A VAN BENTHEM/ROSEN THEOREM FOR COALGEBRAIC PREDICATE LOGIC

LUTZ SCHRÖDER, DIRK PATTINSON, AND TADEUSZ LITAK

ABSTRACT. Coalgebraic modal logic serves as a unifying framework to study a wide range of modal logics beyond the relational realm, including probabilistic and graded logics as well as conditional logics and logics based on neighbourhoods and games. Coalgebraic predicate logic (CPL), a generalization of a neighbourhood-based first-order logic introduced by Chang, has been identified as a natural first-order extension of coalgebraic modal logic, which in particular coincides with the standard first-order correspondence language when instantiated to Kripke-style relational modal operators. Here, we generalize to the CPL setting the classical van Benthem/Rosen theorem stating that both over arbitrary and over finite models, modal logic is precisely the bisimulation-invariant fragment of first-order logic. As instances of this generic result, we obtain corresponding characterizations for, e.g., conditional logic, neighbourhood logic (i.e., classical modal logic), and monotone modal logic.

1. INTRODUCTION

Van Benthem's [vB76] characterization of modal logic as precisely the bisimulation-invariant fragment of first-order logic over relational models is a pillar of *modal correspondence theory* [vB84, GO06]. This result has been extended to finite models by Rosen [Ros97], and special frame classes have been considered by Dawar and Otto [DO05]. Results of this kind guarantee that modal logic is as strong as first-order logic when it comes to expressing bisimulation-invariant properties, complementing the much simpler upper bound stating that all modal formulas are bisimulation-invariant. From the perspective of first-order logic, the van Benthem/Rosen theorem provides an effective syntax for bisimulation-invariant properties [Ott06].

Here, we extend these results to coalgebraic modal logic [Pat03], thus making initial forays into *coalgebraic correspondence theory*. Coalgebraic modal logic is a generic framework that captures a wide range of modal logics from the literature, e.g. the modal logic of neighbourhood frames (called classical modal logic by Chellas [Che80]), monotone modal logic, normal modal logics [BdRV01], graded and probabilistic modal logics [Fin72, LS91, HM01], and various conditional logics [Che80]; note in particular that coalgebraic logic is not tied to relational semantics. The parameters of the framework are a *type functor*, whose coalgebras serve as models, and a choice of *predicate liftings* defining the modal operators.

An important point to be settled here is the design of the first-order correspondence language. In an earlier version of this work [SP10a], we opted for a multi-sorted first-order logic inspired by the correspondence language for neighbourhood frames of Hansen et al. [HKP09]. This language included, in particular, an explicit sort for subsets of the state space (i.e. *neighbourhoods*). Hence, it required specific tweaks in the semantics to avoid second-order effects invalidating the target result: the bisimulation-invariant fragment of monadic second order logic is the μ-calculus rather than basic modal logic [JW95]. In the meantime, *coalgebraic predicate logic (CPL)* has been identified as a less expressive but more natural correspondence language [LPSS12], and we present our results in terms of this language here. CPL is based on a language for neighbourhood models introduced by Chang [Cha73] extending the first-order syntax with the construct

$$x\heartsuit\lceil y : \phi\rceil,$$

where $\heartsuit$ is a modal operator and $\lceil y : \phi\rceil$ is set comprehension $\{y \mid \phi\}$. Reading $\heartsuit$ informally as 'necessarily/possibly/probably...', $x\heartsuit\lceil y : \phi\rceil$ states that '$\phi(y)$ necessarily/possibly/probably... holds at x', where the free variable y in ϕ ranges over alternative worlds while any other free variables in ϕ are viewed as parameters. In CPL, we retain this syntax but allow $\heartsuit$ to be interpreted as any modal operator with a coalgebraic semantics. For example, in the probabilistic setting, which has modal operators L_p read 'with probability at least p', the formula $xL_p\lceil y : y \neq x\rceil$ expresses that at x, the probability of moving to a different state is $\geq p$. As another example we can take $\heartsuit$ to be the usual $\Diamond$, interpreted over Kripke frames, and thus exactly reproduce the classical relational correspondence language, so that our results literally generalize the classical van Benthem/Rosen theorem. In this setting, $x\Diamond\lceil y : \phi\rceil$ states that there exists a successor y of x satisfying $\phi(y)$; in particular $x\Diamond\lceil y : y = z\rceil$ states that z is a successor of x. A further indication of the naturality of CPL is that, in analogy to the classical case, it is expressively equivalent to several natural variants of coalgebraic hybrid logic [LPS13].

Technically, we adapt the method of Rosen and a related proof by Otto [Ott06] to prove that, under suitable assumptions, coalgebraic modal logic is, both over finite and over arbitrary models, precisely the fragment of CPL characterized by *invariance under behavioural equivalence*. As Rosen's method avoids compactness and saturation, which feature prominently in the original proof of van Benthem's theorem, we can deal also with classes of coalgebras that fail to be first-order axiomatizable. The latter is a fairly typical phenomenon, witnessed, e.g., by conditional or probabilistic interpretations of modal operators [Sch07]. Like Otto's proof, our proof relies on Gaifman locality. As CPL is not immediately equivalent to a standard first-order language, we need to invest some effort to inherit Gaifman locality; we achieve this by enriching CPL with a *support* relation that gives access to a *supporting Kripke frame* [SP09] of the coalgebraic model at hand, and then translate into a standard first-order language extending the language of Hansen et al. [HKP09].

To show that a first-order formula that is invariant under behavioural equivalence is equivalent to a *finitary* modal formula, we (frustratingly) have to assume that the underlying signature functor preserves finite sets. This covers Kripke and neighbourhood semantics, as well as the selection function semantics of conditional logic and a bounded version of graded modal logic, but excludes, e.g., probabilistic modal logic and full graded modal logic. For the general case, we do provide a characterization result in terms of bounded-rank modal formulas with infinitary conjunction. This result applies to essentially all logics of interest. Moreover, the finitary result is actually just an easy corollary for the case of finite similarity types, using no more than the observation that over finite similarity types, there are, up to equivalence, only finitely many modal formulas of a given bounded depth. As an application, we obtain, e.g., that every formula in a natural first order logic with counting quantifiers over multigraphs that is invariant under behavioural equivalence (over finite or arbitrary models) is equivalent (over the same class of models) to a possibly infinitary but bounded-depth formula in graded modal logic.

Related Work. As indicated above, we take the syntax and the basic idea of the semantics of CPL from Chang [Cha73]. Chang's original language applies to neighbourhood frames only, and as such is a fragment of the correspondence language for neighbourhood logic studied by Hansen et al. [HKP09], who in fact prove a van Benthem theorem for their language. The latter hence implies the neighbourhood frame instance of the van Benthem version of our generic result; the Rosen version (over finite models) appears to be new.

First-order formalisms related to Chang's language have been considered for topological spaces, which happen to be particular instances of neighbourhood frames when defined in terms of local neighbourhood bases [MM77, Sgr80, Zie85]. Ten Cate et al. [tCGS09] prove a van Benthem theorem for topological modal logic within a correspondence language called L^t, which properly contains the topological variant of Chang's language [MZ80]. This theorem is similar in spirit to both the van Benthem theorem for neighbourhood logic [HKP09] and to the van Benthem version of the instance of our generic theorem for monotone logic; due to the use of topological models, however, it is formally incomparable to both of these results.

The only other van Benthem type result we know of outside standard relational logics is de Rijke's characterization of graded modal logic over (possibly infinite) *Kripke frames* as the *g-bisimulation*-invariant fragment of first-order logic with counting quantifiers; our own results for graded logic over *multigraphs* are orthogonal to this result.

A different generic first-order logic largely concerned with the Kleisli category of a monad rather than with coalgebras for a functor is introduced and studied by Jacobs [Jac10]; as stated explicitly in op. cit., the nature of the connection, if any, to coalgebraic first order languages remains a subject for future research.

This work is an extended and reworked version of an earlier conference paper [SP10a], in which we used a different and more involved correspondence language as discussed above. In earlier work on CPL [LPSS12] we have proved the van Benthem/Rosen theorem for CPL by embedding CPL into this language; here, we give a native proof, the key additional step being a Gaifman locality result for CPL with explicit support.

2. Coalgebra and Modal Logic

We begin by recalling the basic definitions and examples of coalgebraic modal logic. We fix a modal similarity type Λ consisting of modal operators $\heartsuit$ with given finite arities (possibly 0).

The set $ML(\Lambda)$ of *modal Λ-formulas* is then given by the grammar

$$\phi, \psi ::= \bot \mid \phi \wedge \psi \mid \neg\phi \mid \heartsuit(\phi_1, \dots, \phi_n)$$

where $\heartsuit \in \Lambda$ is n-ary. We denote the infinitary language that admits conjunctions of arbitrary sets of formulas rather than just binary conjunctions by $ML_\infty(\Lambda)$. We write $\mathit{rank}(\phi)$ for the maximal nesting depth of modal operators in the formula ϕ, defined formally as $\mathit{rank}(\bot) = 0$, $\mathit{rank}(\bigwedge \Phi) = \sup_{\phi\in\Phi} \mathit{rank}(\phi)$, $\mathit{rank}(\neg\phi) = \mathit{rank}(\phi)$ and $\mathit{rank}(\heartsuit(\phi_1, \dots, \phi_n)) = 1 + \max\{\mathit{rank}(\phi_1), \dots, \mathit{rank}(\phi_n)\}$. Thus, the rank of a formula in $ML_\infty(\Lambda)$ is an element of $\mathbb{N} \cup \{\infty\}$, while the rank of a formula in $ML(\Lambda)$ is finite.

Remark 1. In the interest of generality, $ML(\Lambda)$ is not assumed to come equipped with a supply of propositional atoms. These can be emulated as nullary modalities in Λ. This allows covering more formalisms such as Hennessy-Milner Logic (which does not have propositional atoms), and simplifies the development of the generic theory.

We interpret modal Λ-formulas over *coalgebras*, that is, pairs (C, γ) consisting of a carrier set C and a transition function $\gamma : C \to TC$ where $T : \mathbf{Set} \to \mathbf{Set}$ is an endofunctor on the category $\mathbf{Set}$ of sets and maps. We refer to elements of TC as *structured successors*; they should be thought of as data structures organizing successor states, e.g. sets of successors or distributions over successors. The basic example is $T = \mathcal{P}$, the *(covariant) powerset functor* – in this case, a coalgebra $\gamma : C \to \mathcal{P}(C)$ is easily seen to be just a Kripke frame, i.e., a binary relation on C. To interpret the modal operators, we additionally fix, for each $\heartsuit \in \Lambda$, a *predicate lifting* $[\![\heartsuit]\!]$, i.e., a natural transformation

$$[\![\heartsuit]\!] : \mathcal{Q}^n \to \mathcal{Q} \circ T^{op}$$

where n is the arity of $\heartsuit$ and $\mathcal{Q} : \mathbf{Set}^{op} \to \mathbf{Set}$ denotes the contravariant powerset functor (i.e., $\mathcal{Q}(X)$ is the powerset of X, and for a map $f : Y \to X$, $\mathcal{Q}(f) : \mathcal{Q}(X) \to \mathcal{Q}(Y)$ takes preimages). That is, $[\![\heartsuit]\!]$ is a family of maps $[\![\heartsuit]\!]_X : \mathcal{Q}(X)^n \to \mathcal{Q}(T(X))$ subject to *naturality*, i.e.,

$$(Tf)^{-1}[[\![\heartsuit]\!]_X(A_1, \dots, A_n) = [\![\heartsuit]\!]_Y(f^{-1}[A_1], \dots, f^{-1}[A_n])$$

for all functions $f : Y \to X$ and all $A_1, \dots, A_n \subseteq X$. We call the collection of all these data a *structure*, which by abuse of notation we denote just as Λ. Given such a structure, we interpret modal Λ-formulas ϕ over coalgebras (C, γ) in terms of a satisfaction relation $c \models_{(C,\gamma)} \phi$ for $c \in C$, defined recursively by the obvious clauses for Boolean operators and

$$c \models_{(C,\gamma)} \heartsuit(\phi_1, \dots, \phi_n) \qquad \text{iff} \qquad \gamma(c) \in [\![\heartsuit]\!]_C([\![\phi_1]\!]_{(C,\gamma)}, \dots, [\![\phi_n]\!]_{(C,\gamma)})$$

where $[\![\phi]\!]_{(C,\gamma)} = \{d \mid d \models_{(C,\gamma)} \phi\}$.

Example 2. (1) Kripke models over a set P of propositional variables are triples (W, R, σ) where W is a set, $R \subseteq W \times W$ is a binary relation, and $\sigma : \mathsf{P} \to \mathcal{P}(W)$ is a valuation of propositional variables. It is easy to see that Kripke models are in 1-1 correspondence with T-coalgebras for $TX = \mathcal{P}(X) \times \mathcal{P}(\mathsf{P})$. The syntax of the modal logic K comes about via the similarity type $\Lambda = \{\Diamond\} \cup \mathsf{P}$ where $\Diamond$ is unary and each $p \in \mathsf{P}$ doubles as a nullary modality. The language $ML(\Lambda)$ is interpreted over T-coalgebras by virtue of the structure

$$[\![\Diamond]\!]_X(A) = \{(B, C) \in TX \mid B \cap A \neq \emptyset\} \qquad [\![p]\!]_X = \{(B, C) \in TX \mid p \in C\}.$$

Clearly this semantics coincides with the standard textbook semantics of K [BdRV01].

(2) The modal logic of neighbourhood frames (also called *classical modal logic* [Che80] or *neighbourhood logic*) arises via the same similarity type but is interpreted over neighbourhood models, i.e. coalgebras for the functor $TX = \mathcal{Q}(\mathcal{Q}(X)) \times \mathcal{P}(\mathsf{P})$ where again P is a set of propositional variables and $\mathcal{Q}$ denotes contravariant powerset. For a T-coalgebra (C, γ), we say that $A \subseteq C$ *is a neighbourhood of* $c \in C$ if $\gamma(c) = (N, B)$ where $A \in N$. The interpretation of propositional constants (nullary modalities) is as above and the semantics of $\Box$ is given by the predicate lifting

$$[\![\Box]\!]_X(A) = \{(N, B) \in TX \mid A \in N\},$$

which again gives rise to the standard semantics.

(3) *Monotone modal logic* has the same syntax as classical modal logic, but is interpreted over monotone neighbourhood models, i.e., coalgebras for the functor

$$TX = \{A \in \mathcal{Q}(\mathcal{Q}(X)) \mid A \text{ upwards closed}\} \times \mathcal{P}(\mathsf{P})$$

where upwards closure refers to subset inclusion. Thus, monotone modal logic validates the axiom $\Box(a \wedge b) \to \Box a$.

(4) *Conditional logic* has a binary modal operator $\Rightarrow$ that we write in infix notation. Conditional models over a set P of propositional variables come about as coalgebras for the functor

$$TX = \{f : \mathcal{Q}(X) \to \mathcal{P}(X) \mid \ f \text{ a function}\} \times \mathcal{P}(\mathsf{P})$$

(which defines a functor since the powerset on the left of the function space is contravariant) where propositional constants are interpreted as above and the lifting

$$[\![\Rightarrow]\!]_X(A, B) = \{(f, D) \in TX \mid f(A) \subseteq B\}$$

induces the selection function semantics of the conditional logic CK [Che80]. Thus, $\Rightarrow$ behaves like a Kripke modality $\psi \Rightarrow \_$ indexed over propositions ψ, but satisfies only replacement of equivalents (rather than, e.g., antimonotonicity) in the first argument.

(5) We obtain a variant of graded modal logic [Fin72] if we consider the similarity type $\Lambda = \{\langle k\rangle \mid k \geq 0\} \cup \mathsf{P}$ where $\langle k\rangle$ reads as 'more than k successors satisfy ...'. To retain naturality of predicate liftings, we slightly deviate from the traditional semantics [Fin72] and interpret graded modal logic over multigraphs [DV02] where the graded modalities refer to the weighted sum of successors. That is, we use coalgebras for $TX = \mathcal{B}X \times \mathcal{P}(\mathsf{P})$ where $\mathcal{B}$ is the multiset functor defined by $\mathcal{B}C = C \to \mathbb{N} \cup \{\infty\}$, whose elements we view as $\mathbb{N} \cup \{\infty\}$-valued measures μ. In particular, for $A \subseteq X$ we write $\mu(A) = \sum_{x\in A} \mu(x)$, and in this notation define $\mathcal{B}f$ on maps f as taking image measures, i.e. $\mathcal{B}f(\mu)(B) = \mu(f^{-1}[B])$. We extend T to a Λ-structure by stipulating

$$[\![\langle k\rangle]\!]_X(A) = \{(\mu, B) \in TX \mid \mu(A) > k\}$$

to express that more than k successors (counted with multiplicities) have property A. This semantics is equivalent to the standard Kripke semantics w.r.t. satisfiability of formulas, as multigraphs can be converted to Kripke frames by inserting the appropriate number of copies for each successor [Sch07]. All this easily generalizes to linear constraints between numbers of successors, i.e., Presburger modal logic [DL10], with n-ary modal operators $c_1\#(\cdot_1) + \dots c_n \cdot \#(\cdot_n) > c_0$ in arguments $(\cdot_i)$ for $1 \leq i \leq n$, and coefficients $c_0, \dots, c_n \in \mathbb{Z}$.

A variation of graded modal logic arises by limiting the overall weight of successor states. If we consider the subfunctor

$$\mathcal{B}_k X = \{\mu \in \mathcal{B}X \mid \mu(X) \leq k\}$$

for some $k \geq 0$, we may describe k-bounded multigraphs and interpret the sublanguage that only features the modalities $\langle i\rangle$ for $i < k$.

(6) Frame classes can be combined in various ways by standard constructions on functors [CP07, SP11]; e.g. the *fusion* of two logics, which effectively just forms the disjoint union of modal similarity types and axiomatizations, is modelled by taking the product of the associated functors. For example, coalgebras $(C, \gamma : C \to TC)$ where $TC := \mathcal{B}C \times \mathcal{P}C \times \mathcal{P}(\mathsf{P})$ combine a multigraph model and a relational model, and we can interpret the fusion of K and graded modal logic on T-coalgebras by projecting out the respective components.

(7) For probabilistic logics, we prefer to work with subprobabilities for technical reasons, where a (discrete) subprobability distribution on a set X is a discrete measure P on X such that $P(X) \leq 1$. (By a discrete measure we mean one that is defined on the whole powerset, and hence concentrated on a countable set.) The similarity type of the *modal logic of subprobabilities*, a variant of probabilistic modal logic [HM01], contains, apart from propositional variables, the modal operators L_p for rational $p \in [0, 1] \cap \mathbb{Q}$. This language is interpreted over T-coalgebras where $TX = \mathcal{S}(X) \times \mathcal{P}(\mathsf{P})$ and $\mathcal{S}(X)$ is the set of subprobability distributions on X. Up to the use of subprobability distributions instead of probability distributions, a T-coalgebra is, in the terminology used in epistemic views of probabilistic logic, a discrete type space [HM01], and its first component $X \to \mathcal{S}(X)$

a Markov chain. The modalities L_p, read as "in the next step, it holds with probability at least p that", are interpreted via the liftings $[\![L_p]\!]_X(A) = \{(\mu, B) \in TX \mid \mu(A) \geq p\}$, which induces, up to the move to subprobabilities, the standard semantics.

For the sake of readability, we restrict the generic parts of the technical exposition to unary operators from now on (but continue to use conditional logic as an example); everything extends straightforwardly to higher (finite) arities by just writing more indices.

Assumption 3. We assume w.l.o.g. that T preserves injective maps [Bar93]. For ease of notation, we will in fact sometimes assume T is *standard*, i.e., maps subset inclusions $X \hookrightarrow Y$ to subset inclusions $TX \hookrightarrow TY$. Moreover, we assume w.l.o.g. that T is *non-trivial*, i.e., $TX = \emptyset \implies X = \emptyset$ (otherwise, $TX = \emptyset$ for all X).

Using the above assumption, we can give a simple definition of *support*, which will play a role in our proofs:

Definition 4. A set $A \subseteq C$ is a *support* of $t \in TC$ if $t \in TA \subseteq TC$. A *supporting Kripke frame* of a T-coalgebra $\gamma : C \to TC$ is a binary relation R on C such that for each $c \in C$, $R(c) = \{d \in C \mid cRd\}$ is a support of $\gamma(c)$.

Support has played a role in various coalgebraic model constructions, see, e.g., [SP09]. We keep the notion of support as broad as possible, and in particular do not insist on minimality, as the set of supports of $t \in TC$ in general need not have a smallest element with respect to subset inclusion [Gum05].

Example 5.

- For any coalgebra (C, γ), the total relation $R = C \times C$ is a supporting Kripke frame.
- Any Kripke frame is its own smallest supporting Kripke frame. Clearly, if R is a supporting Kripke frame of any coalgebra and $R \subseteq R'$, then R' is also a supporting Kripke frame. Thus, the supporting Kripke frames for a Kripke frame $S \subseteq C \times C$ are precisely the Kripke frames between S and $C \times C$.
- The smallest supporting Kripke frame of an $\mathcal{S}$-coalgebra $\gamma : C \to \mathcal{S}C$ (with $\mathcal{S}$ the subdistribution functor, see Example 2.7) is $\{(c, d) \in C \times C \mid \gamma(c)(d) > 0\}$.
- As an example of a structure where coalgebras do not in general have smallest supporting Kripke frames, consider neighbourhood logic (Example 2.2): given $T = \mathcal{Q} \circ \mathcal{Q}$, a relation $R \subseteq C \times C$ is a supporting Kripke frame of a T-coalgebra $\gamma : C \to TC$ iff for all $c \in C$ and all $A \subseteq C$, $A \in \gamma(c)$ iff $A \cap R(c) \in \gamma(c)$.

Behavioural equivalence. In the relational setting, the expressivity of modal logic is characterized by *bisimilarity*. While generic coalgebraic notions of bisimulation work well in many cases [Rut00, Sta11, GS13], to obtain full generality one needs to work with *behavioural equivalence*, a notion best described in terms of coalgebra morphisms. A *morphism* between T-coalgebras (C, γ) and (D, δ) is a function $f : C \to D$ such that $\delta \circ f = Tf \circ \gamma$. Given T-coalgebras (C, γ) and (D, δ), two states $(c, d) \in C \times D$ are called *behaviourally equivalent*, written $C, c \approx D, d$, if they can be identified by morphisms of T-coalgebras, i.e., there are morphisms $f : (C, \gamma) \to (E, \epsilon)$ and $g : (D, \delta) \to (E, \epsilon)$ into a T-coalgebra (E, ϵ) such that $f(c) = g(d)$.

Remark 6. Given that we will also be interested in behavioural equivalence on *finite* coalgebras, one may wonder if in this case it makes a difference whether the coalgebra E in the above definition is also required to be finite. The answer is no, since morphic images of coalgebras are subcoalgebras and unions of subcoalgebras are subcoalgebras, so that if C and D as above are finite and map into E, then they map into a finite subcoalgebra of E.

A property P of states is *invariant* under behavioural equivalence, in short *behavioural-equivalence invariant* if whenever a state x has property P and y is behaviourally equivalent to x then y has property P.

Lemma 7. *Satisfiablity of formulas from $ML(\Lambda)$ is behavioural-equivalence invariant.*

The simple proof [Pat03] is by induction, with naturality of predicate liftings used in the modal step. A still shorter formulation is that modal Λ-formulas are behavioural-equivalence invariant.

Our main Theorem 34 below extends the result of van Benthem [vB76] to a coalgebraic setting, establishing that all behavioural-equivalence invariant first order formulas in a suitable correspondence language defined in Section 3 are in fact equivalent to (possibly infinitary) modal formulas. The proof follows Rosen [Ros97] and Otto [Ott06], and in particular makes use of the stratification of behavioural equivalence that explicitly accounts for the number of transition steps. From a coalgebraic perspective, this comes about by considering the projections of (states of) coalgebras into the so-called *terminal sequence* of the underlying functor (see [Pat04] for a detailed exposition from the logical viewpoint). The terminal sequence consists of the objects $T^n 1$ where $1 = \{*\}$ is a one-element set and T^n denotes the n-fold iteration of T (in particular, $T^0 1 = 1$), connected by maps $p_n : T^{n+1}1 \to T^n 1$ defined by taking p_0 to be the unique map $T1 \to 1$ and putting $p_{n+1} = Tp_n$. Every T-coalgebra (C, γ) defines a cone over the terminal sequence by $\gamma_0 : C \to 1$ and $\gamma_{n+1} = T\gamma_n \circ \gamma : C \to T^{n+1}1$:

$$\begin{tikzcd}
C \arrow[r, "\gamma"] \arrow[d, "\gamma_n"'] \arrow[dr, "\gamma_{n+1}"] & TC \arrow[d, "T\gamma_n"] \\
T^n 1 & T^{n+1}1. \arrow[l, "p_n = Tp_{n-1}"]
\end{tikzcd}$$

Given two T-coalgebras C and D, we now call a pair $(c, d) \in C \times D$ *n-step equivalent*, in symbols $C, c \approx_n D, d$, if $\gamma_n(c) = \delta_n(d)$. The following lemma relates n-step equivalence and behavioural equivalence:

Lemma 8. *Let C, D be T-coalgebras, and let $n \geq k \in \omega$. For $(c, d) \in C \times D$, $c \approx d$ implies that $c \approx_n d$, and $c \approx_n d$ implies that $c \approx_k d$.*

Separation. We recall the notion of separation, which has been used as a sufficient (and, under mild additional assumptions, necessary) criterion to establish the Hennessy-Milner property for coalgebraic modal logics [Pat04, Sch08]. We require a few notational items:

Definition 9. Given a set Z, we denote by $\Lambda(Z)$ the set of formal expressions $\heartsuit z$ for $\heartsuit \in \Lambda$, $z \in Z$. A *one-step formula* over X is a Boolean expression over $\Lambda(\mathcal{P}(X))$. As a special case, a *one-step conjunctive clause* is a conjunctive clause over $\Lambda(\mathcal{P}(X))$, where a conjunctive clause is a conjunction of literals; like in the case of modal Λ-formulas, we sometimes consider infinitary conjunctions, explicitly designated as such. A one-step formula ϕ over X is interpreted as a subset $\llbracket \phi \rrbracket$ of TX by extending the assignment $\llbracket \heartsuit A \rrbracket = \llbracket \heartsuit \rrbracket_X(A)$ using the Boolean algebra structure of $\mathcal{P}(TX)$. For $t \in TX$, we write $t \models \phi$ if $t \in \llbracket \phi \rrbracket$; in particular, $t \models \heartsuit A$ if $t \in \llbracket \heartsuit \rrbracket_X(A)$.

Definition 10. We say that Λ is *separating* if for all sets X, every $t \in TX$ is uniquely determined by the set $\{\heartsuit A \in \Lambda(\mathcal{P}(X)) \mid t \models \heartsuit A\}$.

Example 11. All structures mentioned so far are well-known (and easily seen) to be separating. In fact, non-separating structures are not particularly common; one well-known example of a non-separating structure is coalition logic [SP09]. As a simple example on how to establish separation, consider the structure for the basic modal logic K (Example 2.1). We have to show that $t \in \mathcal{P}(X)$ is uniquely determined by $\{\Box A \mid t \models \Box A\}$. But $t \models \Box A$ iff $t \subseteq A$ and it is clearly true that every subset t of X is uniquely determined by the set of its supersets.

Separation allows us to construct modal definitions for properties that are invariant under n-step equivalence:

Definition 12. A property P of states is *$\approx_n$-invariant* if whenever a state x has property P and $x \approx_n y$ then y has property P.

Lemma 13. *If Λ is separating then every $\approx_n$-invariant property is definable by an infinitary modal Λ-formula of modal rank $\leq n$.*

Proof. Induction on n. For $n = 0$, all states are equivalent under $\approx_0$, so a $\approx_0$-invariant property holds either for all states or for no state and is hence defined by either $\top$ or $\bot$. For the step from n to $n + 1$, it suffices to show that we can define all $\approx_{n+1}$-equivalence classes by infinitary modal Λ-formulas of rank at most $n + 1$; the claim then follows, as we obtain definitions of all $\approx_{n+1}$-invariant properties using (possibly infinitary) disjunctions, preserving the bound $n + 1$ on the rank. Thus, let c be a state in a T-coalgebra (C,γ). By separation, there is a (possibly infinitary) one-step conjunctive clause ϕ over $\Lambda(\mathcal{P}(T^n 1))$ uniquely describing $\gamma_{n+1}(c) = T\gamma_n(\gamma(c))$ as an element of $T^{n+1}1 = T(T^n 1)$. We then obtain a (possibly infinitary) modal Λ-formula ψ defining the $\approx_{n+1}$-equivalence class of c and having the claimed modal rank by replacing every set $A \in \mathcal{P}(T^n 1)$ occurring in ϕ with the formula ψ_A defining the class of all states d in T-coalgebras (D,δ) such that $\delta_n(d) \in A$ – this class is clearly closed under n-step equivalence and hence, by induction, definable by a possibly infinitary modal Λ-formula of rank at most n. Now note that by naturality of predicate liftings, for each coalgebra (D,δ), the conjunctive clause $\bar{\phi}$ over $\Lambda(\mathcal{P}(D))$ obtained from ϕ by replacing each $A \in \mathcal{P}(T^n 1)$ with $\gamma_n^{-1}[A]$ defines the set $\{t \in TD \mid T\gamma_n(t) \models \phi\} = \{t \in TD \mid T\gamma_n(t) = \gamma_{n+1}(c)\}$. Thus, ψ defines, on (D,δ), the set of all $d \in D$ such that $T\delta_n(\delta(d)) = \gamma_{n+1}(c)$; since $T\delta_n\delta = \delta_{n+1}$, this means that ψ defines the $\approx_{n+1}$-equivalence class of c. □

3. Coalgebraic Predicate Logic

We next recall the definition of *coalgebraic predicate logic (CPL)*, introduced in earlier work [LPSS12, LPS13] as a semantic generalization of Chang's modal first-order language [Cha73]. Formulas of CPL over a modal similarity type Λ and a first-order predicate signature Σ are given by the grammar

$$\mathsf{CPL}(\Lambda,\Sigma) \ni \phi,\psi ::= \quad y_1 = y_2 \mid P(\vec{x}) \mid \bot \mid \phi \to \psi \mid \forall x.\,\phi \mid x\heartsuit\lceil y_1 : \phi_1\rceil \dots \lceil y_n : \phi_n\rceil$$

where $\heartsuit \in \Lambda$ is an n-ary modal operator, $P \in \Sigma$ a k-ary predicate symbol, and x, y_i are individual variables from a fixed set iVar, which we keep implicit. Further Boolean connectives and the existential quantifier are defined in the standard way. We do not include function symbols, which can be added at no extra cost [Cha73]. The *quantifier rank* $\mathsf{qr}(\phi)$ of a formula ϕ is the maximal nesting depth of binders in ϕ, where a binder is either a quantifier or a modality.

Remark 14. We will often consider languages with $\Sigma = \emptyset$, i.e., without predicate symbols (so that formulas are built from Λ alone); we refer to CPL without predicate symbols as *pure CPL*, and omit mention of Σ in pure CPL. Indeed the actual correspondence language for coalgebraic modal logic is pure CPL: we will see below that the standard translation of coalgebraic modal logic into CPL does not require predicate symbols. We continue to hide propositional atoms within the *modal* similarity type: for an atomic proposition p, which would be modelled as a unary predicate in the standard correspondence language, CPL has the *modal* formula xp, with p a nullary modality. We include predicate symbols on the one hand to clarify the relation to Chang's original language, and on the other hand since for technical reasons, we will later need to extend pure CPL with a support predicate anyway.

In the i-th component $\lceil y_i : \phi_i\rceil$ of a modal formula $x\heartsuit\lceil y_1 : \phi_1\rceil \dots \lceil y_n : \phi_n\rceil$, y_i is used as a comprehension variable, i.e., $\lceil y_i : \phi_i\rceil$ denotes a subset of the carrier of the model, to which modal operators can be applied in the usual way. In $x\heartsuit\lceil y_1 : \phi_1\rceil \dots \lceil y_n : \phi_n\rceil$, x is free and y_i is bound in ϕ_i; otherwise the notions of free and bound variables are standard.

In the same way as for coalgebraic modal logic, the semantics of CPL is parametrized in terms of a structure Λ, i.e., a choice of a coalgebraic type functor T and an assignment of predicate liftings $[\![\heartsuit]\!]$ to the modalities $\heartsuit \in \Lambda$. A pair $\mathfrak{M} = (C,\gamma,I)$ consisting of a coalgebra $\gamma : C \to TC$ and a predicate interpretation $I : \Sigma \to \bigcup_{n\in\omega} \mathcal{P}(C^n)$ respecting arities of symbols will be called a *(coalgebraic) model*. In other words, a coalgebraic model consists simply of a $\mathbf{Set}$-coalgebra and an ordinary first-order model whose universe coincides with the carrier of the coalgebra. Given a model $\mathfrak{M} = (C,\gamma,I)$ and a valuation $v : \mathsf{iVar} \to C$, we define satisfaction $\mathfrak{M}, v \models \phi$ in the standard way for first-order connectives and for $\heartsuit$ by the clause

$$\mathfrak{M}, v \models x\heartsuit\lceil y_1 : \phi_1\rceil \dots \lceil y_n : \phi_n\rceil \iff \gamma(v(x)) \models \heartsuit([\![\phi_1]\!]^{y_1}_{\mathfrak{M},v}, \dots, [\![\phi_n]\!]^{y_n}_{\mathfrak{M},v})$$

where $[\![\phi]\!]^{y}_{\mathfrak{M},v} := \{c \in C \mid \mathfrak{M}, v[c/y] \models \phi\}$ and $v[c/y]$ is v modified by mapping y to c. We use the standard shorthand $\mathfrak{M} \models \phi(\vec{c})$ for $\mathfrak{M}, v \models \phi$ where $\vec{c}$ is a vector of values for the free variables of ϕ and v maps the free variables of ϕ to these values.

We discuss a few examples in this setting [LPSS12].

Example 15. (1) *Relational first-order logic.* Starting from the structure for the basic modal logic K (Example 2.1), we obtain an instance of CPL that is expressively equivalent to the usual first-order correspondence language $\mathcal{CL}$ for modal logic, i.e., the first order language with one binary predicate R and a unary predicate p for each propositional atom. Specifically, we map $\mathcal{CL}$ into CPL by translating atoms $R(x, y)$ to $x\Diamond\lceil z : z = y\rceil$ and $p(x)$ to $x\,p$ (recall that propositional atoms are nullary modalities). Conversely, we map CPL into $\mathcal{CL}$ by translating $x\Diamond\lceil y : \phi\rceil$ to $\exists y.\,(R(x, y) \wedge \phi)$ and xp to $p(x)$.

(2) *Social Situations and Neighbourhood Frames.* The instance of CPL for the structure defining neighbourhood logic (Example 2.2) coincides (up to minor syntactic modifications [LPSS12]) exactly with Chang's original modal first-order language [Cha73]. As noted by Chang himself, the resulting language is particularly well-tailored for reasoning about social situations and relationships between an individual and sets of individuals. In the presence of a binary relation $S(x, y)$ that we read as 'x speaks to y', and interpreting $\Box$ as 'enjoyable', we read the formula $\exists y_1.\,\exists y_2.\,(x\Box\lceil z : S(z, y_1)\rceil \wedge x\Box\lceil z : S(z, y_2)\rceil \wedge y_1 \neq y_2)$ as 'there are at least two people such that x finds it enjoyable to speak to them' where x determines the truth of this sentence by inspecting the set $\{z : S(z, y_i)\}$ of people speaking to y_i.

(3) *Party Invitations and Non-Monotonic Conditionals.* In the spirit of Chang's original examples concerning social situations, we may read the antecedent of the conditional as 'invites' and the consequent as 'makes happy', noting that selection function semantics can be read as determining a proposition-indexed family of transition relations, where the index is determined as the extension of the antecedent of a conditional, and $\phi \Rightarrow \_$ is then just the box modality for the transition relation indexed by ϕ. We denote the right-hand dual of $\Rightarrow$ as $>$, i.e. $\phi > \psi \equiv \neg(\phi \Rightarrow \neg\psi)$ so that $\phi > \_$ is the diamond modality for the transition relation indexed by ϕ. Given a binary relation ff ('facebook friend') the formula $\exists y_1, y_2.\,(y_1 \neq y_2 \wedge x(\lceil z : \mathsf{ff}(x, z)\rceil > \lceil z : z = y_1\rceil) \wedge x(\lceil z : \mathsf{ff}(x, z)\rceil > \lceil z : z = y_2\rceil))$ states that there are at least two persons y_1, y_2 – possibly Mark Zuckerberg and Sheryl Sandberg – who are made happy by x if x invites *precisely* her facebook friends to her birthday party. (The formula makes no statement about y_1 and y_2's emotional state if the set of invitees differs in any way from the set of facebook friends, i.e. if either some of the facebook friends are not invited or some of the invitees are not among the facebook friends.)

(4) *Facebook Friends and Graded Modal Logic.* Given a $\mathcal{B}$-coalgebra $(C, \gamma : C \to \mathcal{B}C)$, we can think of elements of C as individuals, and of $\gamma(c)(d)$ as the number of 'likes' (in the sense of facebook) that d has received from c. In other words, $\gamma(c)(d) = n$ models the fact that c has pressed the 'like'-button on d's page n times. In the presence of the above binary relation $\mathsf{ff}(x, y)$ expressing that y is a facebook-friend of x, the formula $x\langle k\rangle\lceil z : \exists y.\,\mathsf{ff}(x, y) \wedge \mathsf{ff}(y, z)\rceil$ expresses that x likes more than k activities of friends of her friends.

On a more sober note, CPL with graded modalities is similar to first order logic with counting quantifiers $\exists_{\geq k}$ but has some semantic differences due to the weighting of successors enabled by the multigraph semantics. In a nutshell, $x\langle k\rangle\lceil y : \phi\rceil$ will behave like $\exists_{\geq k} y.\,\phi$ except that it will not just count the number of domain elements y satisfying ϕ but it will count each y with the multiplicity assigned to it locally at x. We can in fact embed standard first order logic with counting quantifiers into CPL by disabling non-trivial multiplicities: we just need to add to the indicated translation the axiom $\forall x.\,\forall y.\,(x\langle 0\rangle\lceil z : z = y\rceil \wedge \neg x\langle 1\rangle\lceil z : z = y\rceil)$ stating that the transition multiplicity between any two states in the model is 1.

(5) *Combination of Frame Classes.* We can take $T = \mathcal{B} \times \mathcal{QQ}$ and combine operators for the facebook sense of 'like' and Chang's modalities for social situations. A formula $\neg x\Box\lceil y : y\langle 2\rangle\lceil y : y = z\rceil\rceil$ then expresses that x does not fancy the perspective of liking strictly more than 2 of the facebook activities of z (or, to be more precise, the general company of people who do so). The reader may find it entertaining to compare our facebook examples with those of [SLG11].

(6) *Probabilistic first-order logic* CPL over probabilistic modalities L_p is similar to Halpern's *type-1 probabilistic first-order logic (PFOL)* [Hal90], up to a semantic difference that lies in the fact that the former works with local distributions at each state, and the latter with a single global distribution. Specifically, type-1 PFOL works over structures consisting of a standard first order model and a probability distribution over the domain of the model, while CPL instead works with a family of distributions over the domain, one for each state in the model (i.e. with a Markov

chain on the domain). Additionally, type-1 PFOL has a more expressive syntax which allows using arbitrary formulas in first-order arithmetic over probabilities of open formulas, and moreover caters for probabilities taken over several free variables. Sticking to the case with one free variable, the term $w_y(\phi(y))$ designates, in type-1 PFOL, the probability of a random element y of the domain to satisfy $\phi(y)$, and an atomic formula $w_y(\phi(y)) \geq p$ corresponds roughly to the CPL formula $xL_p\lceil y : \phi\rceil$ except that in the CPL formula, the probability distribution depends on x. Unlike in the case of graded logic, it does not seem possible to force the probability distribution to be independent of the state by means of a CPL formula. (We remark that the generalization to subprobabilites, on the other hand, is easily axiomatized away by means of the CPL formula $\forall x.\, x\, L_1\lceil y : \top\rceil$.)

Since CPL is itself based on modalities, the translation of modal formulas to CPL is simpler than in the standard setup. We can inductively define the *standard translation* $ST_x(\phi)$ of a modal Λ-formula ϕ as a CPL formula with one free variable x by commutation with all Boolean operators and

$$ST_x(\heartsuit\phi) = x\heartsuit\lceil x : ST_x(\phi)\rceil.$$

Note that we reuse the free variable x as the bound variable in $\lceil x : ST_x(\phi)\rceil$, so that the translation ends up in the single-variable fragment of CPL.

Remark 16. The above definition may create the impression that CPL is more closely related to modal logic than standard first-order logic, hence making the characterization of modal logic as a fragment of CPL a weaker result than in the standard case. However, we have seen that CPL is expressively equivalent to first-order logic in the standard relational setup (Example 15.1) so that in this case, our characterization result does instantiate to yield exactly the original van Benthem and Rosen theorems (Example 44.3).

The standard translation is clearly correct in the following sense.

Definition 17. We say that a (finitary or infinitary) modal Λ-formula ϕ is *equivalent* to a CPL formula $\psi(x)$ over Λ in one variable x if for every coalgebraic model $\mathfrak{M} = (C, \gamma, I)$ and every state $c \in C$, $c \models_{(C,\gamma)} \phi$ iff $\mathfrak{M} \models \psi(c)$.

Lemma 18. *Every finitary modal Λ-formula ϕ is equivalent to its standard translation $ST_x(\phi)$.*

The same result also applies to infinitary formulae provided first-order logic is extended to allow infinitary conjunctions. Moreover, as noted in earlier work [LPS13] we also have

Lemma 19. *The subset of $\phi \in CPL(\Lambda, \emptyset)$ obtained as the image of ST_x for a fixed x consists precisely of the (equality-free) and quantifier-free formulas using only the variable x.*

(We put 'equality-free' in brackets because in the absence of function symbols and with only one variable x available, the only equation in the syntax is $x = x$.) Lemma 19 will allow us to derive a CPL-internal characterization of behaviourally invariant formulas in Corollary 38 below.

Remark 20. For the neighbourhood frame structure (Example 2.2), Hansen, Kupke and Pacuit [HKP09] define a first order correspondence language as follows. The language has two sorts s (*states*), n (*neighbourhoods*). Its predicate symbols are unary predicates P_a representing propositional atoms a, a predicate N where xNu states that u is a neighbourhood of x, and an elementhood predicate, which we write $\in$ for consistency with notation that we introduce later. A neighbourhood model C translates into a first-order model C^o for this signature by interpreting s as C, n as the set of all subsets of C that are neighbourhoods of some state in C, N as the neighbourhood relation, $\in$ as elementhood, and each P_a as the set of states satisfying a. The class N of all first-order models that arise in this way is easily seen to be first-order definable (by right totality of N and extensionality of neighbourhoods w.r.t. $\in$); moreover, C is clearly recoverable from C^o.

We use this notation to lend precision to the statement we made in the introduction, to the effect that the neighbourhood instance of CPL embeds into this two-sorted language. Specifically, we have a translation t that is defined recursively by commutation with Boolean operators and quantifiers, and

$$t(x\Box\lceil z : \phi\rceil) = \exists u : n.\, (xNu \wedge \forall z : s.\, (z \in u \leftrightarrow \phi)).$$

Postcomposing the standard translation into CPL as defined above with t yields the standard translation defined by Hansen et al. It is now easy to see (formally, with largely the same proof as given by Hansen et al. to show correctness of their standard translation) that for a neighbourhood model C and a CPL formula ϕ,

$$C \models \phi \quad \text{iff} \quad C^o \models t(\phi). \tag{1}$$

Hence, the van Benthem version of our characterization theorem (for unrestricted models) is implied by the corresponding theorem proved by Hansen et al.: If a CPL-formula ϕ is invariant under behavioural equivalence, then so is $t(\phi)$, by (1). Therefore $t(\phi)$ is, by the van Benthem Theorem of Hansen et al., equivalent to the standard translation of a modal formula; again by (1), and because t commutes with the standard translations, it follows that ϕ is equivalent to the standard translation of a modal formula.

It does not, on the other hand, seem to be the case that our *generic* characterization theorem follows from the result of Hansen et al. (suitably generalized to allow for multiple modalities of finite arity). The way to go about deriving a generic result from one concerning neighbourhood frames would be to note that a separating set of predicate liftings for T induces an embedding of T into a product N of n-ary neighbourhood functors $N_n(X) = \mathcal{Q}(\mathcal{Q}(X^n))$ [SP10b], so that every T-coalgebra can be seen as an N-coalgebra. One then observes that states in T-coalgebras are behaviourally equivalent iff they are behaviourally equivalent when seen as states in N-coalgebras. However, this latter fact does not appear to imply that CPL-formulas that are behavioural-equivalence invariant over T are also behavioural-equivalence invariant over N (as a state in a T-coalgebra may be behaviourally equivalent to one in an N-coalgebra that does not come from a T-coalgebra), so that the characterization of the behavioural-equivalence invariant fragment does not easily transfer from N back to T.

4. Support and Gaifman Locality

Following Otto [Ott06], we base our characterization of the bisimulation-invariant fragment of coalgebraic predicate logic on the fact that the truth of bisimulation-invariant properties at a world only depends on a local neighbourhood of this point. In (standard) first-order logic, this is expressed in terms of *Gaifman distance*. The Gaifman distance of two points in a model is the length of the shortest path that connects these two points along the interpretation of relation symbols. To define Gaifman distances in an extension of first-order logic with general (not necessarily relational) modal operators, *support* (Definition 4) plays a crucial role. To illustrate the problem, consider for a moment the attempt to define a Gaifman graph R on a T-coalgebra $\gamma : C \to TC$ by putting cRd (and dRc) whenever $\gamma(c) \models \heartsuit A$ and $d \in A$ for some $\heartsuit \in \Lambda$, $A \subseteq C$. In most situations, e.g. as soon as $\heartsuit\top$ is valid for some $\heartsuit \in \Lambda$, this will lead to a Gaifman distance of 1 between any two states, which will clearly render Gaifman locality useless. We can work around this problem by introducing a dedicated support relation into the language, as follows.

We define *support-CPL* to be CPL with a distinguished binary predicate supp, which in each model is required to be interpreted by a supporting Kripke frame of the underlying T-coalgebra; such models are called *supported coalgebraic models*. Naturality of predicate liftings then implies that for each formula ϕ,

$$x\heartsuit\lceil y : \phi\rceil \leftrightarrow x\heartsuit\lceil y : \phi \wedge x \;\mathsf{supp}\; y\rceil \tag{2}$$

is valid in support-CPL. *Pure support-CPL* has the support relation as the only predicate. We make the following assumption to simplify the presentation w.r.t. induced submodels.

Assumption 21. We assume that $T\emptyset \neq \emptyset$. We distinguish an element $\bot_T \in T\emptyset$, and then make T into a pointed functor by equipping every set TC with a distinguished element $Ti(\bot_T)$, also denoted $\bot_T$, where $i : \emptyset \to C$.

Remark 22. All our running examples have $T\emptyset \neq \emptyset$. This was in fact the reason to consider subdistributions instead of distributions in our version of probabilistic logic – there is a unique trivial subdistribution on the empty set, which however is not a distribution.

We note explicitly that T really is pointed, i.e., $\bot_T$ is preserved by all maps Tf:

Lemma 23. *For every map f, $Tf(\bot_T) = \bot_T$.*

Proof. Let $f : X \to Y$; then we have to show $TfTi_X(\bot_T) = Ti_Y(\bot_T)$ where $i_X : \emptyset \hookrightarrow X$ and $i_Y : \emptyset \hookrightarrow Y$. But this is immediate from $fi_X = i_Y$ and functoriality of T. □

Definition 24. Let $\mathfrak{M} = (C, \gamma, I)$ be a supported coalgebraic model, and let $A \subseteq C$. The *submodel* $\mathfrak{M}|_A = (A, \gamma|_A, I|_A)$ *induced* by A is defined (exploiting in the notation that T is standard, cf. Assumption 3) by taking $I|_A(P) = I(P) \cap A^n$ for n-ary P, and

$$\gamma|_A(c) = \begin{cases} \gamma(c) & \text{if } \mathsf{supp}(c) \subseteq A \\ \bot_T & \text{otherwise.} \end{cases}$$

From now on, only two languages will be of interest: pure CPL and pure support-CPL. In particular, our main result – the coalgebraic van Benthem/Rosen theorem – will be stated only for pure CPL, as pure CPL is the appropriate correspondence language for coalgebraic modal logic.

In pure support-CPL, we take the Gaifman graph to be determined by supp alone; formally:

Definition 25. The *Gaifman graph* of a supported coalgebraic model is the undirected graph induced by the interpretation of supp. The *Gaifman distance* between two states in a supported coalgebraic model is the graph distance in the Gaifman graph. For $l \geq 0$, the *l-neighbourhood* of a state c in a model $\mathfrak{M}$ is the (pointed) submodel $\mathsf{N}_l^{\mathfrak{M}}(c)$ induced by the set of all states with Gaifman distance at most l from c. A formula $\phi(x)$ in pure support-CPL with a single free variable x is *l-local* if for every supported coalgebraic model $\mathfrak{M} = (C, \gamma, I)$ and every $c \in C$, $\mathfrak{M}, c \models \phi(x)$ iff $\mathsf{N}_l^{\mathfrak{M}}(c), c \models \phi(x)$. Moreover, $\phi(x)$ is *Gaifman l-local* if for any two states $c, d \in C$ with isomorphic l-neighbourhoods (i.e., the neighbourhoods are isomorphic as pointed T-coalgebras and relations are preserved), $\mathfrak{M}, c \models \phi(x)$ iff $\mathfrak{M}, d \models \phi(x)$.

The main difference between a formula being l-local or Gaifman l-local is that the former property is expressed in terms of two different models (one of them being an induced submodel) whereas the latter is internal to a single model.

Remark 26. The above definition suggests another attempt to define a Gaifman distance for pure CPL (without support), namely as the supremum over Gaifman distances taken over all extensions of a coalgebraic model to a supported coalgebraic model. However, it is easy to construct examples (e.g. using topological frames) where this supremum is infinite everywhere, so that all l-neighbourhoods are singletons; for such a notion of distance, formulas will clearly not in general be Gaifman l-local for any finite l.

Our goal is now to show that the standard Gaifman theorem extends to pure support-CPL; i.e.,

Theorem 27 (Gaifman theorem for pure support-CPL)**.** *Every formula $\phi(x)$ in pure support-CPL with a single free variable is Gaifman l-local for some $l \geq 0$, exponentially bounded in the quantifier rank of ϕ.*

The strategy we pursue to prove this theorem is to inherit the statement from a standard first-order language into which we suitably translate pure support-CPL. Recall that the standard Gaifman theorem [Gai82, Lib04] makes exactly the same statement as above for standard first-order formulas with a single free variable, where the Gaifman graph of a first-order model connects two points if they occur together in some tuple in the interpretation of a predicate. The usual formulation in single-sorted logic readily extends to multiple sorts, using the standard encoding of multiple sorts as unary predicates. Note that the latter does not affect Gaifman distance as all the newly introduced predicates are unary and hence do not induce additional edges in the Gaifman graph.

The language we use is an extension of the correspondence language for neighbourhood models introduced by Hansen et al. [HKP09], the additional feature being, again, the support relation:

Definition 28. The *(coalgebraic) Henkin signature* $\mathcal{H}(\Lambda)$ associated with the modal similarity type Λ consists of two sorts s, n of *states* and *(modal) neighbourhoods*, respectively, together with the sorted relation symbols

- $\heartsuit \subseteq s \times n$ for all $\heartsuit \in \Lambda$ (the modal operators)
- $\in \subseteq s \times n$ (membership of points in neighbourhoods)
- $\mathsf{supp} \subseteq s \times s$ (the support relation).

We translate pure support-CPL formulas ψ over Λ into first-order formulas $t(\psi)$ over $\mathcal{H}(\Lambda)$, defined recursively by

$$t(x\heartsuit\lceil y : \phi\rceil) = \exists u : n. \, (x\heartsuit u \wedge \forall y. \, (y \in u \leftrightarrow (t(\phi) \wedge x \, \mathsf{supp} \, y)))$$

and commutation with all other syntactic constructs.

By (2), the above definition of $t(x\heartsuit\lceil y:\phi\rceil)$ captures the semantics of $x\heartsuit\lceil y:\phi\rceil$. Notice that $t(\psi)$ has the same free variables as ψ, in particular has no free variables of sort n. Moreover, the quantifier rank of $t(\psi)$ is at most twice that of ψ, as the latter counts modalities as quantifiers.

The motivation for the term Henkin signature is that, in remote analogy to Henkin models of higher-order logic, the sort n of neighbourhoods can be interpreted by *any* set; we do not require that this set contains all sets of states, or even satisfies extensionality w.r.t. $\in$. In fact, we will explicitly use non-extensional models.

A supported coalgebraic model $\mathfrak{M}=(C,\gamma,I)$ induces an $\mathcal{H}(\Lambda)$-structure $\mathcal{H}(\mathfrak{M})$ as follows. We interpret s by C, and n by the set

$$\{(c,B)\mid B\subseteq\mathsf{supp}(c)\}.$$

Here, the restriction to $B\subseteq\mathsf{supp}(c)$ serves to circumvent the above-mentioned problems with short Gaifman distances; technically, it is needed to validate the neighbourhood compatibility lemma proved below (Lemma 30). Then, $\in$ is interpreted by elementhood, i.e., by the relation $\{(c,(d,B))\mid c\in B\}$ (in particular, $\mathcal{H}(\mathfrak{M})$ is non-extensional), and the interpretation of supp is kept. Finally, we interpret $\heartsuit\in\Lambda$ in $\mathcal{H}(\mathfrak{M})$ by

$$\{(c,(c,B))\mid\gamma(c)\models\heartsuit B\}.$$

With this notion of model conversion, the translation of pure support-CPL into $\mathcal{H}(\Lambda)$ preserves satisfaction:

Lemma 29. *For every formula ϕ in support-CPL and every valuation v in a supported coalgebraic model $\mathfrak{M}$,*

$$\mathfrak{M},v\models\phi(\vec{c})\qquad\textit{iff}\qquad\mathcal{H}(\mathfrak{M}),v\models t(\phi).$$

Proof. Induction over ϕ, with trivial steps for Boolean operators and quantifiers. In the case for modal operators, we have

$$\begin{aligned}&\mathfrak{M},v\models x\heartsuit\lceil z:\phi\rceil\\ \iff{}&\gamma(v(x))\models\heartsuit[\![\phi]\!]^{z}_{\mathfrak{M},v}\\ \iff{}&\gamma(v(x))\models\heartsuit([\![\phi]\!]^{z}_{\mathfrak{M},v}\cap\mathsf{supp}(v(x)))\qquad\qquad\text{(naturality of predicate liftings)}\end{aligned}$$

where $\mathfrak{M}=(C,\gamma,I)$. The last condition is equivalent to $(v(x),(v(x),[\![\phi]\!]^{z}_{\mathfrak{M},v}\cap\mathsf{supp}(v(x)))$ being in the interpretation of $\heartsuit$ in $\mathcal{H}(\mathfrak{M})$, which by induction is equivalent to $\mathfrak{M},v\models\exists u:n.\,(x\heartsuit u\wedge\forall y.\,(y\in u\leftrightarrow(t(\phi)\wedge x\,\mathsf{supp}\,y)))$, i.e., to $\mathfrak{M},v\models t(x\heartsuit\lceil z:\phi\rceil)$. □

The key observation is now that if two points are locally indistinguishable in $\mathfrak{M}$, then the same is true in $\mathcal{H}(\mathfrak{M})$. Explicitly:

Lemma 30 (Neighbourhood Compatibility). *If two states have isomorphic $l+1$-neighbourhoods in $\mathfrak{M}$, then they have isormorphic l-neighbourhoods in $\mathcal{H}(\mathfrak{M})$.*

Proof. We begin by showing

Claim A: for each modal neighbourhood (d,B) in $\mathsf{N}^{\mathcal{H}(\mathfrak{M})}_{l}(c)$, we have $B\subseteq\mathsf{N}^{\mathfrak{M}}_{l+1}(c)$.

We necessarily have that (d,B) is reachable in one step from some state $e\in\mathsf{N}^{\mathcal{H}(\mathfrak{M})}_{l-1}(c))$. Once we prove

Claim B: the distance between any two states in $\mathfrak{M}$ is at most that in $\mathcal{H}(\mathfrak{M})$

we have $e\in\mathsf{N}^{\mathfrak{M}}_{l-1}(c)$. We then distinguish two cases:

- The step from e to (d,B) in $\mathcal{H}(\mathfrak{M})$ is via some $\heartsuit\in\Lambda$. Then $e=d$ and $B\subseteq\mathsf{supp}(e)$ by construction of the interpretation of $\heartsuit$ in $\mathcal{H}(\mathfrak{M})$, and we are done (in fact even $B\subseteq\mathsf{N}^{\mathfrak{M}}_{l}(c)$).
- The step from e to (d,B) in $\mathcal{H}(\mathfrak{M})$ is via $\in$. Since $B\subseteq\mathsf{supp}(d)$, this implies that every state in B is, in $\mathfrak{M}$, reachable from e in two steps, via d; since e had distance $l-1$, this proves the claim.

It remains to prove Claim B. All paths between states in $\mathcal{H}(\mathfrak{M})$ consist of single steps using supp, which can be done also in $\mathfrak{M}$, and double steps of one of the following types:

- $c\heartsuit(d,B)\heartsuit^{-}e$, where a superscript $^{-}$ denotes inverse relations (and we use predicate symbols to denote their interpretations in $\mathcal{H}(\mathfrak{M})$). In this case, $c=d=e$ by construction of $\mathcal{H}(\mathfrak{M})$.

- $c \in (d, B) \ni e$ – in this case, $c, e \in \mathsf{supp}(d)$, so in $\mathfrak{M}$ we can reach e from c in two steps using supp, via d.
- $c \heartsuit (d, B) \ni e$, or symmetrically – in this case, $c = d$ and hence $e \in \mathsf{supp}(c)$, that is, we can reach e from c in $\mathfrak{M}$ in one step.

We now proceed to prove the actual claim of the neighbourhood compatibility lemma: Let $\mathfrak{M} = (C, \gamma, I)$ be a supported coalgebraic model, and let $f : \mathsf{N}^{\mathfrak{M}}_{l+1}(c) \to \mathsf{N}^{\mathfrak{M}}_{l+1}(d)$ be an isomorphism (as specified in Definition 25, in particular $f(c) = d$). By Claim B, $\mathsf{N}^{\mathcal{H}(\mathfrak{M})}_{l}(c) \subseteq \mathsf{N}^{\mathfrak{M}}_{l+1}(c)$, so we can restrict f to a map $f_s : \mathsf{N}^{\mathcal{H}(\mathfrak{M})}_{l}(c) \to \mathcal{H}(\mathfrak{M})$ on states (we elide sort indices on carriers). By Claim A, we can moreover define a map $f_n : \mathsf{N}^{\mathcal{H}(\mathfrak{M})}_{l}(c) \to \mathcal{H}(\mathfrak{M})$ on modal neighbourhoods by $f_n(e, B) = (f(e), f[B])$; since f preserves supp, $B \subseteq \mathsf{supp}(e)$ implies $f[B] \subseteq \mathsf{supp}(f(e))$, i.e., $f_n(e, B)$ is indeed a modal neighbourhood in $\mathcal{H}(\mathfrak{M})$. We next show that $\bar{f} = (f_s, f_n)$ is a (two-sorted) homomorphism $\mathsf{N}^{\mathcal{H}(\mathfrak{M})}_{l} \to \mathcal{H}(\mathfrak{M})$: since f is an isomorphism of supported coalgebraic models, $\bar{f}$ preserves supp. To see that $\bar{f}$ preserves $\heartsuit$, let $e \heartsuit (e, B)$ in $\mathsf{N}^{\mathcal{H}(\mathfrak{M})}_{l}(c)$. By Claim B, $e \in \mathsf{N}^{\mathfrak{M}}_{l}(c)$ and hence $\mathsf{supp}(e) \subseteq \mathsf{N}^{\mathfrak{M}}_{l+1}(c)$, so that $\gamma|_{\mathsf{N}^{\mathfrak{M}}_{l+1}(c)}(e) = \gamma(e)$. It follows by construction of $\mathcal{H}(\mathfrak{M})$ that $\gamma|_{\mathsf{N}^{\mathfrak{M}}_{l+1}(c)}(e) \models \heartsuit B$. Since $f : \mathsf{N}^{\mathfrak{M}}_{l+1}(c) \to \mathsf{N}^{\mathfrak{M}}_{l+1}(d)$ is in particular an isomorphism of coalgebras, it follows that $\gamma|_{\mathsf{N}^{\mathfrak{M}}_{l+1}(d)}(f(e)) \models \heartsuit f[B]$. Moreover, since f preserves supp, $f(e) \in \mathsf{N}^{\mathfrak{M}}_{l}(d)$, so by the same argument as above, $\gamma|_{\mathsf{N}^{\mathfrak{M}}_{l+1}(d)}(f(e)) = \gamma(f(e))$, and hence $\gamma(f(e)) \models \heartsuit f[B]$. So, by construction of $\mathcal{H}(\mathfrak{M})$, $f(e) \heartsuit (f(e), f[B])$ in $\mathcal{H}(\mathfrak{M})$. Finally, it is clear that $\bar{f}$ preserves $\in$.

Preservation of all predicates implies that $\bar{f}$ is non-expansive w.r.t. Gaifman distance, and hence is a homomorphism $\mathsf{N}^{\mathcal{H}(\mathfrak{M})}_{l}(c) \to \mathsf{N}^{\mathcal{H}(\mathfrak{M})}_{l}(d)$. Analogously, we obtain from f^{-1} a homomorphism in the opposite direction, and the two homomorphisms are easily seen to be mutually inverse. □

We are now set to prove Gaifman locality for pure support-CPL:

Proof (Theorem 27). Let $\phi(x)$ be a formula in pure support-CPL with a single free variable x. By the standard Gaifman theorem, $t(\phi(x))$ is Gaifman l-local for some l, exponentially bounded in the quantifier rank (in the standard sense) of $(t(\phi))$ and hence also in $\mathsf{qr}(\phi)$. Then $\phi(x)$ is Gaifman $l+1$-local: Let states c, d in a supported coalgebraic model $\mathfrak{M}$ have isomorphic $l+1$-neighbourhoods. By the neighbourhood compatibility lemma, c and d have isomorphic l-neighbourhoods in $\mathcal{H}(\mathfrak{M})$, so that

$$\begin{aligned} \mathfrak{M} \models \phi(c) &\iff \mathcal{H}(\mathfrak{M}) \models t(\phi)(c) && \text{(Lemma 29)} \\ &\iff \mathcal{H}(\mathfrak{M}) \models t(\phi)(d) && \text{(Gaifman locality of } t(\phi(x))\text{)} \\ &\iff \mathfrak{M} \models \phi(d) && \text{(Lemma 29).} \end{aligned}$$

□

From Gaifman locality, we derive a simple corollary that asserts locality of coproduct-invariant formulas. The basic idea is taken from [Ott06], where the same statement is proved for relational models as Lemma 3.5. We apply a simpler if somewhat wholesale argument using Gaifman locality.

Definition 31. A formula $\phi(x)$ in pure support-CPL with a single free variable x is *invariant under coproducts* if for all supported coalgebraic models $\mathfrak{M} = (C, \gamma, I)$, $\mathfrak{N} = (D, \delta, J)$ and all states c in $\mathfrak{M}$, $\mathfrak{M} \models \phi(c)$ iff $\mathfrak{M} + \mathfrak{N} \models \phi(c)$, where $\mathfrak{M} + \mathfrak{N}$ is the coproduct of A and B (given by taking disjoint unions of carriers and support relations, with the coalgebra structure obtained by embedding TC and TD into $T(C + D)$ via the T-images of the coproduct injections).

Corollary 32. *If a formula $\phi(x)$ in pure support-CPL with a single free variable x is invariant under coproducts, then $\phi(x)$ is l-local for some $l \geq 0$, exponentially bounded in the quantifier rank of $\phi(x)$.*

Proof. By Theorem 27, $\phi(x)$ is Gaifman l-local for some $l \geq 0$ of the claimed size; we prove that $\phi(x)$ is in fact l-local. Thus, let $\mathfrak{M}$ be a supported coalgebraic model, let c be a state in $\mathfrak{M}$, and let $\mathfrak{N} = \mathsf{N}^{\mathfrak{M}}_{l}(c)$. We have to prove that $\mathfrak{M} \models \phi(c)$ iff $\mathfrak{N} \models \phi(c)$. Denote the left hand copy of c in $\mathfrak{M} + \mathfrak{N}$ by c_1 and the right hand copy by c_2. Then $\mathfrak{M} \models \phi(c)$ iff $\mathfrak{M} + \mathfrak{N} \models \phi(c_1)$ (by invariance under coproducts) iff $\mathfrak{M} + \mathfrak{N} \models \phi(c_2)$ (by Gaifman l-locality) iff $\mathfrak{N} \models \phi(c)$ (by invariance under coproducts). □

5. A Van Benthem/Rosen Theorem for CPL

We now proceed to prove our main result, the coalgebraic van Benthem/Rosen theorem. Since we need Gaifman locality, we work in pure support-CPL rather than in pure CPL. Of course, the result then holds, a fortiori, also for pure CPL. Formally, the statement of the result is as follows.

Definition 33. A *pointed (supported) coalgebraic model* $\mathfrak{M}, c$ consists of a (supported) coalgebraic model $\mathfrak{M}$ and a state c in $\mathfrak{M}$. We write $\mathfrak{M}, c \approx \mathfrak{N}, d$ for pointed (supported) coalgebraic models $\mathfrak{M}, c$ and $\mathfrak{N}, d$ if c and d are behaviourally equivalent as states in the underlying T-coalgebras; similarly for $\approx_k$. We say that a formula $\phi(x)$ in pure support-CPL with a single free variable x is *behavioural-equivalence invariant* if $\mathfrak{M} \models \phi(c)$ whenever $\mathfrak{N} \models \phi(d)$ and $\mathfrak{M}, c \approx \mathfrak{N}, d$.

Theorem 34 (Van Benthem/Rosen theorem for pure support-CPL)**.** *Let Λ be separating, and let $\phi(x)$ be a formula in pure support-CPL with a single free variable x. Then $\phi(x)$ is behavioural-equivalence invariant (over finite models) iff it is equivalent (over finite models) to an infinitary modal Λ-formula of finite modal rank.*

Before we prove this theorem, we state a few variants as immediate consequences:

Corollary 35 (Van Benthem/Rosen theorem for pure CPL)**.** *Let Λ be separating, and let $\phi(x)$ be a pure CPL formula with a single free variable x. Then $\phi(x)$ is behavioural-equivalence invariant (over finite models) iff it is equivalent (over finite models) to an infinitary modal Λ-formula of finite modal rank.*

Proof. Pure CPL is a sublogic of pure support-CPL, and the standard translation ends up in pure CPL. □

Corollary 36 (Van Benthem/Rosen theorem for finite modal similarity types)**.** *Let Λ be separating and finite, and let $\phi(x)$ be a pure CPL formula with one free variable. Then $\phi(x)$ is behavioural-equivalence invariant (over finite models) iff it is equivalent (over finite models) to a finitary modal Λ-formula.*

Proof. For finite Λ, there are, up to equivalence, only finitely many modal Λ-formulas of a given finite rank. □

Remark 37. Out of the above results, only Corollary 36 can be seen as characterizing a fragment of CPL: *over finite similarity types, the behavioural-equivalence-invariant fragment of pure CPL is coalgebraic modal logic* – for infinite similarity types, Corollary 35 maps the behavioural-equivalence-invariant fragment of pure CPL into an infinitary modal logic that does not embed into CPL. In the case of infinite collections of independent modal operators, such as infinitely many propositional atoms, or boxes for infinitely many unrelated agents, we can immediately and trivially reduce to the finite case (i.e., to the finite collection of operators actually occurring in a given behavioural-equivalence-invariant formula), and in fact proofs of the standard (relational) Rosen theorem begin with exactly such a reduction to finitely many propositional atoms [Ros97, Ott06]. In particular, Corollary 36 does imply the original theorems of van Benthem and Rosen. The problematic case are infinite collections of interdependent operators as, e.g., in graded modal logic; for such cases, we leave a strengthening of Corollary 35 as an open problem. Note that the problem is not settled by a counterexample exhibited in an earlier version of this work [SP10a], which requires the higher expressivity of the correspondence language considered there and does not apply to CPL.

Using Lemma 19 above, we can derive an internal characterization of the behavioural-equivalence-invariant fragment of CPL:

Corollary 38. *The behavioural-equivalence-invariant formulas in one free variable (over finite models) of pure CPL are, up to equivalence (over finite models), precisely the equality-free and quantifier-free formulas in the single-variable fragment of pure CPL.*

Remark 39. Unlike Rosen's proof [Ros97], the original proof of van Benthem's characterization result as well as, e.g., the van-Benthem-type result proved for neighbourhood models in Hansen et al. [HKP09] rely on machinery from classical model theory, in particular compactness and saturation. Over finite models, these methods are, of course, not available; but even over arbitrary models, there are further sources of non-compactness in the coalgebraic setup [Sch07]. For example, T may be finitary, i.e. impose finite branching, which is easily seen to disable compactness. Also in the subdistribution example, the set of formulas

$\{\neg L_1 p\} \cup \{L_{1-1/n}p \mid n \in \mathbb{N}\}$ for a propositional variable p is finitely satisfiable (in finite models) but not satisfiable (in any model). This is one reason to go directly for a Rosen-style proof covering both finite and arbitrary models.

We proceed to develop some facts concerning (partial) tree unravellings of coalgebras, in generalization of corresponding techniques for Kripke frames, including a not entirely trivial coalgebraic generalization of the fact that on trees, behavioural equivalence is equivalent to bounded behavioural equivalence (Lemma 41). The basic notion underlying these concepts is the following.

Definition 40. A pointed supported coalgebraic model $\mathfrak{M}, c$ is a *tree of depth* l if the support relation supp in $\mathfrak{M}$ forms a tree of depth l with root c, i.e., if supp is loop-free and every state is reachable from c by a unique path of length at most l, and moreover all leaves of this tree have the default successor structure $\bot_T$.

Lemma 41. *Let $\mathfrak{M}, c$ and $\mathfrak{N}, d$ be pointed supported coalgebraic models that are trees of depth at most l. Then $\mathfrak{M}, c \approx \mathfrak{N}, d$ iff $\mathfrak{M}, c \approx_l \mathfrak{N}, d$.*

Proof. 'Only if' holds by Lemma 8. To prove 'if', let γ and δ denote the underlying coalgebra structures of $\mathfrak{M}$ and $\mathfrak{N}$, respectively. Similarly as in [Pat03], let $f^0 : 1 \to T1$ be the map induced by the distinguished element $\bot_T$ of $T1$, and put $f^n = T^n f^0 : T^n 1 \to T^n T1 = TT^n 1$, a T-coalgebra on $T^n 1$. We shall show that the fact that $\mathfrak{M}, c$ is a tree of depth at most l implies that

$$(*) \qquad f^l \gamma_l = \gamma_{l+1},$$

which means that $\gamma_l : \mathfrak{M} \to (T^l 1, f^l)$ is a T-coalgebra morphism as $\gamma_{l+1} = T\gamma_l\gamma$; similarly for δ_l. As $\gamma_l(c) = \delta_l(d)$ by assumption, this implies the claim: we can identify c and d by two coalgebra morphisms into $(T^l 1, f^l)$.

It remains to prove $(*)$. We generalize to $f^k\gamma_k = \gamma_{k+1}$ for all $k \geq l$, i.e. commutation of

$$\begin{tikzcd} & \bullet \arrow[dl, "\gamma_k"'] \arrow[dr, "\gamma_{k+1}"] & \\ T^k 1 \arrow[rr, "f^k"'] & & T^{k+1}1, \end{tikzcd}$$

and proceed by induction over l.

$l = 0$: We conduct a further induction over k. For $k = 0$, note that the underlying tree of $\mathfrak{M}, c$ consists only of the root c, and we have

$$f^0\gamma_0(c) = \bot_T = T!(\bot_T) = T!(\gamma_0(c)) = \gamma_1(c)$$

where the second equality is by pointedness of T (Lemma 23). For $k > 0$, and assuming the claim for $k - 1$, we calculate

$$f^k\gamma_k = T(f^{k-1})T\gamma_{k-1}\gamma = T(f^{k-1}\gamma_{k-1})\gamma = T(\gamma_k)\gamma = \gamma_{k+1}$$

where we apply the inductive hypothesis in the second-to-last step.

$l - 1 \to l$: By the inductive hypothesis, the claim holds at all nodes of the tree $\mathfrak{M}$ strictly below the root node c (which have tree depth at most $l - 1$). Thus, we have to show only that $f^k(\gamma_k(c)) = \gamma_{k+1}(c)$. As above, we calculate

$$f^k\gamma_k(c) = T(f^{k-1})T\gamma_{k-1}\gamma(c) = T(f^{k-1}\gamma_{k-1})\gamma(c) = T(\gamma_k)\gamma = \gamma_{k+1}(c)$$

where the second-to-last step holds because $\gamma(c) \in TS \subseteq TC$, where $S = \mathsf{supp}(c) \subseteq C$ is the set of children of c in the underlying tree of $\mathfrak{M}$, and by the inductive hypothesis, $(f^{k-1}\gamma_{k-1})|_S = \gamma_k|_S$, so that $T(f^{k-1}\gamma_{k-1})|_{TS} = T(\gamma_k)|_{TS}$. □

The core construction is described in the following lemma.

Lemma 42 (Unravelling)**.** *Let $\mathfrak{M}, c$ be a (finite) pointed supported coalgebraic model, and let $k \geq 0$. Then there exists a (finite) pointed supported coalgebraic model $\mathfrak{N}, d$ such that $\mathfrak{M}, c \approx \mathfrak{N}, d$, and moreover $\mathsf{N}_k^{\mathfrak{N}}(d), d$ is a tree of depth at most k.*

Proof. Let $\mathfrak{M}$ be based on a coalgebra $\gamma : C \to TC$. We construct a coalgebra D as the disjoint union $D = (\coprod_{i=0}^{k}\{c\} \times C^i) + C$ where we write inl and inr for the left and right injection, respectively. Note that D is finite if C is finite. The idea here is that the left-hand summand is a tree-shaped unfolding of C up to depth k, and the right-hand summand C represents the original coalgebra structure, into which every path runs after $k+1$ steps. Formally, the transition structure $\delta : D \to TD$ is defined by case analysis:

$$\begin{aligned}\delta(\mathit{inr}(c')) &= T\mathit{inr}(\gamma(c')) & \\ \delta(\mathit{inl}(c_0, c_1, \dots, c_k)) &= T\mathit{inr}(\gamma(c_k)) & (c_0 = c)\\ \delta(\mathit{inl}(c_0, c_1, \dots, c_j)) &= Tf(\gamma(c_j)) & (c_0 = c)\end{aligned}$$

where in last clause, $0 \le j < k$ and $f : C \to D$ is defined by $f(c') = \mathit{inl}(c, c_1, \dots, c_j, c')$. We extend (D, δ) to a supported coalgebraic model $\mathfrak{N}$ by interpreting supp as expected, i.e., so that $\mathsf{supp}(x)$ is as in $\mathfrak{M}$ on the right hand summand, $\mathsf{supp}(\mathit{inl}(c_0, \dots, c_j)) = \mathit{inl}[\{(c_0, \dots, c_j)\} \times C]$ for $j < k$, and $\mathsf{supp}(\mathit{inl}(c_0, \dots, c_k)) = \mathit{inr}[C]$. We put $d = \mathit{inr}(c)$.

This turns the projection $\pi : D \to C$ defined by $\pi(\mathit{inr}(c')) = c'$ and $\pi(\mathit{inl}(c_0, \dots, c_i)) = c_i$ into a coalgebra morphism, as

$$\begin{aligned}T\pi(\delta(\mathit{inr}(c'))) &= T\pi(T\mathit{inr}(\gamma(c'))) = \gamma(c') = \gamma(\pi(\mathit{inr}(c'))) \\ T\pi(\delta(\mathit{inl}(c_0, \dots c_k))) &= T\pi(T\mathit{inr}(\gamma(c_k))) = \gamma(c_k) = \gamma(\pi(\mathit{inl}(c_0, \dots c_k))) \\ T\pi(\delta(\mathit{inl}(c_0, \dots, c_j))) &= T\pi(Tf(\gamma(c_j))) = \gamma(c_j) = \gamma(\pi(\mathit{inl}(c_0, \dots, c_j))) \qquad (j < k)\end{aligned}$$

(noting in the last line that $\pi f = id$), so that $\mathfrak{M}, c \approx \mathfrak{N}, d$. By construction, $\mathsf{N}_k^{\mathfrak{N}}(d), d$ is a tree of depth at most k (in fact, exactly k). □

Finally, we note that bounded behavioural equivalence is local in the expected way:

Lemma 43. *Let $\mathfrak{M}, c$ be a pointed supported coalgebraic model, and let $k \ge 0$. Then $\mathfrak{M}, c \approx_k \mathsf{N}_k^{\mathfrak{M}}(c), c$.*

Proof. Let γ denote the transition structure of $\mathfrak{M}$. One shows by induction over $l \ge 0$ that for all l, $\gamma_l(c) \in T^l 1$ depends only on states reachable from c in at most l steps in the supporting Kripke frame and on the successor structures of states reachable in at most $l - 1$ steps. For $l = k$, these data are preserved in the transition from $\mathfrak{M}$ to $\mathsf{N}_k^{\mathfrak{M}}(c)$. □

We are now ready to prove the main result:

Proof (Theorem 34). To prove the non-trivial direction, let $\phi(x)$ be invariant under behavioural equivalence. The pattern of the proof is as in Otto [Ott06]: one first observes that $\phi(x)$ is, in particular, invariant under coproducts (since coproduct injections are coalgebra morphisms), and therefore k-local for some $k \ge 0$ by Corollary 32.

In the next step, we prove that $\phi(x)$ is even $\approx_k$-invariant. The claim of the theorem then follows by Lemma 13.

Thus let $\mathfrak{M}, c \approx_k \mathfrak{N}, d$; we have to prove that $\mathfrak{M} \models \phi(c)$ iff $\mathfrak{N} \models \phi(d)$. By k-locality, it suffices to prove that $\mathsf{N}_k^{\mathfrak{M}}(c) \models \phi(c)$ iff $\mathsf{N}_k^{\mathfrak{N}}(d) \models \phi(d)$. By invariance under behavioural equivalence, this follows if $\mathsf{N}_k^{\mathfrak{M}}(c), c \approx \mathsf{N}_k^{\mathfrak{N}}(d), d$. By the unravelling lemma (Lemma 42), we may assume that $(\mathsf{N}_k^{\mathfrak{M}}(c), c)$ and $(\mathsf{N}_k^{\mathfrak{N}}(d), d)$ are trees of depth at most k. The claim then follows by Lemma 41 once we show that $\mathsf{N}_k^{\mathfrak{M}}(c), c \approx_k \mathsf{N}_k^{\mathfrak{N}}(d), d$, which however follows from $\mathfrak{M}, c \approx_k \mathfrak{N}, d$ and Lemma 43. □

Example 44. Corollary 35 applies to all logics introduced in Sections 2 and 3, see Example 11. Corollary 36 applies to those logics that have only finitely many modalities, i.e., all except (unbounded) graded modal logic and probabilistic modal logic. Explicitly, writing *bisimulation-invariant* in all cases where behavioural equivalence is characterized by known forms of bisimulation [Rut00, Sta11, GS13], we have the following applications.

(1) *Graded modalities:* Both over the class of all models and over the class of finite models, any bisimulation-invariant formula in graded CPL is equivalent to a bounded-depth infinitary graded modal formula. Although first-order logic with counting quantifiers embeds into graded CPL, this result is incomparable to the van-Benthem-type result for the former proved by de Rijke [dR00]: standard FOL with counting quantifiers has a relational semantics, i.e., considers less general

models. Formulas that are bisimulation-invariant over relational models need not be bisimulation-invariant over multigraph models, and an equivalence of a formula with a modal formula that holds over relational models need not hold over multigraph models.

De Rijke's result yields finitary modal equivalents (over general models), and we conjecture that the result for graded CPL can be sharpened similarly. For the time being, we note that for k-bounded graded CPL (see Section 2), Corollary 36 does yield a characterization of the bisimulation-invariant fragment of graded CPL. Notice also that although every formula in graded CPL lives in some bounded fragment, this does not immediately imply a characterization theorem for the whole of graded CPL, as it is not clear that every bisimulation-invariant formula in graded CPL is also bisimulation-invariant in k-bounded graded CPL for some k.

(2) *Probabilistic modalities:* Both over the class of all models and over the class of finite models, any bisimulation-invariant formula in probabilistic CPL is equivalent to a bounded-depth infinitary probabilistic modal formula. Again, we conjecture that the result can be sharpened to replace 'infinitary' with 'finitary'. To our best knowledge, however, this is currently the only known result relating the bisimulation-invariant fragment of a probabilistic first-order logic with probabilistic modal logic.

(3) *Finite similarity types:* For each of
- neighbourhood logic
- monotone neighbourhood logic
- conditional logic
- K

we obtain by Corollary 36 that the behavioural-equivalence-invariant fragment of the respective instance of CPL is the corresponding coalgebraic modal logic, both over finite and over arbitrary models. For K, this just reproves the classical van Benthem / Rosen theorem, as in this case CPL is equivalent to the usual first-order correspondence language. For neighbourhood logic, the van Benthem version of the theorem (i.e., the version for unrestricted models) follows from the corresponding theorem proved by Hansen et al. [HKP09], as neighbourhood CPL embeds into their correspondence language (Remark 20). The Rosen version (for finite models) appears to be new. The characterization theorems we obtain for monotone neighbourhood logic and conditional logic appear to be new.

6. Conclusions

Coalgebraic predicate logic (CPL) [LPSS12] generalizes Chang's modal first-order language to a natural first-order extension of coalgebraic logic. In the current work, we have established a van Benthem/Rosen type theorem for CPL. In its most general form, which applies to any coalgebraic logic as long as the modal similarity type is separating, the result states that both over arbitrary and over finite models, every formula that is invariant under behavioural equivalence is equivalent to an infinitary modal formula *of bounded depth*. As an easy corollary to this, a finitary version is obtained which improves this to equivalence to a finitary modal formula provided that the modal similarity type is finite; in a nutshell, *for finite and separating modal similarity types, coalgebraic modal logic is the behavioural-equivalence-invariant fragment of CPL.* The infinitary result yields for example that a formula in a natural first order logic of multigraphs with counting quantifiers is invariant under behavioural equivalence iff it is equivalent to a bounded-depth infinitary graded modal formula. The finitary result reproduces the classical van Benthem/Rosen theorem, and yields new characterizations of conditional logic, classical modal logic, monotone modal logic, and a bounded version of graded modal logic as the behavioural-equivalence-invariant fragments of the respective instances of CPL, as well as a new Rosen-type theorem for neighbourhood logic, the van-Benthem-type version having already been proved by Hansen et al. [HKP09].

It remains an open problem to extend the finitary result to infinite modal similarity types. This would in particular imply finitary van Benthem/Rosen theorems for graded modal logic over multigraphs [DV02], complementing a van Benthem-type result for graded modal logic over Kripke frames [dR00], and probabilistic modal logic. A further interesting direction for future investigation is to extend the ambient logic. E.g. one might hope to obtain a coalgebraic analogue of the characterization of the modal μ-calculus as the bisimulation-invariant fragment of monadic second order logic due to Janin and Walukiewicz [JW95]; some results to this effect have recently been established by Enqvist, Seifan and Venema [ESV15b, ESV15a]. A fragment of coalgebraic monadic second order logic, so far studied in its neighbourhood instance over topological spaces [MZ80, tCGS09], is the language L^t which imposes particular restrictions on the use of second-order quantification leading to essentially first-order behaviour, and notably has $S4$ as its bisimulation-invariant fragment over topological spaces [tCGS09]. From this, two further problems for future research arise: to sharpen the coalgebraic van Benthem/Rosen theorem to allow for the more expressive language L^t (interpreted coalgebraically) as the ambient correspondence language, and to develop the model theory of CPL with frame conditions.

Acknowledgements. The authors wish to thank Helle Hvid Hansen for useful discussions and pointers to the literature and to Erwin R. Catesbeiana for elaborating on the role of $\bot_T$. In addition, we would also like to repeat our thanks to everybody who discussed with us early ideas on the design of CPL, in particular Alexander Kurz and Drew Moshier, and of course to Katsuhiko Sano for ongoing collaboration on CPL.

REFERENCES

[Bar93] M. Barr. Terminal coalgebras in well-founded set theory. *Theoret. Comput. Sci.*, 114:299–315, 1993.

[BdRV01] P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*. Cambridge University Press, 2001.

[Cha73] C. Chang. Modal model theory. In *Cambridge Summer School in Mathematical Logic*, volume 337 of *LNM*, pages 599–617. Springer, 1973.

[Che80] B. Chellas. *Modal Logic*. Cambridge University Press, 1980.

[CP07] C. Cîrstea and D. Pattinson. Modular construction of complete coalgebraic logics. *Theoret. Comput. Sci.*, 388:83–108, 2007.

[DL10] S. Demri and D. Lugiez. Complexity of modal logics with Presburger constraints. *J. Appl. Logic*, 8:233–252, 2010.

[DO05] A. Dawar and M. Otto. Modal characterisation theorems over special classes of frames. In *Logic in Computer Science, LICS 05*, pages 21–30. IEEE Computer Society, 2005.

[dR00] M. de Rijke. A note on graded modal logic. *Stud. Log.*, 64:271–283, 2000.

[DV02] G. D'Agostino and A. Visser. Finality regained: A coalgebraic study of Scott-sets and multisets. *Arch. Math. Logic*, 41:267–298, 2002.

[ESV15a] S. Enqvist, F. Seifan, and Y. Venema. Expressiveness of the modal μ-calculus on monotone neighborhood structures. *CoRR*, abs/1502.07889, 2015.

[ESV15b] S. Enqvist, F. Seifan, and Y. Venema. Monadic second-order logic and bisimulation invariance for coalgebras. In *Logic in Computer Science, LICS 2015*. IEEE, 2015.

[Fin72] K. Fine. In so many possible worlds. *Notre Dame J. Formal Logic*, 13:516–520, 1972.

[Gai82] H. Gaifman. On local and non-local properties. In *Logic Colloquium 1981*, pages 105–135. North Holland, 1982.

[GO06] V. Goranko and M. Otto. Model theory of modal logic. In P. Blackburn, F. Wolter, and J. van Benthem, editors, *Handbook of Modal Logic*, pages 255–325. Elsevier, 2006.

[GS13] D. Gorín and L. Schröder. Simulations and bisimulations for coalgebraic modal logics. In *Algebra and Coalgebra in Computer Science, CALCO 2013*, volume 8089 of *LNCS*, pages 253–266. Springer, 2013.

[Gum05] H. P. Gumm. From T-coalgebras to filter structures and transition systems. In *Algebra and Coalgebra in Computer Science, CALCO 05*, volume 3629 of *LNCS*, pages 194–212. Springer, 2005.

[Hal90] J. Halpern. An analysis of first-order logics of probability. *Artif. Intell.*, 46:311–350, 1990.

[HKP09] H. Hansen, C. Kupke, and E. Pacuit. Neighbourhood structures: Bisimilarity and basic model theory. *Log. Methods Comput. Sci.*, 5, 2009.

[HM01] A. Heifetz and P. Mongin. Probabilistic logic for type spaces. *Games and Economic Behavior*, 35:31–53, 2001.

[Jac10] B. Jacobs. Predicate logic for functors and monads, 2010.

[JW95] D. Janin and I. Walukiewicz. Automata for the modal μ-calculus and related results. In *Mathematical Foundations of Computer Science, MFCS 1995*, volume 969 of *LNCS*, pages 552–562. Springer, 1995.

[Lib04] L. Libkin. *Elements of finite model theory*. Springer, 2004.

[LPS13] T. Litak, D. Pattinson, and K. Sano. Coalgebraic predicate logic: Equipollence results and proof theory. In G. Bezhanishvili, S. Löbner, V. Marra, and F. Richter, editors, *Logic, Language, and Computation, TbiLLC 2011, Revised Selected Papers*, volume 7758 of *LNCS*, pages 257–276. Springer, 2013.

[LPSS12] T. Litak, D. Pattinson, K. Sano, and L. Schröder. Coalgebraic predicate logic. In A. Czumaj, K. Mehlhorn, A. Pitts, and R. Wattenhofer, editors, *Automata, Languages, and Programming, ICALP 2012*, volume 7392 of *LNCS*, pages 299–311. Springer, 2012.

[LS91] K. Larsen and A. Skou. Bisimulation through probabilistic testing. *Inform. Comput.*, 94:1–28, 1991.

[MM77] J. Makowsky and A. Marcja. Completeness theorems for modal model theory with the Montague-Chang semantics I. *Math. Logic Quarterly*, 23:97–104, 1977.

[MZ80] J. Makowsky and M. Ziegler. Topological model theory with an interior operator: Consistency properties and back-and forth arguments. *Arch. math. Logik*, 20:37–54, 1980.

[Ott06] M. Otto. Bisimulation invariance and finite models. In *Logic Colloquium 02*, volume 27 of *Lect. Notes Log.*, pages 276–298. ASL, 2006.

[Pat03] D. Pattinson. Coalgebraic modal logic: Soundness, completeness and decidability of local consequence. *Theoret. Comput. Sci.*, 309:177–193, 2003.

[Pat04] D. Pattinson. Expressive logics for coalgebras via terminal sequence induction. *Notre Dame J. Formal Logic*, 45:19–33, 2004.

[Ros97] E. Rosen. Modal logic over finite structures. *J. Logic, Language and Information*, 6(4):427–439, 1997.

[Rut00] J. Rutten. Universal coalgebra: A theory of systems. *Theoret. Comput. Sci.*, 249:3–80, 2000.

[Sch07] L. Schröder. A finite model construction for coalgebraic modal logic. *J. Log. Algebr. Prog.*, 73:97–110, 2007.

[Sch08] L. Schröder. Expressivity of coalgebraic modal logic: The limits and beyond. *Theoret. Comput. Sci.*, 390:230–247, 2008.

[Sgr80] J. Sgro. The interior operator logic and product topologies. *Trans. AMS*, 258(1):pp. 99–112, 1980.

[SLG11] J. Seligman, F. Liu, and P. Girard. Logic in the community. In *Logic and Its Applications, ICLA 2011*, volume 6521 of *LNCS*, pages 178–188. Springer, 2011.

[SP09] L. Schröder and D. Pattinson. Strong completeness of coalgebraic modal logics. In *Theoretical Aspects of Computer Science, STACS 09*, Leibniz International Proceedings in Informatics, pages 673–684. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2009.

[SP10a] L. Schröder and D. Pattinson. Coalgebraic correspondence theory. In *Foundations of Software Science and Computations Structures, FOSSACS 2010*, volume 6014 of *LNCS*, pages 328–342. Springer, 2010.

[SP10b] L. Schröder and D. Pattinson. Rank-1 modal logics are coalgebraic. *J. Log. Comput.*, 20:1113–1147, 2010.

[SP11] L. Schröder and D. Pattinson. Modular algorithms for heterogeneous modal logics via multi-sorted coalgebra. *Mathematical Structures in Computer Science*, 21:235–266, 2011.

[Sta11] S. Staton. Relating coalgebraic notions of bisimulation. *Log. Methods Comput. Sci.*, 7, 2011.

[tCGS09] B. ten Cate, D. Gabelaia, and D. Sustretov. Modal languages for topology: Expressivity and definability. *Ann. Pure Appl. Logic*, 159(1-2):146–170, 2009.

[vB76] J. van Benthem. *Modal Correspondence Theory*. PhD thesis, Department of Mathematics, University of Amsterdam, 1976.

[vB84] J. van Benthem. Correspondence theory. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic: Volume II: Extensions of Classical Logic*, pages 167–247. Reidel, Dordrecht, 1984.

[Zie85] A. Ziegler. Topological model theory. In J. Barwise and S. Feferman, editors, *Model-Theoretic Logics*. Springer, 1985.

FAU Erlangen-Nürnberg
E-mail address: Lutz.Schroeder@cs.fau.de

Australian National University
E-mail address: dirk.pattinson@anu.edu.au

FAU Erlangen-Nürnberg
E-mail address: tadeusz.litak@gmail.com