Coalgebraic Correspondence Theory

Lutz Schröder[1,*] and Dirk Pattinson[2,**]

[1] DFKI Bremen and Department of Computer Science, Universität Bremen
[2] Department of Computing, Imperial College London

Abstract. We lay the foundations of a first-order correspondence theory for coalgebraic logics that makes the transition structure explicit in the first-order modelling. In particular, we prove a coalgebraic version of the van Benthem/Rosen theorem stating that both over arbitrary structures and over finite structures, coalgebraic modal logic is precisely the bisimulation invariant fragment of first-order logic.

Introduction

Viewing modal logic as a sub-language of first-order logic via the standard translation is the starting point of modal correspondence theory. One of the core results of this area, van Benthem's theorem [22], states that modal logic is precisely the bisimulation invariant fragment of first-order logic over relational structures. This result has been extended to finite structures by Rosen [18], and special frame classes have been considered in [5]. Results of this kind characterize the expressive power of modal logic – slightly reworded, they state that modal logic can express the same bisimulation-invariant properties as first-order logic.

Here, we extend these results to coalgebraic modal logic [16], thus making initial forays into *coalgebraic correspondence theory*. Coalgebraic modal logic is a generic framework for modal logics that captures a wide range of modal logics from the literature, e.g. the modal logic of neighbourhood frames (called classical modal logic in [3]), normal modal logics [2], graded and probabilistic modal logics [7,14,12], and various conditional logics [3]. The parameters of the framework are a *type functor*, whose coalgebras serve as models, and a choice of *predicate liftings* defining the modal operators. The predicate liftings act as an interface to the type functor, and as such form an integral part of the semantics.

Our correspondence language is a multi-sorted first-order logic, inspired by the correspondence language for neighbourhood frames of [11]. It includes a dedicated sort to represent the type functor and thus provides a full model of coalgebras. Moreover, the language explicitly incorporates predicate liftings, following the semantic principles outlined above. We adapt the method of Rosen (and a related proof by Otto [15]) to prove that, under suitable assumptions, coalgebraic modal logic is, both over finite and over arbitrary structures, precisely the fragment of the coalgebraic correspondence language 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, which is a fairly typical phenomenon.

To show that a first-order formula that is invariant under behavioural equivalence can be characterized by a *finitary* formula, we have to assume that the underling 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. graded and probabilistic modal logic. For the general case, we do provide a characterization result in terms of bounded-rank modal formulas with infinitary conjunction, and a counter-example showing that equivalence to a finitary modal formula fails in general. This result applies to essentially all logics of interest, and indeed covers most of the way to the finitary result in terms of its proof, as the latter is an easy corollary to it in the case of finite signatures. 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 structures is equivalent to a possibly infinitary formula *of bounded depth* in graded modal logic. Although similar results have previously been obtained over the class of *all* structures [6], our result seems to be the first Rosen-type result for graded modal logic over *finite* structures.

We note that the design of the correspondence language used as the setting for our results is a delicate affair: the translation of coalgebraic semantics, like already the translation of neighbourhood semantics used in [11], needs to include a sort of neighbourhoods, i.e. subsets of the state space. On the other hand, we need to avoid the full expressive power of monadic second order logic, in which the van Benthem/Rosen theorem fails to hold, as it contains the μ-calculus (which, in the standard relational case, is in fact its bisimulation-invariant fragment [13]). This forces us to adopt a relaxed interpretation of the neighbourhood sort by suitable subsets of the full powerset. In our setup, the key to a suitable notion of model in this sense is the inclusion of explicit distinguished supports in the language; a particularly pleasant effect of this language extension is that it simultaneously acts as the key to enabling the use of Gaifman locality.

1   Coalgebra and Modal Logic

Throughout the paper, we fix a modal similarity type Λ consisting of modal operators with associated arities. As we will be considering models rather than frames, we express propositional variables as nullary modal operators. The set $\mathcal{F}(\Lambda)$ of Λ-*formulas* is then given by the grammar

$$\mathcal{F}(\Lambda) \ni \phi, \psi ::= \bot \mid \phi \wedge \psi \mid \neg \phi \mid \heartsuit(\phi_1, \ldots, \phi_n)$$

where $\heartsuit \in \Lambda$ is n-ary. We denote the language that admits arbitrary conjunctions of sets of formulas rather than just binary conjunctions by $\mathcal{F}_\infty(\Lambda)$. We write $\operatorname{rank}(\phi)$ for the maximal nesting depth of modal operators in the formula ϕ,

defined formally as $\text{rank}(\bot) = 0$, $\text{rank}(\bigwedge \Phi) = \sup_{\phi \in \Phi} \text{rank}(\phi)$, $\text{rank}(\neg \phi) = \text{rank}(\phi)$ and $\text{rank}(\heartsuit(\phi_1, \ldots, \phi_n)) = 1 + \max\{\text{rank}(\phi_1), \ldots, \text{rank}(\phi_n)\}$. Thus, the rank of a formula in $\mathcal{F}_\infty(\Lambda)$ may be infinite.

Formulas over Λ are interpreted over coalgebras with respect to a Λ-*structure* that consists of an endofunctor $T : \mathsf{Set} \to \mathsf{Set}$ on the category of sets, together with an assignment

$$\llbracket \heartsuit \rrbracket : \mathcal{Q}^n \to \mathcal{Q} \circ T$$

of natural transformations (the *predicate liftings*) where $\mathcal{Q} : \mathsf{Set}^{op} \to \mathsf{Set}$ is the contravariant powerset functor. The functor T is called the *underlying functor* of the structure, and we usually refer to the structure just in terms of its underlying functor, leaving the assignments of predicate liftings implicit.

Assumption 1. We assume w.l.o.g. that T preserves injective maps [1]. For ease of notation, we will in fact sometimes assume that subset inclusions $X \hookrightarrow Y$ are mapped 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 correspondence language:

Definition 2. A set $A \subseteq X$ is a *support* of $t \in TX$ if $t \in TA$.

Support has played a role in various coalgebraic model constructions, see e.g. [21]. We keep the notion of support as broad as possible, and in particular do not insist on minimality, as set of supports of $t \in TX$ does not necessarily have a smallest element with respect to subset inclusion [9].

Given a Λ-structure T, a T-*coalgebra* is a pair (C, γ) where C is a set (of states) and $\gamma : C \to TC$ is a (transition) function. We identify T-coalgebras (C, γ) with their carrier set C in case the transition function is clear from the context. The *semantics* $\llbracket \phi \rrbracket_C \subseteq C$ of a Λ-formula ϕ with respect to a T-coalgebra (C, γ) is given inductively by

$$\llbracket \heartsuit(\phi_1, \ldots, \phi_n) \rrbracket_C = \gamma^{-1} \circ \llbracket \heartsuit \rrbracket_C(\llbracket \phi_1 \rrbracket_C, \ldots, \llbracket \phi_n \rrbracket_C)$$

where $\heartsuit \in \Lambda$ is n-ary, together with the usual clauses for the propositional connectives. We write $(C, c) \models \phi$ if $c \in \llbracket \phi \rrbracket_C$.

Example 3. The following logics are covered by the coalgebraic approach.

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 $\mathcal{F}(\Lambda)$ is interpreted over T-coalgebras by virtue of the structure

$$\llbracket \Diamond \rrbracket_X(A) = \{(B, C) \in TX \mid B \cap A \neq \emptyset\} \qquad \llbracket p \rrbracket_X = \{(B, C) \in TX \mid p \in C\}.$$

Clearly this semantics coincides with the standard textbook semantics of K [2].

2. The modal logic of neighbourhood frames (classical modal logic in [3]) 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 classical modal logic arises via the lifting

$$\llbracket \Box \rrbracket_X(A) = \{(N, B) \in TX \mid A \in N\}$$

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

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

where upwards closure refers to subset inclusion.

4. Conditional logic [3] 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

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

(again, the powerset on the left of the function space is contravariant) where propositional constants are interpreted as above and the lifting

$$\llbracket \Rightarrow \rrbracket_X(A, B) = \{(f, D) \in TX \mid f(A) \subseteq B\}$$

induces the standard semantics of conditional logic.

5. The similarity type of graded modal logic features, apart from propositional constants, an indexed collection of operators $\Diamond_k$ for $k \in \omega$. The intuitive reading of $\Diamond_k \phi$ is that ϕ holds in more than k successors. To retain naturality of predicate liftings, we slightly deviate from the traditional semantics [7] and interpret graded modal logic over coalgebras of the functor T (left) where we use the liftings $\llbracket \Diamond_k \rrbracket$ (right)

$$TX = \{f : X \to \omega\} \times \mathcal{P}(\mathsf{P}) \qquad \llbracket \Diamond_k \rrbracket_X(A) = \{(f, D) \in TX \mid \sum_{x \in A} f(x) > k\}$$

to interpret modal operators. In other words, we interpret graded modal logic over multigraphs [4] where the graded modalities refer to the weighted sum of successors. 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 [19].

A variation of graded modal logic arises by limiting the overall (weighted) sum of successor states. If we consider the sub-functor

$$T_k X = \{f \in TX \mid \sum_{x \in X} f(x) \leq k\} \times \mathcal{P}(\mathsf{P})$$

for some $k \geq 0$, we may describe k-bounded multigraphs as T-coalgebras, and interpret the sub-language that only features the modalities $\Diamond_i$ for $i < k$.

6. For probabilistic logics, we prefer to work with subprobabilities for technical reasons, where a subprobability distribution P on a set X is a discrete measure on X with $P(X) \leq 1$. The similarity type of the *modal logic of subprobabilities*, a variant of probabilistic modal logic [12], contains, apart from propositional variables, the modal operators M_p for rational $p \in [0,1] \cap \mathbb{Q}$. This language is interpreted over T-coalgebras where TX is the set of finitely supported subprobability distributions over X, that is,

$$TX = \{\mu : X \to [0,1] \mid \textstyle\sum_{x \in X} \mu(x) \leq 1\} \times \mathcal{P}(\mathsf{P})$$

where the modalities M_p, read as "with probability of more than p", are interpreted via the liftings $\llbracket M_p \rrbracket_X(A) = \{(\mu, D) \in TX \mid \sum_{x \in A} \mu(x) > p\}$ which induces, up to the move to subprobabilities, the standard semantics.

We note that all similarity types except that of (unbounded) graded modal logic and the modal logic of subprobabilities are finite, provided that we only have finitely many propositional variables.

The aim of this work is to characterize the expressive power of $\mathcal{F}(\Lambda)$ as the fragment of first-order logic that is invariant under behavioural equivalence. The latter is best described in terms of coalgebra homomorphisms. 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 a morphism 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 $f(c) = g(d)$. Formulas of $\mathcal{F}(\Lambda)$ are invariant under behavioural equivalence:

Lemma 4. *Let C, D be T-coalgebras and let $(c, d) \in C \times D$ be behaviourally equivalent. Then $c \models \phi$ iff $d \models \phi$ for all $\phi \in \mathcal{F}(\Lambda)$.*

In other words, modal formulas are invariant under behavioural equivalence. Our main theorem extends [22] to a coalgebraic setting and establishes that all first order formulas in a suitable correspondence language with this property are in fact equivalent to modal formulas. The proof follows Rosen [18] and Otto [15], 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 [17] for a detailed exposition from the logical viewpoint). The objects of the terminal sequence are given by $T_0 = 1$ for an arbitrary one-element set and $T_n = T(T_{n-1})$, and are connected by functions $p_n : T_{n+1} \to T_n$ where $p_0 : T_1 \to 1$ is uniquely determined and $p_n = Tp_{n-1}$. Every T-coalgebra (C, γ) defines a cone over the terminal sequence by $\gamma_0 : C \to 1$ and $\gamma_n = T\gamma_{n-1} \circ \gamma : C \to T_n$. Given two T-coalgebras C and D, we can 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 5. *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 $c \approx_k d$.*

The converse is true if modal operators distinguish "enough" successor states.

Definition 6. *The Λ-structure T is separating if, for all sets X, every $t \in TX$ is uniquely determined by $\{(\heartsuit, A) \mid \heartsuit \in \Lambda \ n\text{-ary}, A \in \mathcal{P}(X)^n, t \in [\![\heartsuit]\!]_X(A)\}$.*

Separation is sufficient to establish the Hennessy-Milner property for coalgebraic modal logics [17,20], and all the structures in Example 3 are indeed separating. In particular, we obtain characteristic formulas for n-step equivalence.

Lemma 7. *If the Λ-structure T is separating and (C, γ) is a T-coalgebra, every $\approx_k$-equivalence class is definable by a formula in $\mathcal{F}_\infty(\Lambda)$ of modal rank $\leq k$.*

2   From Coalgebraic Models to First-Order Structures

The characterization of (coalgebraic) modal logic as a fragment of first-order logic stands or falls with the first-order correspondence language. In general, one needs to balance expressivity of the correspondence language against the characterization results, and the value of our results increases with the expressive power of the correspondence language. The first-order correspondence language that we use here is inspired by [11] as it is multi-sorted and includes specific sorts for states and neighbourhoods. However, it also includes a third sort for *structured successors*, i.e. elements of the set TS where S is the state set. This is, to our taste, not only the most natural first-order modelling of coalgebras, but also strengthens our main result as it increases the expressivity of the correspondence language. Even more expressivity is owed to a slightly surprising feature, whose motivation is more technical in nature: one needs expressive means for the notion of support (Definition 2), which serves the dual purpose of restricting neighbourhoods (thus keeping the logic away from full monadic second-order logic) and on the other hand to avoid vacuity of Gaifman locality (see Remark 10 for details). The following non-essential assumption simplifies the presentation.

Assumption 8. We assume that $T\emptyset$ has a distinguished element $\bot_T$, and hence that every set TX has a distinguished element $Ti(\bot_T)$, also denoted $\bot_T$, where $i : \emptyset \to X$ is inclusion. Moreover, we assume that $\bot_T \notin [\![\heartsuit]\!]_\emptyset(\emptyset, \ldots, \emptyset)$ for every k-ary operator $\heartsuit \in \Lambda$, and hence, by naturality, that $\bot_T \notin [\![\heartsuit]\!]_X(A_1, \ldots, A_k)$ for all sets X and all $A_1, \ldots, A_k \subseteq X$. This is mainly for the sake of readability, as it makes the definition of the standard translation more straightforward. In our running examples, $\bot_T$ can be taken to be

- $\bot_T = \emptyset \in \mathcal{P}(\emptyset)$ for the modal logic K (presented in terms of $\Diamond$);
- $\bot_T = \emptyset \in \mathcal{P}(\mathcal{P}(\emptyset))$ for classical and monotone modal logic;
- $\bot_T = \lambda A. \emptyset \in \mathcal{P}(\emptyset) \to \mathcal{P}(\emptyset)$ for conditional logic.
- $\bot_T = \lambda x. 0$ for graded modal logic (presented in terms of the $\Diamond_k$);
- $\bot_T = \lambda x. 0$ for the modal logic of subprobabilities.

Definition 9. The *(coalgebraic) correspondence language* associated with the modal similarity type Λ is the first order language with equality $\mathcal{L}(\Lambda)$ over three sorts s, t, n of *states*, *successor structures*, and *neighbourhoods*, respectively, consisting of the sorted relation symbols

– tr : $s \times t$ (the coalgebraic transition structure)
– $\heartsuit : t \times n \times \cdots \times n$ (k copies of n) for all k-ary $\heartsuit \in \Lambda$ (the modal operators)
– $\in : s \times n$ (membership of points in neighbourhoods)
– supp : $t \times n$ (support, see Definition 2)

The relation tr represents the transition structure, and is notionally treated as a partial map where undefined represents absence of successors. Whenever we use the term $tr(x)$ for some x, we implicitly assume that $tr(x)$ is defined. The (functional) relation supp encodes a particular choice of support for every $x : t$. Given a first-order $\mathcal{L}(\Lambda)$-structure M, we denote the constituents of M by indexing as usual; e.g. M_s is the state set of M, and M_{tr} the successor relation. We say that M is *based on* a T-coalgebra (C, γ) if the following holds.

– $M_s = C$, $M_t \subseteq TC$ and $M_n \subseteq \mathcal{P}(C)$
– The relation M_{tr} is right-unique, and hence will be written as a partial map. It represents the transition structure γ with default value $\perp_T$; i.e. for each $c \in C$, $\gamma(c) = M_{tr}(c)$ whenever $M_{tr}(c)$ is defined, and $\gamma(c) = \perp_T$ otherwise.
– The relation M_{supp} is functional, and will also be written as a map. It picks a distinguished support (Definition 2) for every $\mu \in M_t$, i.e. $\mu \in T(M_{supp}(\mu))$.
– The relations $M_\heartsuit$ represent the predicate liftings for every $\heartsuit \in \Lambda$ relative to the support, i.e. $M_\heartsuit = \{(\mu, A_1, \ldots, A_n) \in M_t \times \mathcal{P}(M_{supp}(\mu))^n \mid \mu \in [\![\heartsuit]\!]_{M_{supp}(\mu)}(A_1, \ldots, A_n)\}$ for $\heartsuit \in \Lambda$.
– $M_\in$ is membership: $M_\in = \{(s, A) \in C \times M_n \mid s \in A\}$.

We write $\mathsf{Mod}(\mathcal{L}(\Lambda))$ for the class of all $\mathcal{L}(\Lambda)$-structures that are based on some T-coalgebra, briefly referred to as T-*structures*. As every T-structure induces a uniquely defined T-coalgebra (C, γ) we occasionally regard T-structures as T-coalgebras, and in particular use notions such as behavioural equivalence. If M is a first-order structure for $\mathcal{L}(\Lambda)$ and $\phi(x) \in \mathcal{L}(\Lambda)$ is a formula with at most one free variable x of sort s, we write $M, m \models \phi(x)$ if the structure M with the free variable interpreted as m satisfies the formula $\phi(x)$, and $[\![\phi(x)]\!]_M$ is the set of all m such that $M, m \models \phi(x)$. We say that $\phi(x)$ is *invariant under behavioural equivalence* if $M, m \models \phi$ whenever $N, n \models \phi$ and $N, n \approx M, m$.

Our main interest in the present work is to establish results following van Benthem [22] and Rosen [18] which state that every first-order formula which is invariant under behavioural equivalence is equivalent to a modal formula, valid over the class of all structures and over the class of all finite structures, respectively. Occasionally we shall refer to results of the former kind as *van-Benthem-type* theorems, and to results of the latter kind as *Rosen-type* theorems.

Remark 10. Some explanations are in order concerning some aspects of the above definition. We first note that it is crucial that we do not require that

the sort n of neighbourhoods is interpreted by the entire powerset of states. Otherwise, we would essentially arrive at monadic second-order logic, and hence invalidate the main theorem already for the case of the modal logic K as the bisimulation-invariant fragment of monadic second-order logic is the μ-calculus rather than the basic modal logic K [13]. Definition 9 restricts the interpretation of n only by the clause on the interpretation of the modal operators. Technically, this makes less formulas invariant under behavioural equivalence, as the interpretation of n may differ among behaviourally equivalent models. This is the first effect of support: without support, the interpretation of $\heartsuit \in \Lambda$ would need to be defined as something like $M_\heartsuit = \{(\mu, (A_1, \ldots, A_n)) \in M_t \times \mathcal{P}(C)^n \mid \mu \in [\![\heartsuit]\!]_{r_s(\mu)}(A_1, \ldots, A_n)\}$ which would constrain the interpretation of n much more strongly, and e.g. in the case of the modal logic K (with $\heartsuit = \Diamond$) would imply $A \in M_n$ for every $A \in \mathcal{P}(C)$ containing some state that has a predecessor.

The second technical point where support is needed is the following. The core of the proof of Rosen's theorem, as adapted below, is *locality*. In particular, we use Gaifman's theorem stating that every first-order formula is essentially local (see Section 3). Without support, however, locality becomes a void notion in many logics. E.g. in the extension of classical modal logic with necessitation, i.e. with an axiom $\Box\top$, any two points in the model would be connected by a path of length 3 (via the successor structure of the first point and the neighbourhood $\top$). The formalization of support as a functional relation therefore pre-empts this trivialization. As support has already played an important technical role in other contexts [21], and is also at the heart of our unravelling construction, we are beginning to believe it may be more than just a technical nuisance.

Finally, the purpose of the default value in the above definition of T-structure is to deal with substructures that arise by cutting off transitions after a fixed number of steps. In the setting of Kripke frames, this corresponds to all successors of a node x being lost in a substructure, so that x has the empty successor set, and hence fails to satisfy diamond formulas. This notion of cutting off transitions is made explicit by our default mechanism.

The correspondence language contains standard or natural correspondence languages when applied to the running examples, as illustrated next. In most cases, the generic language is even substantially more expressive than the 'natural' correspondence language, and van Benthem/Rosen-type results become *stronger* in the context of more expressive languages, as they apply to more formulas. Some of the examples moreover highlight the importance of support.

Example 11.   1. The coalgebraic correspondence language for K differs rather substantially from the standard first-order correspondence language, the language with a unary predicate p for every propositional variable p and a binary predicate R for the transition relation — it does not explicitly talk about the transition relation, and instead has types for successor sets and neighbourhoods (one of which could be dispensed with in this case). Crucially, the standard correspondence language can be embedded into the language used here, so that our characterization result established in Section 4 does reprove, and in fact strengthen, the classical van Benthem/Rosen theorem. The embedding is defined

by mapping atomic formulas xRy in the standard correspondence language to the formula $\exists A. (\mathsf{tr}(x)\Diamond A \land \forall z. (z \in A \leftrightarrow z = y))$.

2. In the case of classical modal logic with neighbourhood frame semantics, our correspondence language has a sort s of states, a sort t of sets of neighbourhoods, and a sort n of neighbourhoods, with the successor map tr which maps states to sets of neighbourhoods, the relation $\square$ which represents the $\ni$ relation between sets of neighbourhoods and neighbourhoods, membership $\in$ between states and neighbourhoods, and additionally the support map supp from set of neighbourhoods to neighbourhoods. Superficially, this appears to be an extension of the correspondence language for neighbourhood frames used in [11], which has two sorts of states and neighbourhoods, respectively, corresponding to our sorts s and n, and two relations, corresponding to the composite $\mathsf{tr}; \square$ and to $\in$, respectively, in our setting. However, the two languages have a subtly different semantics in that in our language, the sets required to be present in the neighbourhood type are determined by the support, while in [11], the neighbourhood type contains precisely the image of the neighbourhood relation between states and neighbourhoods. At present, it is unclear whether one language can be embedded in the other so that the van Benthem-type result of [11] remains independent of our characterization theorem result below (there is no correspondent in [11] to our Rosen-type result).

We emphasize that our results will apply without further ado to extensions of classical modal logic by rank-1 frame conditions, i.e. axioms where the nesting depth of modal operators is uniformly equal to 1. A simple example is the monotonicity axiom $\square(a \land b) \to \square a$, which axiomatizes monotone neighbourhood frames (Example 3.3). It is unclear to what extent this is possible following [11]. In particular, one cannot interpret the type of neighbourhoods as the set of all sets that are a neighbourhood of some state, as the induced logic would be able to express some (if not all) μ-calculus formulas, such as $(\square\top \land \neg\square\neg\square\bot) \to \mu X. p \land \square X$: as soon as the antecedent $\square\top \land \neg\square\neg\square\bot$ is satisfied, there is a state which has $\emptyset$ as a neighbourhood, so that the neighbourhood type would contain *all* sets of states by monotonicity, thus allowing us to define the fixpoint above by quantification over neighbourhoods. As formulas of this type are invariant under behavioural equivalence but not equivalent to any modal formula, the analogue of the van Benthem/Rosen theorem fails. In the approach presented here, the notion of support enables sufficiently small interpretations of the neighbourhood type and handles rank-1 frame conditions smoothly.

3. One natural correspondence language for graded modal logic, interpreted over multigraphs (thus recovering invariance under behavioural equivalence, which fails over the relational semantics) is an extension of first order logic with counting quantifiers where the counting is relative to local weights induced by the weighting of successors in a multigraph. This induces counting quantifiers $\exists^x_{>k} y. \phi$ read 'in the local weighting at x, there exist more than k y satisfying ϕ'. This language can be mapped into our language by a recursive translation $(\_)^t$, where the clause for a counting quantifier is $(\exists^x_{>k} y. \phi)^t = (\exists A : n. \mathsf{tr}(x)\Diamond_k A \land \forall y. (\phi)^t \leftrightarrow y \in A)$. Support is uncritical in this case (as already for

item 1), as the first-order language with counting quantifiers does not contain a sort for neighbourhoods. The standard translation factors through the language of counting quantifiers via translations $(\_)^s_x$ taking $\Diamond_k \phi$ to $\exists^x_{>k} y. (\phi)^t_y$. As a consequence, the characterization results proved below apply *a forteriori* also to the language with counting quantifiers.

4. Similarly, the correspondence language for the modal logic of subprobabilities contains a sublanguage that speaks about locally determined weights of formulas: borrowing notation from Halpern's first-order logic of so-called type-1 probability structures [10] (which we extend from a single, global probability distribution to Markov chains, i.e. local (sub-)probability distributions), we may write $w^x_y(\phi) \geq p$ to denote that the set of all y satisfying ϕ has, in the local distribution at x, probability at least p.

5. The correspondence language for conditional logic contains the following more natural language, consisting of three sorts s, t, n for states, selection functions, and neighbourhoods, respectively, unary state predicates for the propositional variables, and a ternary relation R of type $s \times n \times s$ giving for each neighbourhood a transition relation on states. As neighbourhoods are explicit, we need to retain the support function supp. This corresponds to the view of a selection function model as a multi-relational Kripke model where relations are indexed over propositions, i.e. a structure of the type $(X, (R_A \subseteq X \times X)_{A \subseteq X})$, where we retain information only about those R_A for which A is in the neighbourhood type. Here, it is again crucial that the neighbourhood type is required to contain only those sets that are contained in the support of some element – without the support, neighbourhoods would be the full powerset, thus affording the full expressive power of monadic second order logic, as every selection function on a set C is contained in $[\![\Rightarrow]\!](A, C)$ for every $A \in \mathcal{P}(C)$.

The translation of modal formulas to first-order logic takes the following form:

Definition 12. The *standard translation* $\mathsf{ST}_x(\phi)$ of a modal formula $\phi \in \mathcal{F}(\Lambda)$ is a first-order formula with one free variable $x : s$ of sort s defined inductively by commutation with all boolean operators and

$$\mathsf{ST}_x(\heartsuit(\phi_1, \ldots, \phi_k)) = \exists A_1, \ldots, A_k : n. (\mathsf{tr}(x)\heartsuit(A_1, \ldots, A_k)) \wedge$$
$$\bigwedge^k_{i=1} \forall y : s. (y \in A_i \leftrightarrow y \in \mathsf{supp}(\mathsf{tr}(x)) \wedge \mathsf{ST}_y(\phi_i))).$$

The default value $\perp_T$ is compatible with the above definition: $\mathsf{ST}_x(\heartsuit(\phi_1, \ldots, \phi_k))$ is not satisfied in case $\mathsf{tr}(x)$ is undefined, which agrees precisely with the behaviour of the default value $\perp_T$ inserted as the successor structure of x in this case. Correctness is a straightforward calculation:

Lemma 13. *Suppose* $\phi = \mathsf{ST}_x(\psi)$ *for a modal formula* $\psi \in \mathcal{F}(\Lambda)$. *Let* M *be a first-order structure based on a coalgebra* C. *Then* $[\![\phi]\!]_M = [\![\psi]\!]_C$. *In particular, ϕ is invariant under behavioural equivalence.*

Remark 14. Unlike Rosen's proof [18], the original proof of van Benthem's characterization result, as well as e.g. the van-Benthem-type result proved

for neighbourhood structures in [11], rely on standard machinery from first-order logic, in particular compactness. There are at least two sources of non-compactness in the overall setup used here: one, of course, rests in the fact that we are aiming for a Rosen-type theorem over *finite models*; and the other is the functor T. Not only may T impose finite branching; e.g. in case T is the probability distribution functor, the set of formulas $\{\neg L_0 \neg p\} \cup \{M_{1-1/n}p \mid n \in \mathbb{N}\}$ for a propositional variable p is finitely satisfiable but not satisfiable, no matter what type of model (finite, infinite, finitely or infinitely branching) we consider.

3   Gaifman's Theorem and Coalgebraic Unravelling

We recall Gaifman's locality theorem [8], and derive a simple corollary that asserts locality of coproduct-invariant formulas (we claim no originality here). The basic idea is taken from [15], where the same statement is proved as Lemma 3.5 for at most binary relational structures. We apply a simpler if somewhat whole-sale argument using Gaifman locality.

Definition 15. The *Gaifman graph* of a relational structure A is the graph whose nodes are the elements of A and contains and edge from a to b iff a and b occur together in one of the tuples in the interpretations of the relation symbols in A. (E.g., the Gaifman graph of a single-relation Kripke structure is just its symmetric closure.) *Gaifman distance* is graph distance in the Gaifman graph. For $l \geq 0$, the *l-neighbourhood* $N_l^A(a)$ of $a \in A$ is the induced substructure of A containing all points with Gaifman distance at most l from a. A first-order formula $\phi(x)$ with a single free variable x is *l-local* if for every relational structure A and every $a \in A$, $A, a \models \phi(x)$ iff $N_l^A(a), a \models \phi(x)$. Moreover, $\phi(x)$ is *Gaifman l-local* if for any two points $a, b \in A$ with isomorphic l-neighbourhoods, $A, a \models \phi(x)$ iff $A, b \models \phi(x)$.

Gaifman locality is weaker than locality in that it admits statements about the global structure of models. We need a special case of Gaifman's theorem:

Theorem 16 (Gaifman [8]). *Every first-order formula $\phi(x)$ is Gaifman l-local for some $l \geq 0$, exponentially bounded in the quantifier rank of ϕ.*

Gaifman's theorem is usually formulated in single-sorted logic, but readily extends to multiple sorts, with the obvious definition of Gaifman distance, using the standard encoding of multiple sorts as unary predicates in single-sorted logic.

Definition 17. A formula $\phi(x)$ with a single free variable x is *invariant under coproducts* if for all relational structures A, B and all points $a \in A$, $A, a \models \phi(x)$ iff $A + B, a \models \phi(x)$, where $A + B$ is the coproduct (disjoint union) of A and B.

Corollary 18. *If $\phi(x)$ is invariant under coproducts, then $\phi(x)$ is l-local for some $l \geq 0$, exponentially bounded in the quantifier rank of $\phi(x)$.*

The previous does not immediately apply to our framework, as neighbourhoods in T-structures are not in general T-structures. The following lemma brings us back into the realm of T-structures, thanks to the default element $\perp_T \in T\emptyset$.

Lemma 19. *Let A be a T-structure, let a be a state in A, and let $k \geq 0$. Then $N_{3k}(a)$ is a T-structure.*

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 21). The basic notion underlying these concepts is the following.

Definition 20. *Let A be a T-structure for the correspondence language. The supporting Kripke frame of A relates states $a, b \in A_s$ iff $b \in A_{\mathsf{supp}}(A_\gamma(a))$; i.e. its transition relation is $A_\gamma; A_{\mathsf{supp}}; A_\ni$ where $A_\ni$ is the inverse relation of $A_\in$. If this Kripke frame is a tree of depth l (with root a), i.e., is loop-free and every state is reachable from a by a unique path of length at most l, and moreover all leaves of this tree have the default successor structure (i.e. they do not have an R-successor) then we say that A (or (A, a)) is a* tree of depth l.

Lemma 21. *Let A, B be T-structures with states $a \in A$, $b \in B$. If (A, a) and (B, b) are trees of depth at most l, then $A, a \approx B, b$ iff $A, a \approx_l B, b$.*

The core construction is described in the following lemma.

Lemma 22 (Unravelling). *Let A be a T-structure, let a be a state in A, and let $k \geq 0$. Then there exists a T-structure B and $b \in B$ such that $A, a \approx B, b$, and moreover $(N_{3k}^B(b), b)$ is a tree of depth at most k.*

Finally, we note that bounded behavioural equivalence is indeed local.

Lemma 23. *Let A be a T-coalgebra, let $a \in A$, and let $k \geq 0$. Then $A, a \approx_k N_{3k}^A(a), a$.*

4 A Coalgebraic van Benthem/Rosen Theorem

The core result proved in relational versions of the van Benthem/Rosen theorem is that every bisimulation-invariant formula can be expressed by a collection of modal formulas of bounded rank. In the relational case, this immediately implies equivalence to a single modal formula, as the set of modal formulas of a given maximal rank is finite up to logical equivalence. Coalgebraically, the situation turns out to be the same as long as the modal similarity type is finite. For infinite modal similarity types, the infinitary version of the van Benthem/Rosen theorem cannot be improved, as we demonstrate by means of a simple counterexample later. We thus tend to regard the infinitary version, stated next, as the most fundamental incarnation of the van Benthem/Rosen theorem. We emphasize that the bound on the rank in the statement of the theorem is the core of the result – without it, the claim is a trivial consequence of the (coalgebraic) Hennessy-Milner property for infinitary languages [17]. As in [15], the theorem below has two readings, for finite and infinite models.

Theorem 24 (Coalgebraic van Benthem/Rosen theorem, infinitary version). *Let Λ be separating. A first-order formula $\phi(x)$ over T with a single free variable x is invariant under behavioural equivalence (over finite models) iff it is equivalent (over finite models) to a modal formula in $\mathcal{F}_\infty(\Lambda)$ with finite modal rank.*

The proof uses the Lemmas established in Section 3 in sequence. As announced, the finitary version of the theorem follows immediately for finite similarity types:

Corollary 25 (Coalgebraic van Benthem/Rosen theorem, finitary version). *Let Λ be finite and separating. Then a first-order formula $\phi(x)$ over T with a single free variable x is invariant under behavioural equivalence (over finite models) iff it is equivalent (over finite models) to a modal formula in $\mathcal{F}(\Lambda)$.*

For the logics introduced in Example 3, the situation is as follows.

Example 26. Theorem 24 applies to all logics of Example 3, and Corollary 25 applies to those logics that only have finitely many modalities, i.e. all of them except (unbounded) graded modal logic and probabilistic modal logic; we note that Corollary 25 does apply to our bounded version of graded modal logic. We emphasize that Theorem 24 does yield a characterization of the behavioural-equivalence-invariant fragment of a first-order logic with counting quantifiers; while a similar van Benthem-type result is known [6], our result seems to be the first Rosen-type result (i.e. over finite structures) for graded modal logic.

As indicated above, a simple example shows that in the full correspondence language for an infinite modal similarity type, one can express properties which are invariant under behavioural equivalence but not expressible by a finitary modal formula, even in the standard coalgebraic modelling of Kripke models with infinitely many variables; in other words, the infinitary version of the van Benthem/Rosen theorem cannot be improved for infinite modal similarity types.

Example 27. Recall that standard Kripke models over the set P of variables are modelled by the functor $TX = \mathcal{P}(X) \times \mathcal{P}(\mathsf{P})$ (Example 3.1). Then the following formula is invariant under behavioural equivalence:

$$\exists y, z : s, Y, Z, A : n. \, (\forall w : s. \, ((w \in Y \leftrightarrow w = y) \wedge (w \in Z \leftrightarrow w = z) \wedge w \in A)$$
$$\wedge \, \mathsf{tr}(x)\Diamond Y \wedge \mathsf{tr}(x)\Diamond Z \wedge \mathsf{tr}(y)\neg\Diamond A \wedge \mathsf{tr}(z)\neg\Diamond A \wedge \mathsf{tr}(y) \neq \mathsf{tr}(z))$$

This formula states that x has two successors y, z which are both deadlocks but disagree on the value of at least one propositional variable. This formula is evidently not equivalent to any finitary modal formula. However, it is expressible by the infinitary modal formula $\bigvee_{p \in \mathsf{P}}(\Diamond(p \wedge \neg\Diamond\top) \wedge \Diamond(\neg p \wedge \neg\Diamond\top))$, which has modal depth 2, thus illustrating Theorem 24.

Note that proofs of the Rosen theorem in a relational setting begin with a (trivial) reduction to finitely many variables, which is possible precisely because the standard correspondence language does not allow one to say that two states agree on all propositional variables. Of course, the example above depends heavily on the use of equality on t. We state the following nagging open question:

Problem 28. Let Λ be separating. Is every formula of the correspondence language that is invariant under behavioural equivalence and does not mention equality on t equivalent to a finitary modal formula?

We note that in the case of infinite collections of independent modal operators, such as infinitely many propositional variables, or boxes for infinitely many unrelated agents, the question is answered positively by a trivial reduction to the finite case. The problematic case are infinite collections of interdependent operators as, e.g., in graded modal logic.

5 Conclusions and Related Work

We have introduced a correspondence language for coalgebraic modal logic, and proved two van Benthem/Rosen type theorems using this language: an infinitary version which applies to *every* separating coalgebraic modal logic, and states that every formula which is invariant under behavioural equivalence is equivalent to an infinitary modal formula *of bounded depth*; and, as an easy corollary to this, a finitary version which improves this to equivalence to a finitary modal formula for *finite* modal similarity types, a condition which in connection with separation implies that the type functor preserves finite sets. The infinitary result yields e.g. 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 yields characterizations of conditional logic, classical modal logic, monotone modal logic, and a bounded version of graded modal logic as the invariant fragment under behavioural equivalence in the respective correspondence languages.

It remains an open problem to extend the finitary result to infinite modal similarity types; a simple example shows that this can work only for a restricted correspondence language that excludes equality on the type of successor structures. This would in particular imply finitary van Benthem/Rosen theorems for graded and probabilistic modal logic. The former would complement a van Benthem-type result for graded modal logic proved in [6] by a Rosen-type result (i.e. over finite structures). A further interesting direction for future investigation is to extend the ambient logic, in particular 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 [13].

References

1. Barr, M.: Terminal coalgebras in well-founded set theory. Theoret. Comput. Sci. 114, 299–315 (1993)
2. Blackburn, P., de Rijke, M., Venema, Y.: Modal Logic. Cambridge University Press, Cambridge (2001)

3. Chellas, B.: Modal Logic. Cambridge University Press, Cambridge (1980)
4. D'Agostino, G., Visser, A.: Finality regained: A coalgebraic study of Scott-sets and multisets. Arch. Math. Logic 41, 267–298 (2002)
5. Dawar, A., Otto, M.: Modal characterisation theorems over special classes of frames. In: Panangaden, P. (ed.) Logic in Computer Science, LICS 2005, pp. 21–30. IEEE Computer Society, Los Alamitos (2005)
6. de Rijke, M.: A note on graded modal logic. Stud. Log. 64, 271–283 (2000)
7. Fine, K.: In so many possible worlds. Notre Dame J. Formal Logic 13, 516–520 (1972)
8. Gaifman, H.: On local and non-local properties. In: Logic Colloquium 1981, pp. 105–135. North Holland, Amsterdam (1982)
9. Gumm, H.P.: From T-coalgebras to filter structures and transition systems. In: Fiadeiro, J.L., Harman, N.A., Roggenbach, M., Rutten, J. (eds.) CALCO 2005. LNCS, vol. 3629, pp. 194–212. Springer, Heidelberg (2005)
10. Halpern, J.Y.: An analysis of first-order logics of probability. Artif. Intell. 46, 311–350 (1990)
11. Hansen, H.H., Kupke, C., Pacuit, E.: Bisimulation for neighbourhood structures. In: Mossakowski, T., Montanari, U., Haveraaen, M. (eds.) CALCO 2007. LNCS, vol. 4624, pp. 279–293. Springer, Heidelberg (2007)
12. Heifetz, A., Mongin, P.: Probabilistic logic for type spaces. Games and Economic Behavior 35, 31–53 (2001)
13. Janin, D., Walukiewicz, I.: Automata for the modal μ-calculus and related results. In: Hájek, P., Wiedermann, J. (eds.) MFCS 1995. LNCS, vol. 969, pp. 552–562. Springer, Heidelberg (1995)
14. Larsen, K., Skou, A.: Bisimulation through probabilistic testing. Inform. Comput. 94, 1–28 (1991)
15. Otto, M.: Bisimulation invariance and finite models. In: Logic Colloquium 2002. Lect. Notes Log., vol. 27, pp. 276–298. ASL (2006)
16. Pattinson, D.: Coalgebraic modal logic: Soundness, completeness and decidability of local consequence. Theoret. Comput. Sci. 309, 177–193 (2003)
17. Pattinson, D.: Expressive logics for coalgebras via terminal sequence induction. Notre Dame J. Formal Logic 45, 19–33 (2004)
18. Rosen, E.: Modal logic over finite structures. J. Logic, Language and Information 6(4), 427–439 (1997)
19. Schröder, L.: A finite model construction for coalgebraic modal logic. J. Log. Algebr. Prog. 73, 97–110 (2007)
20. Schröder, L.: Expressivity of coalgebraic modal logic: The limits and beyond. Theoret. Comput. Sci. 390, 230–247 (2008)
21. Schröder, L., Pattinson, D.: Strong completeness of coalgebraic modal logics. In: Albers, S., Marion, J.-Y. (eds.) Theoretical Aspects of Computer Science, STACS 2009, Leibniz International Proceedings in Informatics, pp. 673–684. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl (2009)
22. van Benthem, J.: Modal Correspondence Theory. PhD thesis, Department of Mathematics, University of Amsterdam (1976)