  for all  $a \in \mathcal{E}$ , then (HP) holds. If in addition  $\mathbb{C}_k$  is idempotent, then (HP#) holds.*

Given an ambient category  $\mathcal{C}$ , a relation  $\bowtie$ , a logical fragment  $\mathcal{L}$  and a formula  $\varphi \in \mathcal{L}$ , we say that  $\varphi$  is *preserved under*  $\bowtie$  if for all  $\mathcal{A}, \mathcal{B} \in \mathcal{C}$ ,  $\mathcal{A} \bowtie \mathcal{B}$  implies that  $\mathcal{A} \models \varphi \Rightarrow \mathcal{B} \models \varphi$ ; we say that  $\varphi$  is *invariant under*  $\bowtie$  if for all  $\mathcal{A}, \mathcal{B} \in \mathcal{C}$ ,  $\mathcal{A} \bowtie \mathcal{B}$  implies that  $\mathcal{A} \models \varphi \Leftrightarrow \mathcal{B} \models \varphi$ .

Since  $\mathbb{M}_k^L$  is an idempotent comonad by Proposition 6.4, we obtain the following result by applying Lemma 7.2.

**Corollary 7.3.** *A linear modal sentence  $\varphi \in \exists \wedge ML_k$  is preserved under trace inclusion in finite structures if, and only if, it is logically equivalent to a formula  $\psi \in \exists^+ \wedge ML_k$ .*

Condition (HP) ensures that the bound on the modal depth between the two fragments does not change.

**Corollary 7.4.** *A linear modal sentence with walk counting modalities  $\varphi \in \# \wedge ML_k$  is preserved under trace inclusion in finite structures if, and only if, it is logically equivalent to a formula  $\psi \in \exists^+ \wedge ML_k$ .*

## 8 Characterisation theorem

One of the most celebrated results in modal logic is the Van Benthem Characterisation theorem [15]. It states that modal logic is the bisimulation invariant fragment of first order logic. In other words, this demonstrates that modal logic is as expressive as first-order logic with respect to bisimulation invariant properties. Rosen [13] extended the result to finite Kripke frames and Otto [12] later provided a generalisation consolidating the two proofs into one.

In this section, we prove a Van Benthem-Rosen characterisation theorem demonstrating that a first-order formula  $\varphi$  of rank  $r$  is invariant under labelled trace equivalence if, and only if, it is logically equivalent to a formula in  $\exists \wedge ML_k$ , where  $k = 2^r$ . The proof will follow a similar strategy to Otto's proof [12]. Namely, as Proposition 7.1 demonstrates, formulas in  $\exists \wedge ML_k$  can be identified with subcategories which are saturated under  $\sim_k^{\text{ltr}}$ , or equivalently,  $\leftrightarrow_k^{\mathcal{C}}$  for  $\mathcal{C} = \mathcal{R}_k^{ML}(\sigma)$ . Thus, to prove the characterisation theorem, one has to show that the class of models satisfying a first-order formula  $\varphi$  is saturated under  $\sim_k^{\text{ltr}}$



whenever  $\varphi$  is invariant under labelled trace equivalence.

**Theorem 8.1** (Van Benthem-Rosen characterisation theorem). *Let  $\varphi$  be a first-order-formula of quantifier rank  $r$ . Then  $\varphi$  is invariant under labelled trace equivalence iff  $\varphi$  is logically equivalent to a formula  $\psi \in \exists \wedge ML_k$ , where  $k = 2^r$ .*

As mentioned above, the characterisation theorem has two versions depending on whether the ambient category of the evaluated pointed structures is infinite i.e.  $\mathbf{Struct}_*(\sigma)$ , or finite i.e.  $\mathbf{Struct}_*^f(\sigma)$ . Following Otto[12], we aim to provide a uniform consolidated proof that captures both versions.

We use the notation  $\mathcal{A} + \mathcal{B}$  for the standard disjoint union of two structures  $\mathcal{A}, \mathcal{B} \in \mathbf{Struct}(\sigma)$ , i.e. the coproduct in  $\mathbf{Struct}(\sigma)$ . Given a sentence  $\varphi$ , we say that  $\varphi$  is *invariant under disjoint extensions* if for all  $(\mathcal{A}, a) \in \mathbf{Struct}_*(\sigma)$  and  $\mathcal{B} \in \mathbf{Struct}(\sigma)$  we have

$$(\mathcal{A}, a) \models \varphi \iff (\mathcal{A} + \mathcal{B}, a) \models \varphi.$$

As the elements of  $\mathcal{B}$  are not reachable from  $a$ , the number of traces does not increase under disjoint extensions and thus  $(\mathcal{A}, a) \sim^{\text{ltr}} (\mathcal{A} + \mathcal{B}, a)$ . Accordingly, we obtain the following:

**Lemma 8.2.** *Invariance under labelled trace equivalence implies invariance under disjoint extensions.*

Given  $\mathcal{A} \in \mathbf{Struct}_*(\sigma)$ , we define a metric  $d$  on  $A$  as  $d(a, b) = n \in \mathbb{N}$  whenever the path distance between  $a$  and  $b$  in the Gaifman graph  $\mathcal{G}(\mathcal{A})$  is  $n$ , and  $d(a, b) = \infty$  whenever there is no path between  $a$  and  $b$  in  $\mathcal{G}(\mathcal{A})$ . We write  $A[a; k]$  for the ball centred on  $a$  with radius  $k$ . Given  $(\mathcal{A}, a) \in \mathbf{Struct}_*(\sigma)$ , we define  $\mathbb{S}_k(\mathcal{A}, a) = (\mathcal{A}[a; k], a)$ , where  $(\mathcal{A}[a; k], a)$  is the substructure induced by  $A[a; k]$ .

Let  $\equiv_r$  denote equivalence in first-order logic up to quantifier rank  $r$ . A key step in our argument for the characterisation theorem is a general result called the Workspace Lemma. Intuitively, we would like to show that a structure  $\mathcal{A}$  is  $\equiv_r$ -equivalent to a local window  $\mathcal{A}[a; k]$  inside of  $\mathcal{A}$ . The obvious problem with demonstrating this through an  $r$ -round Ehrenfeucht-Fraïssé game is that Spoiler can win by playing outside of the local window  $\mathcal{A}[a; k]$ . The Workspace Lemma states that if we expand both  $\mathcal{A}$  and  $\mathcal{A}[a; k]$  by a structure  $\mathcal{B}$  consisting of sufficiently many disjoint copies of  $\mathcal{A} + \mathcal{A}[a; k]$ , Duplicator can use the additional workspace in  $\mathcal{B}$  to evade Spoiler's non-local moves and win the  $r$ -round Ehrenfeucht-Fraïssé game between the expanded structures. The proof of the Workspace Lemma can be found in [2].

**Lemma 8.3** (workspace lemma). *Given  $(\mathcal{A}, a) \in \mathbf{Struct}_*(\sigma)$  and  $r > 0$ , there exists  $\mathcal{B} \in \mathbf{Struct}(\sigma)$  such that  $(\mathcal{A} + \mathcal{B}, a) \equiv_r (\mathcal{A}[a; k] + \mathcal{B}, a)$ , where  $k = 2^r$ . Moreover,  $\mathcal{B}$  is bounded by the size of  $\mathcal{A}$  as  $|\mathcal{B}| \leq 2r|\mathcal{A}|$ . Therefore, if  $\mathcal{A}$  is finite, then so is  $\mathcal{B}$ .*

The key feature of tree models, such as  $\mathbb{M}^L(\mathcal{A}, a)$  or the modal comonad  $\mathbb{M}(\mathcal{A}, a)$ , in proofs of characterisation theorems, is that they satisfy a *companion property*, e.g.  $\mathbb{M}^L(\mathcal{A}, a) \sim^{\text{ltr}} (\mathcal{A}, a)$ . However, since  $\mathbb{M}^L(\mathcal{A}, a)$  is always infinite, this creates an obstacle for a uniform proof of the characterisation theorem that works for both finite and infinite structures. Instead of using  $\mathbb{M}^L(\mathcal{A}, a)$ , we use a structure  $\mathbb{M}^{L(k)}(\mathcal{A}, a)$  which is finite whenever  $(\mathcal{A}, a)$  is finite. Intuitively, this construction unravels the first  $k$ -steps of a trace into elements of  $\mathbb{M}_k^L(\mathcal{A}, a)$  and then proceeds to finish the remainder of the trace in a copy of  $\mathcal{A}$ . We define

$$\mathbb{M}^{L(k)}(\mathcal{A}, a) := \left( \mathbb{M}_k^L(\mathcal{A}, a) + \sum_{(s,k) \in \mathbb{M}_k^L(\mathcal{A}, a), |s|=k} \mathcal{A} \right) / \simeq, \langle a \rangle,$$

where  $\simeq$  is the equivalence relation generated from  $(s, k) \simeq ((s, k), \varepsilon_{\mathcal{A}}(s, k))$  and  $\langle a \rangle$  is the equivalence class with the representative  $a$  under the equivalence relation  $\simeq$ .

In other words, we graft a disjoint copy of  $\mathcal{A}$  onto each leaf of  $\mathbb{M}_k^L(\mathcal{A}, a)$  labelled by the maximal path  $s$  paired with its last index. Note that  $\mathbb{M}^{L(k)}(\mathcal{A}, a)$  is bounded by the size of  $\mathcal{A}$ . In particular, the size of  $\mathbb{M}^{L(k)}(\mathcal{A}, a)$  is at most  $|\mathbb{M}_k^L(\mathcal{A}, a)|(1 + |\mathcal{A}|)$ .

**Proposition 8.4.** *For all  $(\mathcal{A}, a) \in \mathbf{Struct}_*(\sigma)$ ,  $(\mathcal{A}, a) \sim^{\text{ltr}} \mathbb{M}^{L(k)}(\mathcal{A}, a)$ .*

We can relate  $\mathbb{M}_k^L$ ,  $\mathbb{M}^{L(k)}$  and  $\mathbb{S}_k$  in the following way.

**Proposition 8.5.** *Given  $(\mathcal{A}, a) \in \mathbf{Struct}_*(\sigma)$ ,  $\mathbb{S}_k(\mathbb{M}^{L(k)}(\mathcal{A}, a)) \cong \mathbb{M}_k^L(\mathcal{A}, a)$ .*

We are now ready to prove the main part of the characterisation theorem.

**Proposition 8.6.** *Let  $\varphi$  be a first-order formula of quantifier rank  $r$ . If  $\varphi$  is invariant under labelled trace equivalence, then it is equivalent to a formula  $\psi \in \exists \wedge ML_k$ , where  $k = 2^r$ .*

The proof of Theorem 8.1 involves categorical constructions such as invariance under coproducts and the comonads  $\mathbb{M}_k^L$  and  $\mathbb{S}_k$ . More specifically, we use Lemma 8.2 to show that  $\varphi$  is invariant under disjoint extensions. We can then use Lemma 8.3 and Theorem 6.6.(ii) together with propositions 8.4 and 8.5 to derive that  $(\mathcal{B}, b) \models \varphi$  by assuming  $(\mathcal{A}, a) \models \varphi$  and  $(\mathcal{A}, a) \sim_k^{\text{tr}} (\mathcal{B}, b)$ . Since  $\sim^{\text{tr}} \subseteq \sim_k^{\text{tr}}$ , it follows by Theorem 6.6.(ii) that formulas in  $\exists \wedge ML_k$  are invariant under labelled trace equivalence. By combining Theorem 6.6.(ii) and Proposition 8.6, we obtain the characterisation theorem in Theorem 8.1, addressing both the finite and infinite cases.

## 9 Future Work

For future work, we highlight two possible research avenues: First, one can investigate the construction in a linear arboreal category which captures truth preservation in  $\wedge ML_k$ ; this corresponds to a relation reminiscent (albeit incomparable) of *ready trace equivalence* [16]. More generally, one can provide categorical semantics for other relations in the linear-time branching-time spectrum in terms of linear arboreal and arboreal categories. This work will explore how to adapt different notions of coalgebraic bisimulation, such as cospans of open maps, to generalise these relations.

Second, there are many similarities linking our research and the strand of research categorifying behavioural relations via graded monads and fibrations. In particular, linear arboreal categories centre around a notion of paths and utilise open maps as in [18]; furthermore, they employ resource indexing and grading as in [9]. On the other hand, there are some obvious differences.

In particular, these lines of research study coalgebras over endofunctors, monads, and graded monads; whereas we study coalgebras over comonads. The conceptual perspectives behind these formalisms also differ, in a somewhat subtle fashion. Relating these approaches may yield valuable new insights.

## Acknowledgement

The authors would like to thank Luca Reggio for helpful discussions related to linearisable arboreal categories.

## References

- [1] Samson Abramsky, Anuj Dawar, and Pengming Wang. The pebbling comonad in finite model theory. In *Logic in Computer Science (LICS), 2017 32nd Annual ACM/IEEE Symposium on*, pages 1–12. IEEE, 2017. <https://doi.org/10.1109/LICS.2017.8005129>
- [2] Samson Abramsky and Dan Marsden. Comonadic semantics for hybrid logic and bounded fragments, 2021. arXiv preprint. <https://doi.org/10.48550/arXiv.2110.09844>
- [3] Samson Abramsky and Luca Reggio. Arboreal Categories and Resources. In Nikhil Bansal, Emanuela Merelli, and James Worrell, editors, *48th International Colloquium on Automata, Languages, and Programming (ICALP 2021)*, volume 198 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 115:1–115:20, Dagstuhl, Germany, 2021. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. <https://drops.dagstuhl.de/storage/00lipics/lipics-vol198-icalp2021/LIPIcs.ICALP.2021.115/LIPIcs.ICALP.2021.115.pdf>
- [4] Samson Abramsky and Luca Reggio. Arboreal categories and homomorphism preservation theorems, 2022. <https://arxiv.org/abs/2211.15808>
- [5] Samson Abramsky and Luca Reggio. Arboreal Categories: An Axiomatic Theory of Resources. *Logical Methods in Computer Science*, Volume 19, Issue 3, August 2023. <https://doi.org/10.48550/arXiv.2102.08109>

- [6] Samson Abramsky and Nihil Shah. Relating structure and power: Comonadic semantics for computational resources. *Journal of Logic and Computation*, 31(6):1390–1428, August 2021.  
<https://doi.org/10.1093/logcom/exab048>
- [7] Víctor Dalmau. Linear datalog and bounded path duality of relational structures. *Log. Methods Comput. Sci.*, 1(1), 2005.  
<https://doi.org/10.2168/LMCS-1%281%3A5%292005>
- [8] Chase Ford, Stefan Milius, and Lutz Schröder. Behavioural preorders via graded monads. In *Proceedings of the 36th Annual ACM/IEEE Symposium on Logic in Computer Science*, LICS '21, New York, NY, USA, 2021. Association for Computing Machinery.  
<https://doi.org/10.1109/LICS52264.2021.9470517>
- [9] Chase Ford, Stefan Milius, Lutz Schröder, Harsh Beohar, and Barbara König. Graded monads and behavioural equivalence games. In *Proceedings of the 37th Annual ACM/IEEE Symposium on Logic in Computer Science*, LICS '22, New York, NY, USA, 2022. Association for Computing Machinery.  
<https://doi.org/10.1145/3531130.3533374>
- [10] André Joyal, Mogens Nielsen, and Glynn Winskel. Bisimulation from open maps. *Information and Computation*, 127(2):164–185, June 1996.  
<https://doi.org/10.1006/inco.1996.0057>
- [11] Yoàv Montacute and Nihil Shah. The pebble-relation comonad in finite model theory. In *Proceedings of the 37th Annual ACM/IEEE Symposium on Logic in Computer Science*, LICS '22, New York, NY, USA, 2022. Association for Computing Machinery.  
<https://doi.org/10.1145/3531130.3533335>
- [12] Martin Otto. Bisimulation invariance and finite models. In *Logic Colloquium*, volume 2, pages 276–298, 2006.  
<http://dx.doi.org/10.1017/9781316755723.013>
- [13] Eric Rosen. Modal logic over finite structures. *Journal of Logic, Language, and Information*, 6(4):427–439, 1997.  
<https://www.jstor.org/stable/40181601>
- [14] Benjamin Rossman. Homomorphism preservation theorems. *Journal of the ACM (JACM)*, 55(3):1–53, 2008.  
<https://doi.org/10.1145/1379759.1379763>
- [15] Johan van Benthem. *Modal correspondence theory*. PhD thesis, University of Amsterdam, 1976.  
<https://eprints.illc.uva.nl/id/eprint/1838/>
- [16] R. J. van Glabbeek. The linear time - branching time spectrum. In J. C. M. Baeten and J. W. Klop, editors, *CONCUR '90 Theories of Concurrency: Unification and Extension*, pages 278–297, Berlin, Heidelberg, 1990. Springer Berlin Heidelberg.  
<https://doi.org/10.1007/BFb0039066>
- [17] R. J. van Glabbeek. The linear time — branching time spectrum ii. In Eike Best, editor, *CONCUR'93*, pages 66–81, Berlin, Heidelberg, 1993. Springer Berlin Heidelberg.  
[https://doi.org/10.1007/3-540-57208-2\\_6](https://doi.org/10.1007/3-540-57208-2_6)
- [18] Thorsten Wißmann, Jérémy Dubut, Shin-ya Katsumata, and Ichiro Hasuo. Path category for free. In Mikołaj Bojańczyk and Alex Simpson, editors, *Foundations of Software Science and Computation Structures*, pages 523–540, Cham, 2019. Springer International Publishing.  
[https://doi.org/10.1007/978-3-030-17127-8\\_30](https://doi.org/10.1007/978-3-030-17127-8_30)