Modal and Guarded Characterisation Theorems over Finite Transition Systems *

Martin Otto †‡

revised: July 03

Abstract

We explore the finite model theory of the characterisation theorems for modal and guarded fragments of first-order logic over transition systems and relational structures of width two. A new construction of locally acyclic bisimilar covers provides a useful analogue of the well known tree-like unravellings that can be used for the purposes of finite model theory. Together with various other finitary bisimulation respecting model transformations, and Ehrenfeucht-Fraïssé game arguments, these covers allow to upgrade finite approximations for full bisimulation equivalence towards approximations for elementary equivalence. These techniques are used to prove several ramifications of the van Benthem-Rosen characterisation theorem of basic modal logic for refinements of ordinary bisimulation equivalence, both in the sense of classical and of finite model theory.

keywords: finite model theory, modal logic, guarded fragment, bisimulation, preservation and characterisation theorems

1 Introduction

Model theoretic characterisation theorems provide direct links between semantics and syntax. As assertions of the form

> a property satisfies [*a semantic condition*] if and only if it is expressible in [*a syntactic class*],

they express precise semantic-syntactic correspondences. Mostly they are relative to some common syntactic-semantic backdrop, like first-order logic, where the above becomes

> a first-order formula satisfies [*a semantic condition*] if and only if it is expressible in [*a syntactic fragment of first-order logic*].

*This paper extends an earlier conference version [15] presented under the same title at LICS 02

†Computer Science Department, University of Wales Swansea, Swansea SA2 8PP, United Kingdom, m.otto@swan.ac.uk

‡research partially supported by EPSRC grant GR/R11896/01, and by a Ryal Society European Joint Project grant

Classical model theory has many examples [5]; among them the correspondence between preservation under substructures and the universal fragment of first-order; preservation under unions of chains and the Π_2 fragment of first-order; or monotonicity in a predicate and positivity in that predicate.

Many of these correspondences do not translate into theorems of finite model theory. It is known, e.g., that there are first-order sentences which over finite structures are preserved under substructures, but are not equivalent to any universal first-order sentence over finite structures, see [7]. For another failure, which is closer to our concerns here, see also the remarks following Theorem 6 below. Note that the restriction to finite models usually implies a weakening on both sides of the desired equivalence: the semantic condition is only available over finite structures, but the syntactic form also need only apply over finite structures. Also, if a classical characterisation theorem fails as a theorem of finite model theory, it could of course still be that there is an alternative syntactic counterpart which would correspond to the semantic condition over finite models.

A nice example of a characterisation theorem that holds both classically and as a theorem of finite model theory is van Benthem's characterisation of basic modal logic. Here propositional modal logic is regarded as a fragment of first-order logic, interpreted over Kripke structures or transition systems. Over a vocabulary consisting of binary relations R (viewed as accessibility relations, transitions, or actions) and unary predicates P (coding the basic propositions, or state properties), we regard the modal operators $[R]$ and $\langle R\rangle$ as relativised first-order quantifiers according to

$$\begin{aligned} \big([R]\varphi\big)(x) &\equiv \forall y\big(Rxy \rightarrow \varphi(y)\big), \\ \big(\langle R\rangle\varphi\big)(x) &\equiv \exists y\big(Rxy \wedge \varphi(y)\big). \end{aligned}$$

The semantic condition which characterises modal logic as a fragment of first-order logic is that of bisimulation invariance. Bisimulation equivalence is important as a notion of behavioural equivalence between transition systems, or—more classically—as the notion of equivalence induced by the appropriate variant of the infinite back-and-forth Ehrenfeucht-Fraïssé game whose moves capture the relativised pattern of modal quantification.

The classical version of this characterisation theorem is the following, due to van Benthem [18, 19]. We choose a formulation that highlights the harder direction of the equivalence, namely the converse of the (easier) semantic preservation theorem.

Theorem 1 (van Benthem). *Any first-order formula $\varphi(x)$ that is invariant under bisimulation is equivalent to a formula of basic modal logic, and vice versa.*

We let FO stand for first-order logic, ML for propositional modal logic. The above characterisation theorem is then symbolically expressed as the equivalence FO$/\!\sim\,\equiv$ ML, which says that the logic ML precisely expresses those FO-properties that are invariant under bisimulation equivalence, $\sim$.

The finite model theory version of this characterisation result is not an immediate consequence of the classical version, since there are first-order formulae that are bisimulation invariant over finite structures without being bisimulation invariant over all structures. Trivial examples can be generated with the use of some infinity axiom.

Let for instance ψ be the first-order sentence that asserts that the binary relation R is a linear ordering without maximal element. Then any formula $\psi \wedge \varphi(x)$ is trivially bisimulation invariant over finite structures, but not bisimulation invariant over infinite structures if $\varphi(x)$ is satisfiable in any model of ψ.

And indeed, the classical proof of van Benthem's theorem makes use of compactness and saturation techniques that crucially involve infinite models. The characterisation itself, however, does go through in finite model theory, as shown by Rosen [17].

Theorem 2 (Rosen). *Any first-order formula $\varphi(x)$ that is invariant under bisimulation over finite structures is equivalent over finite structures to a formula of basic modal logic, and vice versa.*

While the elegant classical proof of Theorem 1 tells us nothing about the finite model theory version, the rather more constructive argument given by Rosen does apply equally to the classical version, thus providing a new proof there as well. For an alternative, quite elementary and self-contained proof of the van Benthem-Rosen theorem see [14]. In a nutshell, this proof of the van Benthem-Rosen characterisation, which will also point us in the right direction towards our present ramifications, goes as follows (also compare section 2.5 and in particular Theorem 23).

Suppose $\varphi = \varphi(x) \in \mathrm{FO}$ is bisimulation invariant. Let the quantifier rank of φ be q.

By means of analysis of the q-round Ehrenfeucht-Fraïssé game, one can show that $\varphi(x)$ must be ℓ-local around x for $\ell = 2^q - 1$; this means that whether or not φ is satisfied in $\mathfrak{A}, a$ only depends on the substructure induced on the nodes whose distance from a is at most ℓ (see section 2.4). In fact ℓ-locality even follows from just invariance under disjoint unions of transition systems, which itself is a trivial consequence of bisimulation invariance (Lemma 20).

It is a simple observation about bisimulation that φ, being invariant under bisimulation *and* ℓ-local, must then actually be invariant under ℓ-bisimulation, the level ℓ finite approximation to full bisimulation (cf. section 2.2 for bisimulation and ℓ-bisimulation). As a consequence of this, φ is finally seen to be equivalent to a modal logic formula of nesting depth $\ell = 2^q - 1$.

Contrast this with the classical proof, which essentially proceeds indirectly, deriving a contradiction based on a compactness argument. Assuming that φ is bisimulation invariant but not expressible in modal logic at any nesting depth ℓ, compactness yields models $\mathfrak{A}, a$ and $\mathfrak{B}, b$ that are indistinguishable in modal logic (i.e., ℓ-bisimilar for all finite ℓ) but with $\mathfrak{A}, a \models \varphi$ whereas $\mathfrak{B}, b \models \neg\varphi$.

Further passing to sufficiently rich elementary extensions of $\mathfrak{A}, a$ and $\mathfrak{B}, b$, respectively, one arrives at a situation in which moreover modal indistinguishability implies full bisimilarity: a contradiction, as by bisimulation invariance of φ, bisimilar structures must not be distinguished by φ.

The classical proof does not go through in the sense of finite model theory, because it relies on classical theorems and model constructions that are not available in restriction to just finite structures. The game based arguments in the alternative proof, however, go through classically as well as in restriction to finite structures. That proof is also more constructive and yields a bound on the nesting depth of the target formula, which in this case is even optimal.

For our new characterisation results we introduce similar techniques that work classically as well as in restriction to finite models.

They deal with natural refinements of ordinary bisimulation equivalence:

- two-way bisimulation (with backward as well as forward moves along edges).
- global bisimulation (with jumps to any fresh start state).
- global two-way bisimulation (both of the above).
- guarded bisimulation (free moves to overlapping or non-overlapping edges).

These lead to characterisations of more expressive modal and guarded fragments of first-order logic, as indicated in the theorems below. In their naturalness they illustrate the robustness of the close Ehrenfeucht-Fraïssé correspondence between these variants of bisimulation and modal or guarded quantification patterns. They also illustrate the unusually smooth transition between classical and finite model theory of modal logics.

While we here state the theorems as theorems of finite model theory, with the proofs given they apply equally well in the classical context. The two latter theorems can also be stated for sentences rather than for formulae in one free variable. For the precise definitions of the fragments of first-order involved, as well as for the corresponding notions of bisimulation invariance, we refer to the main part of the paper.

Theorem 3. *Any first-order formula $\varphi(x)$ that is invariant under two-way bisimulation in finite structures is equivalent over finite structures to a formula of modal logic with inverse modalities $[R]^-$, and vice versa.*

Theorem 4. *Any first-order formula $\varphi(x)$ that is invariant under global bisimulation in finite structures is equivalent over finite structures to a formula of modal logic with universal modality $(\forall)$, and vice versa.*

Theorem 5. *Any first-order formula $\varphi(x)$ that is invariant under global two-way bisimulation over finite structures is equivalent over finite structures to a formula of modal logic with inverse and universal modalities, and vice versa.*

Theorem 6. *Any first-order formula $\varphi(x)$ in a purely relational vocabulary of width 2 that is invariant under guarded bisimulation over finite structures is equivalent over finite structures to a formula of the guarded fragment of first-order logic, and vice versa.*

It should be noted that the guarded fragment, GF, over a vocabulary of width 2 can (for formulae with no more than two free variables, that is) also be embedded into the 2-variable fragment of first-order logic, FO^2. Interestingly, the classical characterisation theorem of FO^2—as the 2-pebble game invariant fragment of first-order logic—is known to fail in the context of finite model theory. Indeed, the first-order sentence (in three variables) that says of a binary relation R that it is a linear order of the universe, is invariant under 2-pebble game equivalence in restriction to finite structures (but not in general)—and it is easy to see that no first-order sentence with just two variables is equivalent to it over all finite structures. Compare also [3], and, e.g., Example 1.12 in [13].

Whether the characterisation in Theorem 6 extends to vocabularies of widths greater than 2 remains open.

The proofs of the new characterisation theorems extend the alternative proof ideas sketched for the van Benthem-Rosen theorem above. They are based on the underlying Ehrenfeucht-Fraïssé and bisimulation games, and essentially revolve about the idea of *upgrading* corresponding levels of ℓ-bisimulation to levels of approximate, local elementary equivalence that are sufficient to preserve the given first-order φ. This is achieved in model constructions that are also applicable in restriction to finite structures, and respect full bisimulation equivalence while giving local control over first-order properties by making structures locally acyclic. The following serves as a technical cornerstone in these model constructions; for a full statement and the proof compare Proposition 29 in section 3.

Theorem 7. *Every finite transition system admits, for every $k > 3$, a finite globally two-way bisimilar companion that is k-acyclic (has no cycles of lengths less than k).*

For fixed k, the increase in size can be polynomially bounded.

This provides graded analogues, in finite structures, of the well-known but generally infinite acyclic companions obtained as tree unravellings, which play an important role throughout the model theory of modal logics.

Plan of the paper Section 2 firstly reviews some basic definitions; a discussion of the specific differences between our proofs, that work for finite model theory as well as in the classical case, and the classical proof follows in section 2.3; a crucial concept in this context is that of upgrading equivalences (Definition 14); in section 2.4 we review Gaifman locality, with specific ramifications for our purposes, and derived levels of local first-order equivalence, to which we will upgrade finite bisimulation levels; Lemma 22 provides a generic road map for all our proofs of characterisation theorems; section 2.5 discusses the variant proof of the van Benthem-Rosen characterisation in the light of this approach.

The major contribution in terms of finite model constructions is presented in section 3, where the locally acyclic covers are obtained (Theorem 7). This will allow us to upgrade bisimulation equivalence to local first-order equivalence in globally bisimilar companion structures; the corresponding technical upgrading results are presented in section 4. In section 5 these are applied to prove the main modal characterisation theorems, Theorems 4 and 5. Section 6 finally extends the entire development of the previous sections to the level of guarded bisimulation invariance over finite transition systems, including the proof of Theorem 6.

2 Preliminaries and basic definitions

Structures We look at purely relational structures with only unary and binary predicates, often with one distinguished element. Our vocabulary τ will always be finite, relational and of width 2. Writing $\tau = \tau^{(1)} \cup \tau^{(2)}$ for a vocabulary, it is understood that $\tau^{(i)}$ consists of the i-ary predicates in τ.

This format is suitable for rendering transition systems. In this picture, elements of a τ-structure are the states; the unary predicates correspond to the basic propositions: a basic property q holds true of a state s if $s \in P_q$; the binary relations code transitions between states: $(s,t) \in R_i$ means that there is a transition of type i from state s to state t, or that an action i can transform state s into state t. Equally well, we may think of a τ-structure as a Kripke model, with the elements now being possible worlds and the binary relations accessibilities between worlds. Equally well, again, we may just think of edge- and vertex-coloured directed graphs.

τ-structures are represented as in $\mathfrak{A} = (A, (R^{\mathfrak{A}})_{R\in\tau^{(2)}}, (P^{\mathfrak{A}})_{P\in\tau^{(1)}})$, where typically A stands for the universe of $\mathfrak{A}$. The superscripts in the interpretations of predicates R as $R^{\mathfrak{A}}$ and P as $P^{\mathfrak{A}}$ are often dropped. Where we want to refer to a distinguished element we indicate this element explicitly as in $\mathfrak{A}, a$. Although we are mostly interested in finite model theory, all our considerations equally apply to infinite structures. We therefore adopt the convention to mention finiteness explicitly where it matters.

2.1 Some logics

We denote first-order logic as FO, elementary equivalence as $\equiv$. The quantifier rank of first-order formulae is defined as usual, and $\equiv_q$ stands for elementary equivalence up to quantifier rank q, or equivalence in the classical q-round Ehrenfeucht-Fraïssé game (see for instance [8, 7, 16]).

Basic modal logic Propositional modal logic, in its basic form which we denote ML, is based on atomic propositions q (associated with P_q), the usual boolean connectives, and the modal operators $[R]$ and $\langle R \rangle$ (associated with R). For general background we refer to the comprehensive textbook [4]. We here present the syntax in the first-order framework, so that the semantics is just the usual one for first-order.

Syntax. The formulae of ML over vocabulary $\tau = \tau^{(1)} \cup \tau^{(2)}$ are generated as follows:

– for every unary predicate P in $\tau^{(1)}$ and first-order variable x, Px is an atomic formula of ML$[\tau]$.

– if $\varphi(x)$ is a formula of ML$[\tau]$ then so is $\neg\varphi(x)$.

– if $\varphi_1(x)$ and $\varphi_2(x)$ are formulae of ML$[\tau]$ in the same free variable x, then so are $\varphi_1(x) \wedge \varphi_2(x)$ and $\varphi_1(x) \vee \varphi_2(x)$.

– for a binary relation $R \in \tau^{(2)}$, if $\varphi(y)$ is a formula of ML$[\tau]$ in the free variable y and if x is any first-order variable distinct from y, then the following are formulae of ML$[\tau]$ (where we regard the left-hand sides as abbreviations):

$$\begin{aligned} \big([R]\varphi\big)(x) &= \forall y\big(Rxy \rightarrow \varphi(y)\big), \\ \big(\langle R\rangle\varphi\big)(x) &= \exists y\big(Rxy \wedge \varphi(y)\big). \end{aligned}$$

Semantics. The semantics for ML is the usual one for first-order logic.

The guarded fragment The guarded fragment GF extends the modal quantification pattern to a more general form of relativised first-order quantification. As in modal logic, the relativisation is effected by ground atoms. Some of the power of the generalisation from ML to GF is seen only over vocabularies of width greater than 2, where ground atoms can cover more than two elements. We present the syntax in the general format of arbitrary relational vocabularies, but keep in mind that we shall only deal with GF in the setting of width 2 vocabularies where the similarity with ML is closer. GF was introduced by Andréka, van Benthem and Németi in [2], as a powerful yet tractable generalisation of ML, compare in particular also [10].

Syntax. The formulae of GF over vocabulary $\tau = \tau^{(1)} \cup \tau^{(2)}$ are generated as follows:

– all atomic τ-formulae are formulae of GF$[\tau]$.

– GF$[\tau]$ is closed under the boolean connectives $\neg$, $\wedge$ and $\vee$.

– if $\varphi(\bar{x},\bar{y})$ is a formula of GF$[\tau]$ and if $\alpha(\bar{x},\bar{y})$ is a τ-atom (also allowing equality) such that free$(\varphi) \subseteq$ var(α), then the following are formulae of GF$[\tau]$ (where we regard the left-hand sides as abbreviations):

$$\begin{aligned}(\forall\bar{y}.\alpha)\,\varphi(\bar{x},\bar{y}) &= \forall\bar{y}\big(\alpha(\bar{x},\bar{y}) \rightarrow \varphi(\bar{x},\bar{y})\big),\\ (\exists\bar{y}.\alpha)\,\varphi(\bar{x},\bar{y}) &= \exists\bar{y}\big(\alpha(\bar{x},\bar{y}) \wedge \varphi(\bar{x},\bar{y})\big).\end{aligned}$$

The atom α in the last clause is called the *guard* of the (universal or existential) quantification. It is useful to associate with a formula φ of GF a *nesting depth* of guarded quantification, which turns out to be more indicative than its plain first-order quantifier rank. The nesting depth behaves like quantifier rank on atomic formulae and with respect to boolean connectives; however, it increases by just 1 with every guarded quantification (whereas ordinary quantifier rank would go up by the length of the quantified tuple).

Semantics. The semantics for GF is the usual one for first-order logic.

Clearly ML $\subseteq$ GF. The inclusion is proper even in the case of width 2 vocabularies. In particular, GF has equality, so that, e.g., the following is in GF (but clearly not in ML):

$$\forall y(Rxy \rightarrow x = y).$$

Equality can also be used as a guard, whence GF has global universal quantification over any formula $\varphi(y)$ in a single free variable:

$$\forall y(y{=}y \rightarrow \varphi(y)).$$

In modal logics this feature is associated with a *global modality*, whose accessibility relation is the full binary relation over the universe.

In similar terms, GF has what in modal logics would correspond to *inverse modalities*, simply because the guard atoms α have no sense of direction,

$$\forall y(Ryx \rightarrow \varphi(y))$$

is a formula of GF just as $\forall y(Rxy \rightarrow \varphi(y))$ is.

Modal logic with inverse and universal modalities Common extensions of basic modal logic go some way towards capturing the two last features of GF mentioned above.

Universal modality. Extending the syntax of ML, we close under universal and existential quantification, and allow all formulae without free variables as additional constituents for the boolean operations and modal operators.

Inverse modalities. Further extending the syntax of ML, we also allow modal operators w.r.t. to the inverses of the binary relations R:

$$\begin{aligned} \big([R]^{-}\varphi\big)(x) &= \forall y\big(Ryx \rightarrow \varphi(y)\big), \\ \big(\langle R\rangle^{-}\varphi\big)(x) &= \exists y\big(Ryx \wedge \varphi(y)\big). \end{aligned}$$

Definition 8. We denote as ML, ML^{-}, $\mathrm{ML}^{\forall}$, and $\mathrm{ML}^{-\forall}$, respectively, basic modal logic and its extensions with inverse modalities, universal modality, and both.

The inclusion structure is as indicated in the following diagram. It is easy to see that all inclusion are strict, even in restriction to finite structures. Separations, from top to bottom, in terms of properties of a single binary R, and treating $\top$ as universally true propositional constant: transitivity is known not to be expressible in GF; reflexivity, $\forall x Rxx$, is in GF but not in $\mathrm{ML}^{-\forall}$; $\forall x \exists y Rxy \equiv \forall \langle R\rangle \top$ is in $\mathrm{ML}^{\forall}$ (and $\mathrm{ML}^{-\forall}$) but not expressible in either ML or ML^{-}; $\exists y Ryx \equiv \langle R\rangle^{-}\top$ is in ML^{-} (and $\mathrm{ML}^{-\forall}$) but not expressible in ML or $\mathrm{ML}^{\forall}$.

$$\begin{tikzcd}[arrows=-]
 & \mathrm{FO} \arrow[d] & \\
 & \mathrm{GF} \arrow[d] & \\
 & \mathrm{ML}^{-\forall} \arrow[dl] \arrow[dr] & \\
\mathrm{ML}^{-} \arrow[dr] & & \mathrm{ML}^{\forall} \arrow[dl] \\
 & \mathrm{ML} &
\end{tikzcd}$$

2.2 Bisimulations

Modal bisimulations Bisimulations capture notions of behavioural equivalence between transition systems. They can equivalently be presented either in terms of games or in terms of back-and-forth systems. It is instructive to think of bisimulation as the Ehrenfeucht-Fraïssé style notion of equivalence associated to modal logics.

Many variations of the basic notion of plain bisimulation equivalence have been considered. We here only deal with plain bisimulation equivalence (in which, starting from a distinguished state, one can make forward moves along transitions) and its variation involving unrestricted moves to fresh start states (cf. global modality) and backward traversal of transitions (cf. inverse modalities). The standard definitions in terms of a back-and-forth system are based on the following. A description in terms of games will be given below.

Let $Z, Z' \subseteq A \times B$, A and B sets equipped with binary relations $R^{\mathfrak{A}}$ and $R^{\mathfrak{B}}$, respectively. We say that Z' satisfies the *back-and-forth conditions* with respect to R for Z if (*forth*:) for any $(a,b) \in Z$ and any $a' \in A$ such that $(a,a') \in R^{\mathfrak{A}}$, there is some $b' \in B$ such that $(b,b') \in R^{\mathfrak{B}}$ and $(a',b') \in Z'$.

(*back*:) for any $(a,b) \in Z$ and any $b' \in B$ such that $(b,b') \in R^{\mathfrak{B}}$, there is some $a' \in B$ such that $(a,a') \in R^{\mathfrak{A}}$ and $(a',b') \in Z'$.

Z itself satisfies the back-and-forth conditions with respect to R if the above are satisfied for $Z' = Z$.

Definition 9. Let $\mathfrak{A}$ and $\mathfrak{B}$ be τ-structures, $Z \subseteq A \times B$ non-empty.

Z is a *bisimulation* between $\mathfrak{A}$ and $\mathfrak{B}$ if $(a,b) \in Z$ implies that $\mathfrak{A} \models Pa \Leftrightarrow \mathfrak{B} \models Pb$, for all unary $P \in \tau^{(1)}$, and if Z satisfies the back-and-forth conditions w.r.t. all the binary predicates R of τ. In symbols: $Z \colon \mathfrak{A} \sim \mathfrak{B}$.

Z is a *two-way bisimulation* if in addition Z satisfies the back-and-forth conditions w.r.t. the inverses R^{-1} for all binary predicates $R \in \tau^{(2)}$. In symbols: $Z \colon \mathfrak{A} \sim_{-} \mathfrak{B}$.

An (ordinary or two-way) bisimulation Z between $\mathfrak{A}$ and $\mathfrak{B}$ is a *global bisimulation*, if in addition $\pi_1(Z) = A$ and $\pi_2(Z) = B$. In symbols: $Z \colon \mathfrak{A} \sim_{\forall} \mathfrak{B}$ or $Z \colon \mathfrak{A} \approx \mathfrak{B}$, respectively.

Two structures are bisimilar in the corresponding sense, $\mathfrak{A} \sim \mathfrak{B}$, $\mathfrak{A} \sim_{-} \mathfrak{B}$, $\mathfrak{A} \sim_{\forall} \mathfrak{B}$, or $\mathfrak{A} \approx \mathfrak{B}$, if there is a corresponding bisimulation. Two structures with distinguished nodes are bisimilar in the corresponding sense, indicated as in $\mathfrak{A}, a \sim \mathfrak{B}, b$, if there is a corresponding bisimulation Z such that $(a,b) \in Z$.

It is not hard to see that the semantics of ML is invariant under bisimulation, in the sense that for all φ in ML:

$$\mathfrak{A}, a \sim \mathfrak{B}, b \qquad \Rightarrow \qquad \big(\mathfrak{A}, a \models \varphi \;\Leftrightarrow\; \mathfrak{B}, b \models \varphi\big).$$

Similar preservation properties obtain for ML^{-}, $\mathrm{ML}^{\forall}$, $\mathrm{ML}^{-\forall}$, with respect to $\sim_{-}$, $\sim_{\forall}$, and $\approx$, respectively.

Unravellings and tree models Among the most central model theoretic consequences of bisimulation invariance, is that it guarantees the existence of tree models. The well-known *tree unravelling* of a transition system yields a bisimilar companion structure that is a tree.

Let $\mathfrak{A}, a$ be a transition system of type $\tau = \tau^{(1)} \cup \tau^{(2)}$. Its tree unravelling from a, $\mathfrak{A}^*_a$, is based on the set of all finite directed paths from a in $\mathfrak{A}$, including the empty path of length 0 from a which we identify with a itself. If $\sigma = a, a_1, \ldots, a_\ell$ is a path in $\mathfrak{A}$ of length ℓ we let $\pi(\sigma) = a_\ell$ be the last vertex along this path. For $P \in \tau^{(1)}$, we put $\sigma \in P$ in $\mathfrak{A}^*_a$ iff $\pi(\sigma) \in P^{\mathfrak{A}}$. $R \in \tau^{(2)}$ is interpreted in $\mathfrak{A}^*_a$ as the set of all pairs $(\sigma, \sigma\hat{\ }b)$ where $(\pi(\sigma), b) \in R^{\mathfrak{A}}$. It is readily checked that in this way $\mathfrak{A}^*_a, \sigma \sim \mathfrak{A}, \pi(\sigma)$ for all σ, i.e., π induces a global bisimulation so that in particular $\mathfrak{A}^*_a, a \sim_{\forall} \mathfrak{A}, a$.

Similarly, for acyclic two-way bisimilar companions, one can use a *two-way unravelling*, based on the set of all undirected non-degenerate paths from a (paths that may traverse edges in either direction, excluding, however, traversals of the same edge in opposite directions in consecutive steps). The unary predicates are interpreted as above, and $R \in \tau^{(2)}$ is interpreted as the set of all pairs $(\sigma, \sigma\hat{\ }b)$ where $(\pi(\sigma), b) \in R^{\mathfrak{A}}$ and all pairs $(\sigma\hat{\ }b, \sigma)$ where $(b, \pi(\sigma)) \in R^{\mathfrak{A}}$. Then $\mathfrak{A}^*_a, \sigma \sim_{-} \mathfrak{A}, \pi(\sigma)$ and in particular $\mathfrak{A}^*_a, a \approx \mathfrak{A}, a$.

Note that in both cases, the unravelling is infinite if the original system does have (directed, respectively undirected) cycles; hence the interest in certain substitutes for full tree unravellings that provide finite companions with some measure of acyclicity that will concern us in section 3.

Guarded bisimulations Guarded bisimulations are the adequate counterparts to bisimulations in the context of GF. Quantification in GF allows direct access only to the following tuples and subsets over a τ-structure $\mathfrak{A}$.

Definition 10. Let $\mathfrak{A}$ be a τ-structure, τ relational. A subset $s \subseteq A$ is *guarded* if s is a singleton set $s = \{a\}$ for some $a \in A$, or if $s = \{a_1, \ldots, a_k\}$ where $(a_1, \ldots, a_k) \in R$ for some relation $R \in \tau$. A tuple $\bar{a}$ over $\mathfrak{A}$ is *guarded* if its components are elements of some common guarded subset.

Note that in vocabularies of width 2, guarded subsets have one or two elements; two-element guarded subsets correspond to symmetrised relational edges (or edges in the Gaifman graph, see below).

Let $Z, Z' \subseteq \mathrm{Part}(\mathfrak{A}, \mathfrak{B})$ be sets of partial (local) isomorphisms between τ-structures $\mathfrak{A}$ and $\mathfrak{B}$. We say that Z' satisfies the guarded *back-and-forth conditions* for Z if

(*forth*:) for any $p \in Z$ and any guarded subset s' of $\mathfrak{A}$, there is some $p' \in Z'$ with $\mathrm{dom}(p') = s'$ such that p and p' agree on their common domain.

(*back*:) for any $p \in Z$ and any guarded subset t' of $\mathfrak{B}$, there is some $p' \in Z'$ with $\mathrm{im}(p') = t'$ such that the inverses of p and p' agree on their common domain.

Z itself satisfies the guarded back-and-forth conditions if the above are satisfied for $Z' = Z$.

Definition 11. Let $\mathfrak{A}$ and $\mathfrak{B}$ be τ-structures, $Z \subseteq \mathrm{Part}(\mathfrak{A}, \mathfrak{B})$ a non-empty set of local isomorphisms between $\mathfrak{A}$ and $\mathfrak{B}$.
Z is a *guarded bisimulation* between $\mathfrak{A}$ and $\mathfrak{B}$, $Z \colon \mathfrak{A} \sim_{\mathrm{g}} \mathfrak{B}$, if for every $p \in Z$, the domain and image of p are guarded subsets of $\mathfrak{A}$ and $\mathfrak{B}$, respectively, and if Z satisfies the guarded back-and-forth conditions.

We write $Z \colon \mathfrak{A}, \bar{a} \sim_{\mathrm{g}} \mathfrak{B}, \bar{b}$ to indicate that $p \colon \bar{a} \mapsto \bar{b}$ for some $p \in Z$. Note that this implies that we are dealing with parameter tuples that are guarded.

The semantics of GF is invariant under guarded bisimulations. For all φ in GF:

$$\mathfrak{A}, \bar{a} \sim_{\mathrm{g}} \mathfrak{B}, \bar{b} \qquad \Rightarrow \qquad \big(\mathfrak{A}, \bar{a} \models \varphi \;\Leftrightarrow\; \mathfrak{B}, \bar{b} \models \varphi\big)$$

Similar to the modal case, there is a characterisation theorem for GF, [2].

Theorem 12 (Andréka, van Benthem, Németi). *Any first-order formula $\varphi(\bar{x})$ that is invariant under guarded bisimulation is equivalent to a formula of* GF*, and vice versa.*

Our investigations here will culminate in the proofs of Theorems 3, 4, 5 and 6. These various levels of bisimulation invariance discussed so far precisely correspond to the naturally associated syntactic fragments of first-order logic, level by level, over all but also in restriction to just finite transition systems.

$$
\begin{tikzcd}[column sep=small]
& \mathrm{FO}/{\sim_{\mathrm{g}}} \equiv \mathrm{GF} \arrow[d, dash] & \\
& \mathrm{FO}/{\approx} \equiv \mathrm{ML}^{-\forall} \arrow[dl, dash] \arrow[dr, dash] & \\
\mathrm{FO}/{\sim_{-}} \equiv \mathrm{ML}^{-} \arrow[dr, dash] & & \mathrm{FO}/{\sim_{\forall}} \equiv \mathrm{ML}^{\forall} \arrow[dl, dash] \\
& \mathrm{FO}/{\sim} \equiv \mathrm{ML} &
\end{tikzcd}
$$

Bisimulation games The above notions of modal and guarded bisimulation can naturally be captured by Ehrenfeucht-Fraïssé games. We only give a brief outline.

The ordinary *modal bisimulation game* on $\mathfrak{A}$ and $\mathfrak{B}$ is played by two players, Player I and Player II. There are two pebbles, one for each structure, which throughout the game mark one element in each structure. It is Player II's task to maintain the condition that the correspondence between the currently marked elements preserves all the unary predicates. In a play on $\mathfrak{A}, a$ and $\mathfrak{B}, b$, the pebbles are initially placed on the distinguished nodes a and b.

In each round of the game, Player I selects one of the two structures and an R-edge that goes out of the node currently pebbled in that structure, for one of the binary relations R, and moves the pebble along that edge. Player II has to match this move in the opposite structure, by moving the pebble in that structure along an R-edge (the same R) to a node such that the new correspondence again preserves all unary predicates. A player who cannot move, loses the game; otherwise, i.e., if the game continues indefinitely, Player II wins the infinite game.

It is easy to see that a bisimulation $Z\colon \mathfrak{A}, a \sim \mathfrak{B}, b$ is nothing but a formalisation of a winning strategy for Player II in the infinite bisimulation game on $\mathfrak{A}, a$ and $\mathfrak{B}, b$.

The variations for global or two-way bisimulation are obvious. The "two-way" requirement corresponds to giving Player I the option to move a pebble backwards along some R-edge, in which case Player II has to do likewise; the "global" requirement means that Player I can also choose to make a move in which the pebble may be taken to any node, not just along an edge, in which case Player II similarly may move anywhere in the opposite structure. It is not hard to see, though, that without loss of generality one may restrict this choice of making a global move to just the first round of the game, without affecting the existence of a winning strategy.

For the *guarded bisimulation game* one uses two labelled sets of pebbles, one for each structure. In each structure, these pebbles will always be placed on elements inside some guarded set, i.e., mark a guarded tuple. It is Player II's task to make sure that the correspondence between pebbled tuples always is a local isomorphism.

In each round, Player I can determine in which structure to play and also how many of the currently placed pebbles to keep fixed, and how many of the others to place—with the only constraint that the new pebble configuration must again be guarded. Player II, in the opposite structure, needs to keep fixed the pebble(s) corresponding to those that Player I kept fixed and must place pebbles corresponding to those placed by I so as to achieve a correspondence that is a local isomorphism. In our setting of relational structure of width 2, the guarded game really only needs two pebbles over each structure.

Finite approximations Beside strategies in the infinite bisimulation games one can also consider strategies in corresponding games with a fixed finite number ℓ of rounds. Player II wins any play in which ℓ rounds are completed. The ℓ round games induce finite approximations to full bisimulation equivalence, at successively refined finite levels $\ell \in \mathbb{N}$.

At level ℓ, *ℓ-bisimulation* captures the situation where the second player has a winning strategy for ℓ rounds of the respective bisimulation game. We denote these approximations by superscripts as in $\sim^{\ell}$.

In each case, the relationship between ℓ-bisimulation and bisimulation is analogous to that between ℓ-isomorphism (cf. ℓ-round classical Ehrenfeucht-Fraïssé game) and partial isomorphism (infinite Ehrenfeucht-Fraïssé game).

In terms of back-and-forth systems, an ℓ-bisimulation between $\mathfrak{A}$ and $\mathfrak{B}$ consists of a sequence $Z_0, Z_1, \ldots, Z_\ell$ of non-empty sets, where each Z_{i-1} has the back-and-forth property for Z_i. We call such systems *stratified back-and-forth systems of depth ℓ*. The obvious variations capture the finite approximations of $\sim_{-}$, $\sim_{\forall}$, $\approx$, and $\sim_{\mathrm{g}}$.

$$\begin{tikzcd}[arrows=-]
 & \sim_{\mathrm{g}} \arrow[d] & \\
 & \approx \arrow[dl] \arrow[dr] & \\
\sim_{-} \arrow[dr] & & \sim_{\forall} \arrow[dl] \\
 & \sim &
\end{tikzcd}$$

with finite approximations at level ℓ

$$\begin{tikzcd}[arrows=-]
 & \sim_{\mathrm{g}}^{\ell} \arrow[d] & \\
 & \approx^{\ell} \arrow[dl] \arrow[dr] & \\
\sim_{-}^{\ell} \arrow[dr] & & \sim_{\forall}^{\ell} \arrow[dl] \\
 & \sim^{\ell} &
\end{tikzcd}$$

The corresponding Ehrenfeucht-Fraïssé and Karp theorems are summed up in the following. Let $\leftrightharpoons$ stand for any one of the full back-and-forth Ehrenfeucht-Fraïssé style equivalences, as captured by the existence of a strategy for the second player in the infinite game or by a corresponding back-and-forth system: $\sim, \sim_{-}, \sim_{\forall}, \approx, \sim_{\mathrm{g}}$. Let $\leftrightharpoons^{\ell}$ be the corresponding ℓ-approximation, captured by strategies in the ℓ-round games or a stratified back-and-forth system of depth ℓ: $\sim^{\ell}, \sim_{-}^{\ell}, \sim_{\forall}^{\ell}, \approx^{\ell}, \sim_{\mathrm{g}}^{\ell}$.

For the logics $\mathcal{L} = \mathrm{ML}, \mathrm{ML}^{-}, \mathrm{ML}^{\forall}, \mathrm{ML}^{-\forall}, \mathrm{GF}$ let $\mathcal{L}^{\ell}$ stand for the fragment of formulae of nesting depth up to ℓ. We let $\equiv_{\mathcal{L}^{\ell}}$ stand for the logical equivalence induced by formulae in $\mathcal{L}^{\ell}$:

$$\mathfrak{A}, a \equiv_{\mathcal{L}^{\ell}} \mathfrak{B}, b \quad \text{iff} \quad \text{for all } \varphi \in \mathcal{L}^{\ell} : \ \mathfrak{A} \models \varphi[a] \ \Leftrightarrow \ \mathfrak{A} \models \varphi[a].$$

For the full (infinite game) equivalences $\leftrightharpoons$, we correspondingly look at logical equivalences induced by the infinitary variants of these logics. Let $\mathcal{L}_{\infty}$ stands for the extension of $\mathcal{L}$ that allows arbitrary (finite or infinite) conjunctions and disjunctions.

$$\mathfrak{A}, a \equiv_{\mathcal{L}_\infty} \mathfrak{B}, b \quad \text{iff} \quad \text{for all } \varphi \in \mathcal{L}_\infty : \mathfrak{A} \models \varphi[a] \Leftrightarrow \mathfrak{A} \models \varphi[a].$$

With each of the above readings for $\leftrightarrows, \leftrightarrows^\ell, \equiv_{\mathcal{L}_\infty}, \equiv_{\mathcal{L}^\ell}$ we have the following equivalences, for all structures $\mathfrak{A}, a$ and $\mathfrak{B}, b$, and all ℓ:

$$\begin{array}{lclcccc} \mathfrak{A}, a & \leftrightarrows & \mathfrak{B}, b & \Leftrightarrow & \mathfrak{A}, a & \equiv_{\mathcal{L}_\infty} & \mathfrak{B}, b \\ \mathfrak{A}, a & \leftrightarrows^\ell & \mathfrak{B}, b & \Leftrightarrow & \mathfrak{A}, a & \equiv_{\mathcal{L}^\ell} & \mathfrak{B}, b \end{array}$$

While the second equivalence is the corresponding variant of the Ehrenfeucht-Fraïssé theorem, the first equivalence corresponds to the classical theorem of Karp that associates partial isomorphism $\simeq_p$ with equivalence in $\mathcal{L}_{\infty\omega}$.

Full (finitary) $\mathcal{L}$-equivalence, $\equiv_{\mathcal{L}}$, is captured by the common refinement of the finite levels $\equiv_{\mathcal{L}^\ell}$ for all $\ell \in \mathbb{N}$. On the side of the games, let $\leftrightarrows^\omega$ stand for the equivalence induced by the existence of strategies for Player II in all bounded games of finite lengths; $\leftrightarrows^\omega = \bigcap_\ell \leftrightarrows^\ell$, the least common refinement of the $\leftrightarrows^\ell$. Note that all the equivalences $\equiv_{\mathcal{L}^\ell}$ and $\equiv_{\mathcal{L}}$ in question are coarser than elementary equivalence $\equiv$, and hence captured by first-order theories and preserved in model constructions that respect elementary equivalence. Classical model theory in particular provides for elementary extensions that are sufficiently saturated to realise all (first-order theories of) finite configurations that are not explicitly forbidden by the first-order theory of a given structure: so-called ω-saturated models. While we do not need to go into these any further it is interesting to observe the purpose these can serve in classical proofs of preservation theorems like ours. Over ω-saturated structures, $\leftrightarrows^\omega$ coincides with $\leftrightarrows$. Thus, while the infinitary levels of game equivalence—corresponding to equivalence in $\mathcal{L}_\infty$—are not in general first-order, they can be harnessed by first-order means in ω-saturated models.

Over finite structures, however, the fact that $\leftrightarrows^\omega$ coincides with $\leftrightarrows$ follows more constructively, as a consequence of a simple cardinality argument as follows. Over any two individual finite structures the sequence of refinements $\leftrightarrows^0 \supseteq \leftrightarrows^1 \supseteq \leftrightarrows^2 \supseteq \cdots$ must become stationary at some finite level ℓ, and it follows that in restriction to these two fixed structures even $\leftrightarrows^\ell$ captures $\leftrightarrows$.

As in classical Ehrenfeucht-Fraïssé analysis, one finds that over the class of structures of fixed finite relational vocabulary τ, and for each ℓ, the respective equivalence relation $\leftrightarrows^\ell$ has finite index. Furthermore, each $\leftrightarrows^\ell$ equivalence class is definable by a formula of $\mathcal{L}^\ell$, i.e., in the corresponding fragment of first-order logic at nesting depth ℓ.

2.3 Characterisation theorems and their approximations

The classical characterisation theorem, Theorem 1, as well as its variants for the other fragments and equivalences including Theorem 12 for GF, have ℓ-approximations, which establish level-by-level correspondences between invariance $\leftrightarrows^\ell$ (ℓ-bisimulations of the respective kinds) and $\mathcal{L}$-formulae of nesting depth ℓ. It should be stressed that these approximations do by no means prove the full characterisation theorems. Unlike the full characterisation theorems, their ℓ-approximations admit simple inductive proofs, in complete analogy with classical Ehrenfeucht-Fraïssé analysis. Also unlike the full characterisation theorems, the ℓ-approximations are trivially valid also in restriction to just finite structures.

This suggests the following perspective on proving the classical characterisation theorems in a manner that is potentially valid in finite model theory. We let $\leftrightarroweq$ stand for one of the bisimulation notions considered above, or indeed any other back-and-forth equivalence that has corresponding finite approximations $\leftrightarroweq^\ell$ at finite levels ℓ .

Observation 13. *Let $\leftrightarroweq$ have finite approximations $\leftrightarroweq^\ell$, $\ell \in \mathbb{N}$. Assume that each $\leftrightarroweq^\ell$ has finite index for every fixed finite relational vocabulary. Let $\mathcal{L} = \bigcup_\ell \mathcal{L}^\ell$ be a logic, each stratum $\mathcal{L}^\ell$ closed under disjunctions. Assume that each $\mathcal{L}^\ell$ is invariant under $\leftrightarroweq^\ell$ and that each $\leftrightarroweq^\ell$-class is definable by a formula of $\mathcal{L}^\ell$.*

Then the following are equivalent, both in the sense of classical model theory and of finite model theory:

(i) Every first-order formula that is $\leftrightarroweq$-invariant is invariant under $\leftrightarroweq^\ell$ for some ℓ.

(ii) Every first-order formula that is $\leftrightarroweq$-invariant is equivalent to some formula in $\mathcal{L}$.

Note that the ℓ-approximations to a characterisation theorem that links $\mathcal{L}$ to $\leftrightarroweq$-invariance directly follow from the assumptions of the observation: a property is definable by a formula of $\mathcal{L}^\ell$ if and only if it is invariant under $\leftrightarroweq^\ell$.

Proof. (ii) $\Rightarrow$ (i): φ $\leftrightarroweq$-invariant implies φ is equivalent to some $\psi \in \mathcal{L}$ by (ii); if $\psi \in \mathcal{L}^\ell$, we find that ψ, and therefore φ, is invariant under $\leftrightarroweq^\ell$.

(i) $\Rightarrow$ (ii): φ $\leftrightarroweq$-invariant implies that φ is invariant under $\leftrightarroweq^\ell$ for some ℓ by (i). Then φ is equivalent to the disjunction over the $\mathcal{L}^\ell$-formulae defining those $\leftrightarroweq^\ell$ equivalence classes whose members are models of φ. This disjunction is finite, since $\leftrightarroweq^\ell$ has finite index. $\square$

In the light of this observation, the crux of the proof of a characterisation theorem $\mathrm{FO}/{\leftrightarroweq} \equiv \mathcal{L}$—both classically and in finite model theory—lies in establishing condition (i) of Observation 13.

Classically, this condition is established indirectly using compactness and sufficiently rich (ω-saturated) models, over which $\leftrightarroweq^\omega$ (simultaneous $\leftrightarroweq^\ell$ equivalence for all finite ℓ) coincides with full $\leftrightarroweq$ equivalence. I.e., classically one relies on model constructions that allow us to upgrade $\leftrightarroweq^\omega$ to $\leftrightarroweq$ while preserving $\varphi \in \mathrm{FO}$.

Here, on the other hand, we proceed orthogonally. We now look at finitary model constructions that fully preserve $\leftrightarroweq$ (and therefore any $\leftrightarroweq$-invariant φ) and allow us to upgrade $\leftrightarroweq^\ell$, for a specific level ℓ, to some approximation $\doteq$ of elementary equivalence $\equiv$ that is strong enough to preserve φ.

Definition 14. Let $\leftrightarroweq^\ell$ and $\doteq$ be equivalence relations between τ-structures, $\leftrightarroweq$ a refinement of $\leftrightarroweq^\ell$. We say that $\leftrightarroweq^\ell$ can be upgraded to $\doteq$ modulo $\leftrightarroweq$ (in finite structures) if for any two (finite) $\mathfrak{A}, a \leftrightarroweq^\ell \mathfrak{B}, b$ there are (finite) $\hat{\mathfrak{A}}, \hat{a} \leftrightarroweq \mathfrak{A}, a$ and $\hat{\mathfrak{B}}, \hat{b} \leftrightarroweq \mathfrak{B}, b$ such that $\hat{\mathfrak{A}}, \hat{a} \doteq \hat{\mathfrak{B}}, \hat{b}$.

$$
\begin{tikzcd}
\mathfrak{A}, a \arrow[rr, dash, "\sim^{\ell}"] \arrow[d, dash, "\sim"] & & \mathfrak{B}, b \arrow[d, dash, "\sim"] \\
\hat{\mathfrak{A}}, \hat{a} \arrow[rr, dash, "\dot{\equiv}"] & & \hat{\mathfrak{B}}, \hat{b}
\end{tikzcd}
$$

With this intuition, our proofs of the crucial condition (i) in Observation 13 proceed as follows. For a given $\sim$-invariant φ, we determine a suitable approximation $\dot{\equiv}$ of full elementary equivalence such that φ is preserved under $\dot{\equiv}$ for essentially syntactic reasons, and a finite level ℓ such that $\sim^{\ell}$ can be upgraded to $\dot{\equiv}$ (in finite models) modulo $\sim$. This implies that φ is $\sim^{\ell}$-invariant, straight from the diagram.

The appropriate levels of $\dot{\equiv}$ for this argument are obtained from a Gaifman representation of the given FO-formula φ; the relevant ℓ will essentially be the *locality rank* of φ in the sense of Gaifman's locality theorem (see below).

For all ramified cases of modal characterisation theorems, i.e., all cases with the exception of the van Benthem-Rosen theorem itself, the actual upgrading result will revolve around combinatorial constructions of certain 'nice' $\sim$-equivalent finite companion structures, over which FO can locally be controlled. These will be provided in section 3 in the form of *locally acyclic covers*.

2.4 Locality

Recall that the Gaifman graph $G(\mathfrak{A})$ of a relational structure $\mathfrak{A} = (A, \ldots)$ is the symmetric graph with universe A and edges linking any two distinct elements of A that occur together in a common ground atom of a relation in $\mathfrak{A}$. *Gaifman distance* d on $\mathfrak{A}$ is the metric induced by ordinary graph distance in $G(\mathfrak{A})$.

Definition 15. Let $\mathfrak{A}$ be a relational structure.

(a) The *neighbourhood of radius* ℓ about a in $\mathfrak{A}$ is the subset $N^{\ell}(a) = \{a' \in A \colon d(a, a') \leqslant \ell\}$.

(b) A set of elements in $\mathfrak{A}$ is *ℓ-scattered* if the mutual distance between any two distinct members of the set is greater than 2ℓ.

The ℓ-neighbourhoods of any two distinct members of an ℓ-scattered set are disjoint.

Gaifman distance d is first-order definable, for every fixed finite relational signature, in the sense that for every ℓ there is a first-order formula expressing that $d(x, y) \leqslant \ell$.

Definition 16. (i) A formula $\psi(x)$ is *ℓ-local* if it is logically equivalent to its relativisation to $N^{\ell}(x)$.

(ii) A *basic ℓ-local sentence* is one that asserts the existence of an ℓ-scattered set of m elements x all of which satisfy the same ℓ-local formula $\psi(x)$, for some m and ψ.

(iii) A first-order formula $\varphi(x)$ is in *Gaifman form* if it is a boolean combination of local formulae and basic local sentences.

(iv) The *locality rank* of a formula in Gaifman form is the minimal ℓ such that all its constituent local formulae (including those occurring in basic local sentences) are ℓ-local. Its *local quantifier rank* is the maximal quantifier rank in any of the constituent local formulae; its *scattering rank* is the maximal size of a scattered set asserted in any of its constituent basic local sentences.

(v) A *simple local sentence* is a sentence in Gaifman form of scattering rank 1; i.e., one that asserts the existence of an element x satisfying some local formula $\psi(x)$.

Note that, due to definability of Gaifman distance, ℓ-local formulae $\varphi(x)$ are equivalent to formulae whose quantifiers are explicitly relativised to the ℓ-neighbourhood of x; and simple ℓ-local sentences are equivalent to existentially quantified formulae of this kind.

Gaifman's theorem [9] says that first-order logic is essentially local.[1]

Theorem 17 (Gaifman). *Any first-order formula $\varphi(x)$ is equivalent to one in Gaifman form.*

We shall mostly apply Gaifman's theorem in more specific circumstances. The crucial specialisations for our purposes deal with formulae whose semantics is invariant under disjoint sums (unions). We write $\mathfrak{A} + \mathfrak{B}$ for the disjoint sum of two structures $\mathfrak{A}$ and $\mathfrak{B}$ over the same purely relational vocabulary τ. We also write $q \cdot \mathfrak{A}$ for the q-fold direct sum of $\mathfrak{A}$ with itself. We usually regard $\mathfrak{A}$ itself as a substructure of $\mathfrak{A} + \mathfrak{B}$ and $q \cdot \mathfrak{A}$, and use notation like $\mathfrak{A} + \mathfrak{B}, \bar{a}$ for parameters $\bar{a}$ from $\mathfrak{A}$. Note that $\mathfrak{A} + \mathfrak{B} \sim \mathfrak{A}$, and $q \cdot \mathfrak{A} \approx \mathfrak{A}$ as well as $q \cdot \mathfrak{A} \sim_{\mathrm{g}} \mathfrak{A}$ (also with parameters from $\mathfrak{A}$).

Definition 18. A τ-formula φ is

(i) *invariant under disjoint sums* (meaning: with arbitrary other τ-structures) if for all $\mathfrak{A}, \bar{a}$ and $\mathfrak{B}$: $\quad \mathfrak{A}, \bar{a} \models \varphi \;\Leftrightarrow\; \mathfrak{A} + \mathfrak{B}, \bar{a} \models \varphi$.

(ii) *invariant under disjoint copies* (meaning: of the same τ-structure) if for all $\mathfrak{A}, \bar{a}$ and $q \geqslant 1$: $\quad \mathfrak{A}, \bar{a} \models \varphi \;\Leftrightarrow\; q \cdot \mathfrak{A}, \bar{a} \models \varphi$.

Clearly, bisimulation invariance implies invariance under disjoint sums, while invariance under global or guarded bisimulation implies invariance under disjoint copies. The following specialisations of Gaifman's theorem can therefore be brought into play.

In the context of classical model theory and with a view to combinatorial applications, Compton [6] has—independently of Gaifman [9]— obtained several closely related results, which interestingly moreover also anticipate the idea of guarded quantification.

Proposition 19. *Both in the sense of classical and finite model theory:*

(a) If $\varphi = \varphi(x) \in \mathrm{FO}$ is invariant under disjoint sums, then $\varphi(x)$ is local about x.

(b) If $\varphi = \varphi(x) \in \mathrm{FO}$ is invariant under disjoint copies, then φ is equivalent to a boolean combination of local formulae about x and simple local sentences.

Proof. We explicitly prove these statements in their reading for finite model theory. The classical case follows exactly the same lines. We use $\psi_1 \models_{\mathrm{fin}} \psi_2$ or $\psi_1 \equiv_{\mathrm{fin}} \psi_2$ to explicitly

[1] We state the special case for formulae in one free variable. The theorem holds for formulae in arbitrary free variables, but one has to admit slightly more general local formulae in the free variables. These are required to be equivalent to their relativisation to the union of ℓ-neighbourhoods around all their free variables, for some ℓ.

indicate the restriction to just finite models, of semantic implication and bi-implication between formulae.

(a) Assume that over finite structures, $\varphi(x)$ is invariant under disjoint sums. According to Gaifman's theorem, φ is equivalent (over all structures) to a formula of the form

$$\varphi(x) \equiv \bigvee_i \left(\varphi_0^i(x) \wedge \chi^i\right),$$

where the $\varphi_0^i(x)$ are ℓ_i-local about x and the χ^i are boolean combinations of basic local sentences such that without loss of generality

(i) every $\varphi_0^i(x) \wedge \chi^i$ is satisfiable.

(ii) any two distinct $\varphi_0^i(x)$ are mutually exclusive.

We may delete any disjuncts $\varphi_0^i(x) \wedge \chi^i$ that have no finite models, and still retain a formula that is equivalent to φ over all finite structures, where even

(i') every $\varphi_0^i(x) \wedge \chi^i$ is satisfiable in a finite model.

We claim that then necessarily

$$\varphi(x) \equiv_{\mathrm{fin}} \bigvee_i \varphi_0^i(x).$$

Clearly $\varphi(x) \models_{\mathrm{fin}} \bigvee_i \varphi_0^i(x)$. Conversely we show that also $\bigvee_i \varphi_0^i(x) \models_{\mathrm{fin}} \varphi(x)$.

Let to this end $\mathfrak{A}, a \models \bigvee_i \varphi_0^i(x)$ be a finite model of $\bigvee_i \varphi_0^i(x)$. From (i') we obtain finite models $\mathfrak{B}_i, b_i \models \varphi_0^i(x) \wedge \chi^i$. Note that this implies $\mathfrak{B}_i, b_i \models \varphi(x)$. Let $\mathfrak{B}$ be the disjoint union of the $\mathfrak{B}_i$ and $\mathfrak{A}$. From invariance under disjoint sums we get $\mathfrak{B}, b_i \models \varphi(x)$ for each i. As $\mathfrak{B}, b_i \models \neg\varphi_0^j(x)$ for all $j \neq i$ by (ii), inspection of $\varphi(x)$ shows that necessarily $\mathfrak{B}, b_i \models \varphi_0^i(x) \wedge \chi^i$ for each i. Therefore $\mathfrak{B} \models \bigwedge_i \chi^i$. So $\mathfrak{B}, a \models \bigvee_i \varphi_0^i(x) \wedge \bigwedge_i \chi^i$. The latter formula clearly implies $\varphi(x)$. So $\mathfrak{B}, a \models \varphi(x)$, and using invariance under disjoint sums again, also $\mathfrak{A}, a \models \varphi(x)$.

(b) Let $\varphi(x)$ be invariant under disjoint copies over finite structures. Using Gaifman's theorem we obtain a presentation of $\varphi(x)$ of the following form:

$$\varphi(x) \equiv_{\mathrm{fin}} \bigvee_i \Big(\psi^i \wedge \bigvee_j \left(\varphi_0^{ij}(x) \wedge \chi^{ij}\right)\Big)$$

where

(i) the formulae $\varphi_0^{ij}(x)$ are local about x.

(ii) the sentences χ^{ij} are boolean combinations of basic local sentences talking about scattered sets of size greater than 1.

(iii) the sentences ψ^i are complete boolean combinations of all formulae of the form $\exists y \rho(y)$ where ρ ranges over all the local formulae that occur in any of the ψ^i or χ^{ij}.

(iv) Any two ψ^i and $\psi^{i'}$ are mutually exclusive for $i \neq i'$.

(v) For every i: $\varphi_0^{ij}(x)$ and $\varphi_0^{ij'}(x)$ are mutually exclusive whenever $j \neq j'$.

(vi) For every i, j: $\psi^i \wedge (\varphi_0^{ij}(x) \wedge \chi^{ij})$ is satisfiable in a finite model.

We consider $\varphi^i(x) := \bigvee_j \big(\varphi_0^{ij}(x) \wedge \chi^{ij}\big)$ in restriction to models of ψ^i, and claim that

$$\psi^i \wedge \varphi^i(x) \equiv_{\mathrm{fin}} \psi^i \wedge \bigvee_j \varphi_0^{ij}(x).$$

Clearly this proves the claim of part (b). Moreover, the left-hand side clearly implies the right-hand side. For the converse implication let $\mathfrak{A}, a \models \psi^i \wedge \bigvee_j \varphi_0^{ij}(x)$ be a finite model of the right-hand formula. We need to show that $\mathfrak{A}, a \models \varphi^i$. Choose $q \in \mathbb{N}$ to be greater than the cardinality of any scattered set mentioned in the χ^{ij}. By (vi) we find finite models $\mathfrak{B}_j, b_j \models \psi^i \wedge \varphi_0^{ij}(x) \wedge \chi^{ij}$. As $\mathfrak{B}_j, b_j \models \varphi(x)$ and by invariance of φ under disjoint copies, we have $q \cdot \mathfrak{B}_j, b_j \models \varphi(x)$. Clearly still $q \cdot \mathfrak{B}_j, b_j \models \psi^i$. Therefore, $q \cdot \mathfrak{B}_j, b_j \models \varphi^i(x)$ and as clearly also still $q \cdot \mathfrak{B}_j, b_j \models \varphi_0^{ij}(x)$, condition (v) implies that $q \cdot \mathfrak{B}_j, b_j \models \chi^{ij}$. It follows that χ^{ij} can only make positive existential claims about large scattered sets whose members satisfy some local formula that is realised according to ψ^i; conversely, any negative statement in the χ^{ij} can only forbid (large) scattered sets satisfying some local formula that according to ψ^i cannot be realised at all.

It follows that also $q \cdot \mathfrak{A} \models \chi^{ij}$, and—as this is the case for all choices of j—in fact $q \cdot \mathfrak{A} \models \bigwedge_j \chi^{ij}$. But then $q \cdot \mathfrak{A}, a \models \psi^i \wedge \bigvee_j \varphi_0^{ij} \wedge \bigwedge_j \chi^{ij}$. Therefore $q \cdot \mathfrak{A}, a \models \varphi$ and, by invariance of φ, also $\mathfrak{A}, a \models \varphi$. As the ψ^i are mutually exclusive, (iv), it must be that $\mathfrak{A}, a \models \varphi^i(x)$, as desired. ☐

In fact one can improve on part (a) of Proposition 19 by giving a quantitative bound on the locality rank ℓ. A proof of the following lemma, based entirely on an elementary Ehrenfeucht-Fraïssé game argument without appeal to Gaifman's theorem, is presented in [14]. This argument relies on an analysis of the q-round game on structures $q\cdot\mathfrak{A}+q\cdot\mathfrak{B}+\mathfrak{A}, a$ versus $q\cdot\mathfrak{A}+q\cdot\mathfrak{B}+\mathfrak{B}, b$ in the situation where $\mathfrak{A}{\restriction}N^\ell(a), a \equiv_q \mathfrak{B}{\restriction}N^\ell(b), b$, for $\ell = 2^q - 1$. Exhibiting a strategy for the second player is actually a nice exercise in Ehrenfeucht-Fraïssé games.

Lemma 20. *Both classically and in the sense of finite model theory: a first-order formula $\varphi(x)$ of quantifier rank q that is invariant under disjoint sums is ℓ-local for $\ell = 2^q - 1$.*

The tightness of this bound is illustrated by the following example. There are straightforward FO formalisations of the bisimulation invariant property that "there is a red node within distance $2^q - 1$ of x" in quantifier rank q. But any modal formula to this effect must have modal quantifier rank $2^q - 1$, since modal formulae of quantifier rank ℓ are ℓ-local.

All our proofs of characterisation theorems will establish the crucial condition in Observation 13 with an argument about upgrading ℓ-bisimulation invariance to a level of local first-order equivalence that is strong enough to preserve the given formula φ in its Gaifman form. The following lemma serves to encapsulate the generic pattern of these proofs. For the definition of the relevant levels of FO-equivalence compare Definition 16, especially item (iv). For the notion of upgrading recall Definition 14.

Definition 21. For τ-structures with distinguished parameters $\mathfrak{A}, a$ and $\mathfrak{B}, b$, and $\ell, q, n \in \mathbb{N}$: $\mathfrak{A}, a \equiv_{q,n}^{(\ell)} \mathfrak{B}, b$ if for every $k \leqslant \ell$, for every k-local formula $\psi(x)$ of quantifier rank q, and for every $m \leqslant n$:

(i) $\mathfrak{A} \models \psi[a] \Leftrightarrow \mathfrak{B} \models \psi[b]$.

(ii) $\mathfrak{A}$ has a k-scattered subset of size m for ψ iff $\mathfrak{B}$ has.

Note that $\equiv^{(\ell)}_{q,n}$ has finite index and that any FO-formula $\varphi(x)$ in Gaifman form is invariant under $\equiv^{(\ell)}_{q,n}$ if its locality rank, local quantifier rank and scattering rank are bounded by ℓ, q and n, respectively.

Lemma 22. *Both classically and in the sense of finite model theory. Let $\varphi(x)$ be in Gaifman form of locality rank ℓ, local quantifier rank q and scattering rank n. Suppose that φ is invariant under $\leftrightharpoons$. If $\leftrightharpoons^{\ell}$ can be upgraded to $\equiv^{(\ell)}_{q,n}$ modulo $\leftrightharpoons$, then φ is invariant under $\leftrightharpoons^{\ell}$.*

For this compare Observation 13 and Definition 14, and the discussion in section 2.3. The use of Proposition 19 in this context merely is to give a natural a priori bound on the locality and scattering ranks of bisimulation invariant formulae. [2]

2.5 The case of basic modal logic revisited

A high-level sketch of an alternative proof of the van Benthem-Rosen theorem, which also yields an exponential bound on the nesting depth of the target ML formula, was indicated in the introduction. We are now in a position to make this argument precise, and it may serve as an instructive, particularly simple application of the generic proof idea in Observation 13 and of upgrading. The particular simplification derives from the tight locality guaranteed by Lemma 20.

Theorem 23. *Both classically and in the sense of finite model theory: any first-order formula $\varphi(x)$ of quantifier rank q that is invariant under bisimulation is equivalent to a formula of basic modal logic whose modal nesting depth is less than 2^q.*

As remarked above, the bound on the nesting depth is tight; the example given right after Lemma 20 above, illustrates the fact that FO can be exponentially more succinct than ML for bisimulation invariant properties.

We apply a version of Lemma 22 where the target equivalence $\equiv^{(\ell)}_{q,n}$ is replaced by the following equivalence $\sim^{(\ell)}$. Define $\mathfrak{A}, a \sim^{(\ell)} \mathfrak{B}, b$ as $\mathfrak{A}{\restriction}N^{\ell}(a), a \sim \mathfrak{B}{\restriction}N^{\ell}(b), b$. Note that this is a local version of full bisimulation equivalence, not to be confused with ℓ-bisimulation. To prove Theorem 23 we show the following.

Lemma 24. *Modulo $\sim$, $\sim^{\ell}$ can be upgraded to $\sim^{(\ell)}$ (also in finite structures).*

$$\begin{tikzcd}
\mathfrak{A}, a \arrow[r, dash, "\sim^{\ell}"] \arrow[d, dash, "\sim"] & \mathfrak{B}, b \arrow[d, dash, "\sim"] \\
\hat{\mathfrak{A}}, \hat{a} \arrow[r, dash, "\sim^{(\ell)}"] & \hat{\mathfrak{B}}, \hat{b}
\end{tikzcd}$$

[2] It turns out that for the ramified characterisation results, concerning global forms of bisimulation, one does not actually have to appeal to Proposition 19 (b), if one uses simple disjoint copies in an additional step that further upgrades from $\equiv^{(\ell)}_{q,1}$ (scattering rank 1, corresponding to simple local sentences) to $\equiv^{(\ell)}_{q,n}$ (arbitrary scattering rank n); compare Lemma 37 in section 4.

Proof. Let $\mathfrak{A}, a \sim^\ell \mathfrak{B}, b$. Let $\hat{\mathfrak{A}}$ be the result of unravelling $\mathfrak{A}$ from a, restricting the resulting tree to depth ℓ, and identifying leaf nodes in this truncated tree with corresponding nodes in disjoint isomorphic copies of $\mathfrak{A}$. Let $\hat{\mathfrak{A}}_\ell = \hat{\mathfrak{A}} \restriction N^\ell(a)$ be the truncated unravelling of depth ℓ of $\mathfrak{A}, a$ with no attachments to the cut-off points at the leaves. Clearly $\hat{\mathfrak{A}}$ and $\hat{\mathfrak{A}}_\ell$ are finite if $\mathfrak{A}$ is finite. By construction, $\hat{\mathfrak{A}}, a \sim \mathfrak{A}, a$ and $\hat{\mathfrak{A}}_\ell, a \sim^{(\ell)} \hat{\mathfrak{A}}, a$.

Let $\hat{\mathfrak{B}}$ and $\hat{\mathfrak{B}}_\ell$ be similarly obtained from $\mathfrak{B}, b$. Then $\hat{\mathfrak{A}}_\ell \sim \hat{\mathfrak{B}}_\ell$, as $\mathfrak{A}, a \sim^\ell \mathfrak{B}, b$ and as both $\hat{\mathfrak{A}}_\ell$ and $\hat{\mathfrak{B}}_\ell$ are trees of depth ℓ. It follows that $\hat{\mathfrak{A}}, a \sim^{(\ell)} \hat{\mathfrak{B}}, b$, as desired. □

Proof of Theorem 23. Lemma 24 and Observation 13 now prove the theorem. If $\varphi(x)$ of quantifier rank q is bisimulation invariant, it is also ℓ-local for $\ell = 2^q - 1$ by Lemma 20, and hence invariant under $\sim^{(\ell)}$. Upgrading according to Lemma 24, as indicated in the diagram, proves that φ is invariant under $\sim^\ell$, hence expressible in ML at modal nesting depth ℓ. □

In connection with Lemma 24, it should be pointed out that $\sim^\ell$ can in fact be upgraded (modulo $\sim$ and also in finite models) to local isomorphism $\simeq^{(\ell)}$, according to $\mathfrak{A}, a \simeq^{(\ell)} \mathfrak{B}, b$ iff $\mathfrak{A} \restriction N^\ell(a), a \simeq \mathfrak{B} \restriction N^\ell(b), b$. To achieve this, one enriches finite bisimilar companions $\hat{\mathfrak{A}}$ and $\hat{\mathfrak{B}}$ from the above proof with sufficiently many copies of each 'sub-tree' to boost both structures to have equal numbers of realisers for each $(\ell - \ell' - 1)$-bisimulation type adjacent to any node at depth $\ell' < \ell$. This well-known construction is also used as part of Rosen's proof in [17], which then proceeds to upgrade to full elementary equivalence, using Hanf's theorem and some more intricate surgery on finite structures.

The van Benthem-Rosen theorem can easily be adapted to cover the case of two-way bisimulation $\sim^-$ and ML^-. For the above arguments, this involves the following observations and slight modifications. Clearly the locality results of Proposition 19 or Lemma 20 go through, as $\sim^-$-invariance also implies invariance under disjoint sums. For the analogue of Lemma 24, one adapts the proof given in Lemma 24 by using (truncated) two-way unravellings. We then obtain the following.

Corollary 25. *Both classically and in the sense of finite model theory: any first-order formula $\varphi(x)$ of quantifier rank q that is invariant under two-way bisimulation is equivalent to a formula of ML^- whose modal nesting depth is less than 2^q.*

It is apparent from the above that (ordinary as opposed to global) bisimulation invariance implies a very strong form of locality; namely, locality about the distinguished parameter. The picture is quite different, however, when we consider global (and possibly two-way) bisimulation, which does take into account the local behaviour not just around the distinguished parameters but also around any other point. Model constructions that are to respect any form of global bisimulation equivalence therefore have to be much more uniform. Partial or truncated unravellings are not good enough. The distinguishing feature of tree-like unravellings is their acyclicity (acyclicity of the underlying Gaifman graphs). But clearly acyclicity cannot be had in *finite* bisimilar companion structures of any structure that is not itself already acyclic. To the extent that one is still only concerned about the local behaviour in neighbourhoods of some bounded radius, however, it makes sense to approximate acyclicity uniformly but locally by avoiding *short cycles* in the Gaifman graph, i.e., to at least keep small local neighbourhoods acyclic. This is exactly what the locally acyclic bisimilar covers to be discussed in the following section achieve.

3 Locally acyclic bisimilar covers

Recall that $\tau = \tau^{(1)} \cup \tau^{(2)}$ is a finite, purely relational vocabulary of width 2, with binary relations $R \in \tau^{(2)}$. We denote by E the combined edge relation

$$E = \bigcup_{R \in \tau^{(2)}} R.$$

The edge relation in the Gaifman graph $G(\mathfrak{A})$ is the symmetric and irreflexive version of E.

Definition 26. Let $\mathfrak{A}$ be a τ-structure, $G(\mathfrak{A})$ its Gaifman graph.

(i) A *cycle* (of length ℓ) in $\mathfrak{A}$ is an ℓ-cycle in $G(\mathfrak{A})$ in the graph theoretic sense: a sequence of vertices $a_0, \dots, a_{\ell-1}$, where for each consecutive pair of indices $(i, i+1)$ (cyclically understood in the sense of $\mathbb{Z}_\ell$) we have $(a_i, a_{i+1}) \in R$ or $(a_{i+1}, a_i) \in R$ for some $R \in \tau^{(2)}$. A cycle of length 1 is a *loop*.

(ii) A cycle is *non-degenerate* if always $a_{i-1} \neq a_{i+1}$.

(iii) $\mathfrak{A}$ is *acyclic* if it is loop-free and has non-degenerate cycles.

(iv) $\mathfrak{A}$ is *k-acyclic* if it is loop-free and has no non-degenerate cycles of lengths $< k$.

Note that a k-acyclic structure is locally acyclic in the sense that the substructures induced on ℓ-neighbourhoods of its elements are acyclic if $k \geqslant 2\ell + 2$. In graph theoretic terms, (iv) may be rephrased by saying that the girth of $G(\mathfrak{A})$ is at least k.

All structures are 3-acyclic. Degenerate cycles cannot be avoided at all, as every edge gives rise to a degenerate cycle of length 2. In order to capture all other degeneracies in the presence of several directed edge relations R we introduce the following notion of a simple transition system.

Definition 27. A τ-structure $\mathfrak{A}$ is *simple* if the $R^{\mathfrak{A}}$ are mutually disjoint for $R \in \tau^{(2)}$ and if their union $E^{\mathfrak{A}}$ is anti-symmetric and irreflexive.

In graph theoretic terms one might consider a simple structure as an edge-partitioned and vertex-coloured tournament.

A special and very natural kind of bisimulations—familiar, e.g., form the bisimilar companion structures obtained as unravellings—are those induced by homomorphisms.

Definition 28. (a) A homomorphism $\pi \colon \hat{\mathfrak{A}} \to \mathfrak{A}$ is a *bisimilar cover* of $\mathfrak{A}$ by $\hat{\mathfrak{A}}$ if $Z_\pi = \big\{(\hat{a}, a) \colon a = \pi(\hat{a})\big\}$ is a global two-way bisimulation between $\hat{\mathfrak{A}}$ and $\mathfrak{A}$.

(b) A bisimilar cover $\pi \colon \hat{\mathfrak{A}} \to \mathfrak{A}$ is called *faithful* if, for every $\hat{a}$ and each $R \in \tau^{(2)}$, π restricts to a bijection between $\{\hat{a}' \in \hat{A} \colon (\hat{a}, \hat{a}') \in R^{\hat{\mathfrak{A}}}\}$ and $\{a' \in A \colon (a, a') \in R^{\mathfrak{A}}\}$, as well as between $\{\hat{a}' \in \hat{A} \colon (\hat{a}', \hat{a}) \in R^{\hat{\mathfrak{A}}}\}$ and $\{a' \in A \colon (a', a) \in R^{\mathfrak{A}}\}$.

Consider the example of faithful bisimilar covers obtained from two-way unravellings of transition systems. Suppose without loss of generality that $\mathfrak{A}$ is simple and connected (each connected component may be considered separately). The two-way unravelling of $\mathfrak{A}$ from some element a of $\mathfrak{A}$, as discussed in section 2.2, together with the natural projection that maps an undirected path $\sigma = a, a_1, \dots, a_\ell$ to its last element $\pi(\sigma) = a_\ell$, provides a faithful bisimilar cover of $\mathfrak{A}$ by a simple acyclic transition system, albeit generally an infinite one. As pointed out above, no cyclic finite $\mathfrak{A}$ can have a finite acyclic bisimilar companion. Our aim in this section is the following.

Proposition 29. *Every finite transition system $\mathfrak{A}$ admits, for every $k > 3$, a faithful bisimilar cover $\pi \colon \hat{\mathfrak{A}} \to \mathfrak{A}$ by a finite k-acyclic simple transition system $\hat{\mathfrak{A}}$. For fixed k, the size of $\hat{\mathfrak{A}}$ can be polynomially bounded in terms of the size of $\mathfrak{A}$ ([3]).*

Before we give a proof of the proposition, note that truncated tree-like unravellings with branches linked back into initial segments of the tree-like unravelling do not in general give rise to locally acyclic covers because acyclicity is understood in terms of undirected cycles (cycles in $G(\mathfrak{A})$) rather than directed cycles.

The case of simple transition systems The assumption of simplicity simplifies the proof of the proposition. The general case will then been reduced to this case.

Let $\mathfrak{A}$ be simple, $k \in \mathbb{N}$. Suppose $(G, \circ)$ is a finite group into which E can be embedded as $\mathbf{g} \colon E \to G$; $e \mapsto g_e$, such that $\{g_e \colon e \in E\} \cap \{g_e^{-1} \colon e \in E\} = \emptyset$.

Recall that the Cayley graph associated with $G, (g_e)_{e \in E}$ is the undirected graph with vertex set G and edges $\{h, h'\}$ exactly between those h and h' for which $h' = h \circ g_e$ or $h = h' \circ g_e$ for some $e \in E$.

With $\mathfrak{A}$ and $G, \mathbf{g}$ we associate the following structure $\mathfrak{A} \otimes_{\mathbf{g}} G$ with universe $A \times G$. Unary predicates $P \in \tau^{(1)}$ are interpreted in $\mathfrak{A} \otimes_{\mathbf{g}} G$ as $\pi^{-1}(P^{\mathfrak{A}})$ where $\pi \colon A \times G \to A$ is the natural projection. For the binary predicates $R \in \tau^{(2)}$ we put an R-edge from (a, h) to (a', h') if and only if $e = (a, a') \in R^{\mathfrak{A}}$ and $h' = h \circ g_e$.

$$\begin{array}{rcl}
\mathfrak{A} \otimes_{\mathbf{g}} G & = & \big(A \times G, (P^{\mathfrak{A} \otimes_{\mathbf{g}} G})_{P \in \tau^{(1)}}, (R^{\mathfrak{A} \otimes_{\mathbf{g}} G})_{R \in \tau^{(2)}}\big), \\
P^{\mathfrak{A} \otimes_{\mathbf{g}} G} & = & \pi^{-1}(P^{\mathfrak{A}}), \\
R^{\mathfrak{A} \otimes_{\mathbf{g}} G} & = & \big\{((a, h), (a', h \circ g_e)) \colon (a, a') = e \in R^{\mathfrak{A}}\big\}.
\end{array}$$

This clearly turns $\pi \colon \mathfrak{A} \otimes_{\mathbf{g}} G \to \mathfrak{A}$ into a faithful bisimilar cover. $\mathfrak{A} \otimes_{\mathbf{g}} G$ is also itself simple, as $\mathfrak{A}$ is simple and due to the distinctness of the g_e and their inverses.

Any non-degenerate cycle in $\hat{\mathfrak{A}}$ projects to a non-degenerate cycle in the Cayley graph of $G, (g_e)_{e \in E}$. Therefore, $\mathfrak{A} \otimes_{\mathbf{g}} G$ will be k-acyclic if the girth of the Cayley graph associated with $G, (g_e)_{e \in E}$ is at least k. Suitable Cayley graphs have explicitly been constructed, with asymptotically near optimal dependence of the size of the graph (or group) on the required girth and degree. Note that for our application, the degree of the required Cayley graph is $d = 2|E^{\mathfrak{A}}|$. These bounds guarantee k-acyclic bisimilar covers of size polynomial in the size of the given $\mathfrak{A}$, for any fixed k. It is also clear that an exponential growth in terms of k is unavoidable. See [1] for a full discussion of these explicit constructions of Cayley graphs with large girth, and [14] for another intuitive though exponential construction inspired by the idea of local bisimilar unravellings.

[3] This also answers a question left open in the proceedings version [15].

Theorem 30 (Margulis, Imrich). *For d and k there are d-regular Cayley graphs of regular degree d, size $\mathcal{O}(d^{ck})$ (c some fixed constant) and girth no less than k.*

We have proved the following lemma, which covers Proposition 29 for simple transition systems.

Lemma 31. *Any simple $\mathfrak{A}$ admits a faithful cover $\pi\colon \mathfrak{A}\otimes_{\mathbf{g}} G \to \mathfrak{A}$ by a simple k-acyclic structure of size $\mathcal{O}(|E|^{ck})$, for a suitable choice of G.*

The general case For structures $\mathfrak{A} = (A, \bar{R}, \bar{P})$ that are not simple it now suffices to find first a faithful bisimilar cover $\pi\colon \mathfrak{A}' \to \mathfrak{A}$ by some simple $\mathfrak{A}'$, and then apply the above construction to further eliminate short cycles from these.

A simple way to achieve this involves an intermediate encoding in which edges of $\mathfrak{A}$ are replaced by paths of length two that pass through new vertices whose colour characterises the kind of edge involved. In more detail, we associate with $\tau = \tau^{(1)} \cup \tau^{(2)}$ a new vocabulary τ_s consisting of $\tau^{(1)}$ together with new unary predicates Q_R for each $R \in \tau^{(2)}$ and a new binary predicate S.

With an arbitrary transition system $\mathfrak{A}$ of type τ associate a simple τ_s transition system $\mathfrak{A}_s$ over the universe A_s which is the disjoint union of A and the disjoint union of the $R^{\mathfrak{A}}$ for $R \in \tau^{(2)}$. (Note that each individual edge of $\mathfrak{A}$ gives rise to a new element in $\mathfrak{A}_s$.) The $P \in \tau^{(1)}$ are interpreted as in $\mathfrak{A}$: $P^{\mathfrak{A}_s} = P^{\mathfrak{A}}$. The new Q_R mark the elements encoding the $R^{\mathfrak{A}}$-edges: $Q_R^{\mathfrak{A}_s} = R^{\mathfrak{A}} \subseteq A_s$. $S^{\mathfrak{A}_s}$ finally is interpreted to contain exactly all those pairs $(a, e) \in A \times Q_R^{\mathfrak{A}_s}$ and $(e, a') \in Q_R^{\mathfrak{A}_s} \times A$ for which $e = (a, a') \in R^{\mathfrak{A}}$, $R \in \tau^{(2)}$:

$$\begin{aligned} Q_R^{\mathfrak{A}_s} &= \{e\colon e \in R^{\mathfrak{A}}\}, \\ S^{\mathfrak{A}_s} &= \{(a, e), (e, a')\colon (a, a') \in R^{\mathfrak{A}}, R \in \tau^{(2)}\}. \end{aligned}$$

Clearly $\mathfrak{A}_s$ is simple. We may now apply the above construction to obtain a faithful bisimilar cover $\pi_s\colon \hat{\mathfrak{A}}_s \to \mathfrak{A}_s$ by a simple, 5-acyclic τ_s structure $\hat{\mathfrak{A}}_s$. Directed S-paths of length two in $\mathfrak{A}_s$ of the form a, e, a' have unique lifts to any $\hat{a} \in \pi_s^{-1}(a)$ or any $\hat{a}' \in \pi_s^{-1}(a')$. Conversely any length two directed S-path of the form $\hat{a}, \hat{e}, \hat{a}'$ in $\hat{\mathfrak{A}}_s$ with $\hat{e} \in Q_R$ projects to a path a, e, a' in $\mathfrak{A}_s$ with $e \in Q_R$, and therefore corresponds to an R-edge in $\mathfrak{A}$. Note also that $P^{\hat{\mathfrak{A}}_s} \subseteq \pi_s^{-1}(A)$ where we think of A as a subset of A_s.

Any such simple τ_s structure $\hat{\mathfrak{A}}_s$ induces a simple τ structure $\hat{\mathfrak{A}}$, according to the following straightforward reverse transformation:

$$\begin{aligned} \hat{A} &= \pi_s^{-1}(A) \text{ where } A \subseteq A_s, \\ P^{\hat{\mathfrak{A}}} &= P^{\hat{\mathfrak{A}}_s}, \\ R^{\hat{\mathfrak{A}}} &= \{(\hat{a}, \hat{a}')\colon (\hat{a}, \hat{e}), (\hat{e}, \hat{a}') \in S^{\hat{\mathfrak{A}}_s}, \text{ for some } e \in Q_R^{\hat{\mathfrak{A}}_s}\}. \end{aligned}$$

Since $\hat{\mathfrak{A}}_s$ does not have any non-degenerate 4-cycles, $\hat{\mathfrak{A}}$ turns out simple. The above considerations about projections and unique lifts of paths imply that π_s induces a homomorphism $\pi\colon \hat{\mathfrak{A}} \to \mathfrak{A}$ which moreover is a faithful bisimilar cover. We have found a bisimilar cover $\pi\colon \hat{\mathfrak{A}} \to \mathfrak{A}$ as formulated in the following corollary.

Corollary 32. *Any finite transition system $\mathfrak{A}$ admits a faithful bisimilar cover $\pi\colon \hat{\mathfrak{A}} \to \mathfrak{A}$ by a finite simple transition system $\hat{\mathfrak{A}}$ of polynomially bounded size.*

Together with Lemma 31 this proves Proposition 29 in the general case.

4 Upgrading global bisimulation equivalences

We use the results from the previous section to obtain bisimilar companions which can serve to upgrade ℓ-bisimulations between finite transition systems to stronger forms of local first-order equivalence following the idea behind Lemma 22. Compare Definition 14 for upgrading, Definition 21 for the relevant levels $\equiv^{(\ell)}_{q,n}$ of local FO-equivalence, and recall Gaifman's theorem, Theorem 17, or its specific ramifications from Proposition 19.

Recall in particular that any FO-formula $\varphi(x)$ is equivalent to one in Gaifman form, and as such is therefore invariant under $\equiv^{(\ell)}_{q,n}$ for suitable levels of ℓ (its locality rank), q (its local quantifier rank), and n (its scattering rank).

4.1 Upgrading global two-way bisimilarity

The main proposition about upgrading from $\approx^{\ell}$ is the following. Its proof, however, is broken down into a sequence of lemmas that highlight some intermediate upgrading stages in their own right. Locally acyclic covers are used in the central step, Lemma 35.

Proposition 33. *Modulo $\approx$, $\approx^{\ell}$ can be upgraded to $\equiv^{(\ell)}_{q,n}$, for any q and n, classically as well as in finite models.*

For technical reasons we consider a strengthening of two-way bisimulation in which the second player can match multiplicities up to q in responses to the first player's challenges in each individual round, for some fixed q. Formally, the usual back-and-forth requirements are strengthened to corresponding q-back-and-forth requirements according to, for instance,

(*q-forth* along forward R:) for any $(a,b) \in Z$ and any distinct $a'_1,\dots,a'_k \in A$ such that $(a,a'_i) \in R^{\mathfrak{A}}$ for $1 \leqslant i \leqslant k$, where $k \leqslant q$, there are distinct $b'_1,\dots,b'_k \in B$ such that $(b,b'_i) \in R^{\mathfrak{B}}$ and $(a'_i,b'_i) \in Z'$ for $1 \leqslant i \leqslant k$.

We write $\sim_{-}^{\ell;q}$ for the corresponding level of two-way q-back-and-forth ℓ-bisimulation, formally induced by a depth ℓ stratified back-and-forth system with the appropriate two-way q-back-and-forth conditions.

The corresponding variant of $\approx^{\ell}$, global two-way q-back-and-forth ℓ-bisimulation $\approx^{\ell;q}$ is analogously defined, with the additional requirement that the corresponding back-and-forth system covers all of $\mathfrak{A}$ and $\mathfrak{B}$: $\mathfrak{A} \approx^{\ell;q} \mathfrak{B}$ iff for every a in $\mathfrak{A}$ there is some b in $\mathfrak{B}$ such that $\mathfrak{A},a \sim_{-}^{\ell;q} \mathfrak{B},b$, and vice versa.

Lemma 34. *Modulo $\approx$, $\approx^{\ell}$ can be upgraded to $\approx^{\ell;q}$, for any q, classically as well as in finite models.*

Proof. If $\mathfrak{A},a \approx^{\ell} \mathfrak{B},b$, it suffices to blow up all multiplicities in $\mathfrak{A}$ and $\mathfrak{B}$ q-fold to achieve the desired degree of bisimulation equivalence. This is done with the following operation:

$$\mathfrak{A} \otimes q = \big(A \times \{1,\dots,q\}, (\rho^{-1}(R))_{R\in\tau^{(2)}}, (\rho^{-1}(P))_{P\in\tau^{(1)}}\big)$$

where $\rho\colon \mathfrak{A}\otimes q \to \mathfrak{A}$ is the natural projection.

Clearly $\mathfrak{A}\otimes q \approx \mathfrak{A}$ and $\mathfrak{A}\otimes q,(a,1) \approx^{\ell;q} \mathfrak{B}\otimes q,(b,1)$. □

Lemma 35. *Modulo $\approx$, $\approx^{\ell;q}$ can be upgraded to $\equiv^{(\ell)}_{q,1}$, classically as well as in finite models.*

Proof. Let $\mathfrak{A}, a \approx^{\ell;q} \mathfrak{B}, b$, and let $\pi \colon \hat{\mathfrak{A}} \to \mathfrak{A}$ and $\pi \colon \hat{\mathfrak{B}} \to \mathfrak{B}$ be faithful bisimilar covers by $(2\ell+2)$-acyclic simple transition systems, according to Proposition 29. Let $\hat{a}$ and $\hat{b}$ be any representatives in $\pi^{-1}(a)$ and $\pi^{-1}(b)$, respectively. Note that automatically $\hat{\mathfrak{A}}, \hat{a} \approx^{\ell;q} \hat{\mathfrak{B}}, \hat{b}$, as the cover is faithful. We claim that $\hat{\mathfrak{A}}, \hat{a} \equiv^{(\ell)}_{q,1} \hat{\mathfrak{B}}, \hat{b}$.

$\hat{\mathfrak{A}} \approx^{\ell;q} \hat{\mathfrak{B}}$ implies that for every a there is a b such that $\hat{\mathfrak{A}}, a \sim_{-}^{\ell;q} \hat{\mathfrak{B}}, b$, and vice versa.

The latter implies that $\hat{\mathfrak{A}} \restriction U^\ell(a), a \sim_{-}^{\ell;q} \hat{\mathfrak{B}} \restriction U^\ell(b), b$. Any two such $\hat{\mathfrak{A}} \restriction U^\ell(a)$ and $\hat{\mathfrak{B}} \restriction U^\ell(b)$ are acyclic, since $\hat{\mathfrak{A}}$ and $\hat{\mathfrak{B}}$ themselves are $(2\ell+2)$-acyclic.

To establish $\hat{\mathfrak{A}}, \hat{a} \equiv^{(\ell)}_{q,1} \hat{\mathfrak{B}}, \hat{b}$, it therefore suffices to show the following.

Claim 36. *Let $\mathfrak{A}, a$ and $\mathfrak{B}, b$ be simple and acyclic and such that $A \subseteq U^\ell(a)$ and $B \subseteq U^\ell(b)$. Then $\mathfrak{A}, a \sim_{-}^{\ell;q} \mathfrak{B}, b$ implies $\mathfrak{A}, a \equiv_q \mathfrak{B}, b$.*

For the proof of the claim, we exhibit a strategy in the the q-round Ehrenfeucht-Fraïssé game on $\mathfrak{A}, a$ and $\mathfrak{B}, b$. Fix $\mathfrak{A}, a$ and $\mathfrak{B}, b$ as in the claim. For a tuple $\mathbf{a} = (a_1, \ldots, a_k)$ in $\mathfrak{A}$ we let $\mathrm{span}(a, \mathbf{a})$ denote the set of those elements of A that lie on one of the shortest paths connecting a to a_i in the Gaifman graph $G(\mathfrak{A})$ of $\mathfrak{A}$, for $1 \leqslant i \leqslant k$. For a' in $\mathfrak{A}$ we let $d(a, a')$ denote the Gaifman distance (length of the shortest path) from a to a'. Similar notions apply in $\mathfrak{B}, b$. The strategy for Player II consists in maintaining the following condition, in terms of elements $\mathbf{a} = (a_1, \ldots, a_k)$ and $\mathbf{b} = (b_1, \ldots, b_k)$ marked so far in $\mathfrak{A}$ and $\mathfrak{B}$, respectively.

$(*)$ there is an isomorphism $f \colon \mathfrak{A} \restriction \mathrm{span}(a, \mathbf{a}) \simeq \mathfrak{B} \restriction \mathrm{span}(b, \mathbf{b})$ such that for all $a', b' = f(a')$: $\mathfrak{A}, a' \sim_{-}^{\ell';q} \mathfrak{B}, b'$ where $\ell' = \ell - d(a, a') = \ell - d(b, b')$.

Condition $(*)$ is obviously met initially, with empty $\mathbf{a}$ and $\mathbf{b}$ and for $f \colon a \mapsto b$.

Assume $(*)$ is true after round $k < \ell$, let $f \colon \mathfrak{A} \restriction \mathrm{span}(a, \mathbf{a}) \simeq \mathfrak{B} \restriction \mathrm{span}(b, \mathbf{b})$ accordingly, and suppose without loss of generality that Player I selects a' in $\mathfrak{A}$ in the next round and that $a' \notin \mathrm{span}(a, \mathbf{a})$. Let $d(a, a') = t$ and consider the shortest path $a = a'_0, a'_1, \ldots, a'_s, \ldots, a'_t = a'$ from a to a' in $G(\mathfrak{A})$. Let a'_s be the last element on this path that is contained in $\mathrm{span}(a, \mathbf{a})$, so that $\mathrm{span}(a, \mathbf{a}a') = \mathrm{span}(a, \mathbf{a}) \dot{\cup} \{a'_{s+1}, \ldots, a'_t\}$. Let $b'_i = f(a'_i)$ for $i \leqslant s$. By the above condition, $\mathfrak{A}, a'_s \sim_{-}^{\ell-s;q} \mathfrak{B}, b'_s$. Successively exercising the two-way q-forth property we find a matching path $b'_s, b'_{s+1}, \ldots, b'_t$ in $G(\mathfrak{B})$ always using fresh elements b_i for $i > s$, such that also for $s < i \leqslant t$:

- $(b'_{i-1}, b'_i) \in R^{\mathfrak{B}}$ iff $(a'_{i-1}, a'_i) \in R^{\mathfrak{A}}$, and similarly w.r.t. R^{-1}, for all $R \in \tau^{(2)}$,
- $d(b, b'_i) = i$,
- $\mathfrak{A}, a'_i \sim_{-}^{\ell-i;q} \mathfrak{B}, b'_i$.

Let $b' = b'_t$ and $f' \colon \mathfrak{A} \restriction \mathrm{span}(a, \mathbf{a}a') \to \mathfrak{B} \restriction \mathrm{span}(b, \mathbf{b}b')$ the extension of f that sends a'_i to b'_i for $s < i \leqslant t$. Simplicity and acyclicity of $\mathfrak{A}$ and $\mathfrak{B}$ guarantee that f' is an isomorphism; moreover f' satisfies the required bisimulation conditions by construction.

Our choice of b' for a' exemplifies the way in which $(*)$ is maintained in response to a next move of Player I in $\mathfrak{A}$. A challenge played in $\mathfrak{B}$ can be answered in a symmetric fashion. $\square$

Lemma 37. *Modulo $\approx$, $\equiv^{(\ell)}_{q,1}$ can be upgraded to $\equiv^{(\ell)}_{q,n}$ for any n, classically as well as in finite models.*

Proof. Clearly, if $\mathfrak{A}, a \equiv_{q,1}^{(\ell)} \mathfrak{B}, b$, then $n \cdot \mathfrak{A}, a \equiv_{q,n}^{(\ell)} n \cdot \mathfrak{B}, b$, where $n \cdot \mathfrak{A}$ is the n-fold disjoint sum of copies of $\mathfrak{A}$. □

4.2 Upgrading global forward bisimilarity

Lemma 38. *Modulo $\sim_\forall$, $\sim_\forall^{2\ell}$ can be upgraded to $\approx^\ell$, classically as well as in finite models.*

It is easy to see that one cannot achieve a similar upgrade without decreasing the approximation level ℓ. For instance, a two-edge chain is 1-bisimilar (in the sense of $\sim_\forall^1$) to a one-edge chain. But any globally bisimilar companion structures of these would still be of depths 2 and 1, respectively. These therefore cannot be 1-bisimilar in the two-way sense: the former must have nodes with non-zero in- *and* out-degree; the latter cannot have such.

Let for the following $\mathrm{tp}_{\mathfrak{A}}^\ell(a)$ denote the ℓ-bisimulation type ($\sim^\ell$-type) of a in $\mathfrak{A}$. Semantically, $\mathrm{tp}_{\mathfrak{A}}^\ell(a)$ precisely determines the $\sim^\ell$ equivalence class of $\mathfrak{A}, a$. Syntactically $\mathrm{tp}_{\mathfrak{A}}^\ell(a)$ is defined by the corresponding depth ℓ modal Hintikka formula. We note that $\sim^\ell$ has finite index, for any fixed finite τ.

The full bisimulation type ($\sim$-type) of a in $\mathfrak{A}$ is in the following denoted $\mathrm{tp}_{\mathfrak{A}}(a)$.

With a directed path $a_0, \ldots, a_k$ in $\mathfrak{A}$ we associate the string consisting of the ℓ-bisimulation types $\mathrm{tp}_{\mathfrak{A}}^\ell(a_i)$ and the edge types linking a_i to a_{i+1} along this path,

$$\mathrm{tp}_{\mathfrak{A}}^\ell(a_0), R_0, \mathrm{tp}_{\mathfrak{A}}^\ell(a_1), R_1, \ldots, R_{k-1}, \mathrm{tp}_{\mathfrak{A}}^\ell(a_k),$$

where $(a_i, a_{i+1}) \in R_i^{\mathfrak{A}}$.

Definition 39. A string $\mathrm{tp}_{\mathfrak{A}}^\ell(a_0), R_0, \ldots, R_{k-1}, \mathrm{tp}_{\mathfrak{A}}^\ell(a_k)$ associated with a directed path $(a_i, a_{i+1}) \in R_i^{\mathfrak{A}}$ is an *ℓ-history* of $a = a_k$ in $\mathfrak{A}$ if either $k = \ell$ (we refer to a *proper ℓ-history*), or $k < \ell$ and the path is not backward extendible, i.e., a_0 has in-degree zero (we refer to a *short ℓ-history*).

We say that a in $\mathfrak{A}$ has a *unique ℓ-history* if all ℓ-histories of a in $\mathfrak{A}$ are identical (in particular they are all short of the same length $k < \ell$, or all proper); in this case $\mathrm{hist}_{\mathfrak{A}}^\ell(a)$ stands for this unique ℓ-history.

We say that $\mathfrak{A}$ has *unique ℓ-histories* if every node in $\mathfrak{A}$ has a unique ℓ-history.

Note that tree structures in particular do have unique histories. Also note that the ℓ-history (or ℓ-histories) of a node determines its ℓ'-histories for any $\ell' \leqslant \ell$.

Let us say that a bisimulation $\mathfrak{A} \sim_\forall^\ell \mathfrak{B}$ respects zero in-degree, if for every node a in $\mathfrak{A}$ of in-degree zero there is a node b in $\mathfrak{B}$ of in-degree zero such that $\mathfrak{A}, a \sim^\ell \mathfrak{B}, b$ and vice versa. Note that if $\mathfrak{A} \sim_\forall^\ell \mathfrak{B}$ do not satisfy this condition, we can still always pass to companions $\mathfrak{A}' \sim_\forall \mathfrak{A}$ and $\mathfrak{B}' \sim_\forall \mathfrak{B}$ where $\mathfrak{A}' \sim_\forall^\ell \mathfrak{B}'$ does respect zero in-degree. Simply let $\mathfrak{A}'$ be the disjoint union of all structures $\mathfrak{A}_a$ obtained by adding a new copy of a with outgoing edges into $\mathfrak{A}$ just as from a but without any incoming edges, for each a in $\mathfrak{A}$. If $\mathfrak{B}'$ is similarly obtained from $\mathfrak{B}$, then $\mathfrak{A}'$ and $\mathfrak{B}'$ realise exactly the same bisimulation types as $\mathfrak{A}$ and $\mathfrak{B}$, respectively, and in each of them any bisimulation type realised at all is also realised by a node of zero in-degree. This crude construction does however not preserve uniqueness of histories.

Lemma 40. *Let $\mathfrak{A}$ and $\mathfrak{B}$ have unique ℓ-histories. If $\mathfrak{A} \sim_{\forall}^{2\ell} \mathfrak{B}$ respects zero in-degree, then $\mathfrak{A} \approx^{\ell} \mathfrak{B}$.*

Proof. Let $\mathfrak{A}$ and $\mathfrak{B}$ be as in the lemma. We show for instance that for any a in $\mathfrak{A}$ there is some b in $\mathfrak{B}$ such that $\mathfrak{A}, a \approx^{\ell} \mathfrak{B}, b$.

Assume first that the ℓ-history of a is proper. Let a_0 be a node in $\mathfrak{A}$ from which a is reachable on a path $a_0, a_1, \ldots, a_\ell = a$ of length ℓ. Choose b_0 such that $\mathfrak{A}, a_0 \sim_{\forall}^{2\ell} \mathfrak{B}, b_0$. Exercising the forth property ℓ times, following the path from a_0 to a in $\mathfrak{A}$, we find a path $b_0, b_1, \ldots, b_\ell$ for which $\mathfrak{A}, a_i \sim^{2\ell-i} \mathfrak{B}, b_i$, for $0 \leqslant i \leqslant \ell$. Choosing $b := b_\ell$ we have found an element in $\mathfrak{B}$ whose ℓ-history is identical with that of a.

In case that the ℓ-history of a is short, we work with this short history and, since the given 2ℓ-bisimulation respects zero in-degree, similarly find a matching b that has the same short ℓ-history as a.

It now suffices to argue that $\mathrm{hist}^{\ell}_{\mathfrak{A}}(a) = \mathrm{hist}^{\ell}_{\mathfrak{B}}(b)$ implies $\mathfrak{A}, a \approx^{\ell} \mathfrak{B}, b$. To this end consider the stratified system $(Z_m)_{0 \leqslant m \leqslant \ell}$ where

$$Z_m := \big\{(a,b) \in A \times B \colon \mathrm{hist}^{m}_{\mathfrak{A}}(a) = \mathrm{hist}^{m}_{\mathfrak{B}}(b)\big\}.$$

By the above, $\pi_1(Z_\ell) = A$, and by symmetry also $\pi_2(Z_\ell) = B$. In order to show that $(Z_m)_{0 \leqslant m \leqslant \ell} \colon \mathfrak{A} \approx^{\ell} \mathfrak{B}$ it remains to establish that this stratified system satisfies the two-way back-and-forth properties. For this observe that for $1 \leqslant k \leqslant \ell$, if $\mathrm{hist}^{k}_{\mathfrak{A}}(a) = \mathrm{hist}^{k}_{\mathfrak{B}}(b)$, then

(i) $\mathfrak{A}, a \sim^{k} \mathfrak{B}, b$.

(ii) if a' and b' are obtained as corresponding back-and-forth extensions of (a,b) along edges (a,a') and (b,b') in the sense of $\mathfrak{A}, a \sim^{k} \mathfrak{B}, b$, then $\mathrm{hist}^{k-1}_{\mathfrak{A}}(a') = \mathrm{hist}^{k-1}_{\mathfrak{B}}(b')$.

(iii) a has zero in-degree iff b has; otherwise, if a' and b' are predecessors along corresponding edges (a',a) and (b',b), then also $\mathrm{hist}^{k-1}_{\mathfrak{A}}(a') = \mathrm{hist}^{k-1}_{\mathfrak{B}}(b')$.

Of these, (i) is trivial by agreement of $\sim^{k}$-types in a and b in particular. (ii) follows from the fact that a' and b' have unique ℓ-histories, whence they in particular also have unique $(k-1)$-histories; the latter are exemplified by the length $(k-2)$ suffixes of the $(k-1)$-histories of a and b (which are identical) expanded by the edge type of (a,a') and (b,b') and $\mathrm{tp}^{k-1}_{\mathfrak{A}}(a') = \mathrm{tp}^{k-1}_{\mathfrak{B}}(b')$ (identical according to the back-and-forth choice of a' and b').

For (iii): as a and b have identical unique k-histories, one of them can be short of length zero only if the other is. If they are not of length zero, these identical k-histories are, as unique histories, exemplified by k-histories involving a' and b' as immediate predecessors, respectively. The identical $(k-1)$-prefixes of these k-histories imply the desired identity of $(k-1)$-histories. □

Proof of Lemma 38. We provide partner structures $\tilde{\mathfrak{A}} \sim_{\forall} \mathfrak{A}$ and $\tilde{\mathfrak{B}} \sim_{\forall} \mathfrak{B}$ that have unique ℓ-histories and realise in nodes of zero in-degree all $\sim^{\ell}$-types that are realised at all. The latter condition implies in particular that any maximal global ℓ-bisimulation between $\tilde{\mathfrak{A}}$ and $\tilde{\mathfrak{B}}$ will respect zero in-degree. It then follows from Lemma 40 that $\tilde{\mathfrak{A}} \approx^{\ell} \tilde{\mathfrak{B}}$, whence we have upgraded $\sim_{\forall}^{2\ell}$ to $\approx^{\ell}$ in $\sim_{\forall}$ equivalent companion structures as required. The construction is explicitly carried out for $\mathfrak{A}$.

Let H be the finite set of all proper ℓ-histories realisable in any τ-structure, $|H| = n$. For a in $\mathfrak{A}$ let $\mathfrak{A}_a$ be the result of unravelling $\mathfrak{A}$ to depth $\ell + 1$ from a. In other words, we restrict the usual tree unravelling $\mathfrak{A}_a^*$ of $\mathfrak{A}$ from a to $\mathfrak{A}_a^*\restriction N^{\ell+1}(a)$. We let $\tilde{\mathfrak{A}}'$ consist of the disjoint sum of $n+1$ copies each of all these $\mathfrak{A}_a$ for $a \in A$. Label the $n+1$ copies of $\mathfrak{A}_a$ as $\mathfrak{A}_{a,h}$ for $h \in H$ and $\mathfrak{A}_{a,\emptyset}$ for the one extra.

Some surgery is necessary to produce $\tilde{\mathfrak{A}}$ from $\tilde{\mathfrak{A}}'$. Note that leaf nodes (nodes at distance $\ell+1$ from the root a) in copies of $\mathfrak{A}_a$ do not realise the appropriate $\sim^\ell$-types (unless they happen to be derived from nodes of zero out-degree in $\mathfrak{A}$).

This is set right if we now identify any such leaf node c with the root c in any copy of $\mathfrak{A}_c$ in $\tilde{\mathfrak{A}}'$. In order to preserve uniqueness of ℓ-histories through this process, though, the target copies are determined according to the ℓ-history that c has in the unrestricted unravelling of $\mathfrak{A}$. In more detail, let for a leaf node c in $\mathfrak{A}_a = \mathfrak{A}_a^*\restriction N^{\ell+1}(a)$

$$h(c) := \mathrm{hist}^{\ell}_{\mathfrak{A}_a^*}(c)$$

be the ℓ-history of c in $\mathfrak{A}_a^*$. Note that this history is proper and also that it attributes to c itself the $\sim^\ell$-type that it should have. Now $\tilde{\mathfrak{A}}$ is obtained from $\tilde{\mathfrak{A}}'$ through identification of any leaf node c in any copy of any $\mathfrak{A}_a$ with the root in $\mathfrak{A}_{c,h(c)}$.

It is clear that $\tilde{\mathfrak{A}} \sim_\forall \mathfrak{A}$; that $\tilde{\mathfrak{A}}$ has unique ℓ-histories; and that any ℓ-bisimulation type realised in $\tilde{\mathfrak{A}}$ is realised by some a in $\mathfrak{A}$ and therefore realised by the root a in $\mathfrak{A}_{a,\emptyset}$, a node of zero in-degree in $\tilde{\mathfrak{A}}$. □

5 Characterisation theorems

To finish the arguments for Theorem 4 and 5 we follow the pattern outlined in Observation 13 and Lemma 22 and finally prove the following. It may also be instructive to compare this with the simpler case of the van Benthem-Rosen theorem as proved in section 2.5.

Proposition 41. *Both classically and in the sense of finite model theory, for $\varphi(x) \in \mathrm{FO}$ of locality rank ℓ:*

(i) if φ is invariant under global two-way bisimulation $\approx$ then φ is in fact invariant under $\approx^\ell$.

(ii) if φ is invariant under global bisimulation $\sim_\forall$ then φ is invariant under $\sim_\forall^{2\ell}$.

Proof. By upgrading as in Lemma 22, using Proposition 33 for (i) and additionally Lemma 38 for (ii).

For (i): as a FO-formula of locality rank ℓ, φ is preserved under $\equiv_{q,n}^{(\ell)}$ for suitable q and n. (In fact, Proposition 19 even tells us that φ is preserved under $\equiv_{q,1}^{(\ell)}$.) By Proposition 33 $\approx^\ell$ can be upgraded modulo $\approx$ to any such level, classically and in finite models. Therefore φ is invariant under $\approx^\ell$.

For (ii), the required upgrading is modulo $\sim_\forall$ and needs to take us from $\sim_\forall^{2\ell}$ to $\equiv_{q,n}^{(\ell)}$ (or just to $\equiv_{q,1}^{(\ell)}$ if Proposition 19 is invoked). This is achieved by first upgrading to $\approx^\ell$ according to Lemma 38, and then proceeding as in case (i). □

$$
\begin{tikzcd}
\mathfrak{A}, a \arrow[r, dash, "\approx^{\ell}"] \arrow[d, dash, "\approx"'] & \mathfrak{B}, b \arrow[d, dash, "\approx"] \\
\hat{\mathfrak{A}}, \hat{a} \arrow[r, dash, "\equiv^{(\ell)}_{q,n}"] & \hat{\mathfrak{B}}, \hat{b}
\end{tikzcd}
\qquad
\begin{tikzcd}
\mathfrak{A}, a \arrow[r, dash, "\sim_{\forall}^{2\ell}"] \arrow[d, dash, "\sim_{\forall}"'] & \mathfrak{B}, b \arrow[d, dash, "\sim_{\forall}"] \\
\tilde{\mathfrak{A}}, \tilde{a} \arrow[r, dash, "\approx^{\ell}"] \arrow[d, dash, "\approx"'] & \tilde{\mathfrak{B}}, \tilde{b} \arrow[d, dash, "\approx"] \\
\hat{\mathfrak{A}}, \hat{a} \arrow[r, dash, "\equiv^{(\ell)}_{q,n}"] & \hat{\mathfrak{B}}, \hat{b}
\end{tikzcd}
$$

Corollary 42. *Both classically and in the sense of finite model theory. Let $\varphi(x) \in \mathrm{FO}$ be invariant under $\approx$. If φ is of locality rank ℓ, then it can equivalently be expressed in $\mathrm{ML}^{\bar{}\,\forall}$ at modal nesting depth ℓ.*

For φ invariant under $\sim_{\forall}$ our proof only yields expressibility in $\mathrm{ML}^{\forall}$ at nesting depth 2ℓ where ℓ is the locality rank of φ. This seems to be sub-optimal, and may be an artifact of the particular upgrading strategy employed.

6 The guarded picture

An investigation of guarded bisimulation invariance over (finite) transition systems can be carried out analogously to what has been done for global two-way bisimulation invariance above. In particular, we provide faithful (finite) locally acyclic guarded covers for (finite) transition systems in section 6.1; we show how these can be used to upgrade guarded ℓ-bisimulation to appropriate levels $\equiv^{(\ell)}_{q,n}$ of local first-order equivalence in section 6.2; and finally put these results together to prove Theorem 6 in section 6.3.

The main tool to bridge the gap between global two-way bisimulation $\approx$ and guarded bisimulation $\sim_{\mathrm{g}}$ over relational structures of width two involves an encoding of guarded quantifier free types as transition relations. We fix some terminology for this purpose.

A *non-degenerate* 2*-type* over τ is a full description of the isomorphism type of a two element τ-structure in variables x, y, which may be formalised as a conjunction over a maximally consistent set of atomic and negated atomic τ-formulae in variables x and y including the conjunct $x \neq y$. We write $p(x, y)$ for 2-types, and $\mathrm{tp}_{\mathfrak{A}}(a, a')$ for the unique 2-type satisfied by (a, a') in $\mathfrak{A}$, for $a \neq a'$.

A 1*-type* over τ similarly is a full description of a one element τ-structure (which apart from monadic information contains the information about loops w.r.t. the binary predicates). Obvious notation like $\mathrm{tp}_{\mathfrak{A}}(a) = q$ applies.

For a 2-type $p = p(x, y)$ we let p_x and p_y be the unique 1-types obtained as the restrictions of p to its x-part or y-part, respectively. Let p^{-1} stand for the result of swapping x and y in p. A 2-type p is *symmetric* if $p = p^{-1}$, asymmetric otherwise.

Definition 43. A 2-type $p(x, y)$ over τ is *guarded* if it includes a conjunct Rxy or Ryx for some $R \in \tau^{(2)}$. In other words, guarded 2-types are those 2-types that are realised by non-degenerate guarded pairs.

6.1 Locally acyclic guarded covers

Definition 44. A homomorphism $\pi\colon \hat{\mathfrak{A}} \to \mathfrak{A}$ is a *guarded cover* of $\mathfrak{A}$ by $\hat{\mathfrak{A}}$ if $Z_\pi = \{\pi \restriction \hat{s}\colon \hat{s} \subseteq \hat{A} \text{ guarded in } \hat{\mathfrak{A}}\}$ is a guarded bisimulation between $\hat{\mathfrak{A}}$ and $\mathfrak{A}$.

The guarded cover π is *faithful* if, for every $\hat{a}$ and every guarded 2-type $p = p(x, y)$, π restricts to a bijection between $\{\hat{a}' \in \hat{A}\colon \mathrm{tp}_{\hat{\mathfrak{A}}}(\hat{a}, \hat{a}') = p\}$ and $\{a' \in A\colon \mathrm{tp}_{\mathfrak{A}}(a, a') = p\}$.

The above construction of faithful, locally acyclic bisimilar covers of transition systems naturally lends itself to the extension to guarded covers in relational vocabularies of width two. One merely has to encode all non-degenerate quantifier-free 2-types by new binary relations which can be interpreted so as to form a simple transition system which faithfully encodes the underlying relational structure. Similar considerations and translations for guarded logics on graphs are presented in [11].

Let Λ_2 be a fixed maximal set of guarded 2-types over τ containing all symmetric guarded 2-types, and precisely one of p or p^{-1} for every asymmetric guarded 2-type. Let Λ_1 be the set of all 1-types over τ.

We associate with τ a new vocabulary τ_g consisting of new unary predicates P_q for every $q \in \Lambda_1$ and new binary R_p for every $p \in \Lambda_2$.

In order to deal with the encoding of symmetric 2-types in a simple transition system, which cannot have undirected edges, we break the symmetry by means of an arbitrary auxiliary ordering on the universe. Let $\mathfrak{A}$ be a τ-structure, $<$ an arbitrary linear ordering $<$ on A. With $(\mathfrak{A}, <)$ associate the following simple τ_g transition system $\mathfrak{A}_g = (A, (P), (Q))$ on universe A:

$$\begin{aligned}
P_q^{\mathfrak{A}_g} &= \{a\colon q = \mathrm{tp}_{\mathfrak{A}}(a)\} && \text{(for each } q \in \Lambda_1\text{)},\\
R_p^{\mathfrak{A}_g} &= \{(a, a')\colon a < a' \text{ and } \mathrm{tp}_{\mathfrak{A}}(a, a') = p\} && \text{(for each symmetric } p \in \Lambda_2\text{)},\\
R_p^{\mathfrak{A}_g} &= \{(a, a')\colon \mathrm{tp}_{\mathfrak{A}}(a, a') = p\} && \text{(for each asymmetric } p \in \Lambda_2\text{)}.
\end{aligned}$$

Clearly $\mathfrak{A}_g$ is simple and satisfies the following compatibility conditions:

(a) the P_q partition the universe.

(b) if $(a, a') \in R_p$ then $a \in P_q$ for $q = p_x$ and $a' \in P_q$ for $q = p_y$.

(c) for any non-degenerate pair (a, a'), at most one binary relation R_p can link a with a' (simplicity).

Note that (a) and (b) are preserved under global bisimulation.

Conversely, for any simple τ_g transition system $\mathfrak{B}_g$ satisfying (a) and (b) there is a unique associated τ-structure $\mathfrak{B}$. The universe of $\mathfrak{B}$ is that of $\mathfrak{B}_g$. Monadic and binary predicates from τ are interpreted so as to be consistent with the 1- and 2-types prescribed by the P_q and the R_p, and such that a non-degenerate pair (b, b') will be guarded in $\mathfrak{B}$ if and only if b and b' are linked by some R_p in $\mathfrak{B}_g$.

A small subtlety arises with respect to loops. A loop $(a, a) \in R^{\mathfrak{A}}$ in a transition system is eliminated in acyclic bisimilar covers, but clearly cannot and must not be eliminated in a guarded cover. Correspondingly, the information about loops has been shifted into monadic predicates associated with the 1-types. But in order to get our criteria for *acyclicity* right in this context, we explicitly have to allow loops in k-acyclic covers. Deviating from Definition 26 we now do not insist on loop-freeness.

With this it is not hard to check the following.

Lemma 45. *Let $\mathfrak{B}_g$ and $\mathfrak{B}'_g$ be simple τ_g transition systems. Let $\mathfrak{B}_g \approx \mathfrak{B}'_g$ and let $\mathfrak{B}_g$ satisfy conditions (a) and (b) above. Then $\mathfrak{B}'_g$ also satisfies (a) and (b), and $\mathfrak{B} \sim_{\mathrm{g}} \mathfrak{B}'$ for the associated τ-structures.*

Let $\mathfrak{A}$ be a τ-structure with an associated τ_g transition system $\mathfrak{A}_g$. Let $\pi\colon \hat{\mathfrak{A}}_g \to \mathfrak{A}_g$ be a bisimilar cover of $\mathfrak{A}_g$ by a simple τ_g transition system $\hat{\mathfrak{A}}_g$. Then $\hat{\mathfrak{A}}_g$ satisfies (a) and (b) and for the induced τ-structure $\hat{\mathfrak{A}}$:

(i) $\pi\colon \hat{\mathfrak{A}} \to \mathfrak{A}$ is a guarded cover,

(ii) if $\hat{\mathfrak{A}}_g$ is k-acyclic then $\hat{\mathfrak{A}}$ is k-acyclic (apart from necessary loops);

(iii) if $\pi\colon \hat{\mathfrak{A}}_g \to \mathfrak{A}_g$ is faithful then so is $\pi\colon \hat{\mathfrak{A}} \to \mathfrak{A}$.

Putting this together with the covering results for (simple) transition systems obtained above we get the following.

Corollary 46. *Let τ be any finite relational vocabulary of width two, $\mathfrak{A}$ a finite τ-structure and $k > 3$. Then there is a faithful guarded cover $\pi\colon \hat{\mathfrak{A}} \to \mathfrak{A}$ by a finite structure $\hat{\mathfrak{A}}$ that is k-acyclic apart from necessary loops. For fixed k, the size of $\hat{\mathfrak{A}}$ can be polynomially bounded in terms of the size of $\mathfrak{A}$.*

An open issue related to this result concerns potential extensions to the setting of arbitrary relational vocabularies. We do not know whether one can similarly achieve finite guarded covers of finite relational structures that avoid short chordless cycles. See [12] for a discussion. In that paper another aspect of acyclicity (in hypergraphs)—to do with the avoidance of bad cliques (rather than cycles) in the Gaifman graph—is shown to be realisable in finite guarded covers, with applications to the clique guarded fragment and extension theorems for partial isomorphisms.

6.2 Upgrading guarded bisimilarity

Lemma 47. *Modulo $\sim_{\mathrm{g}}$, $\sim^{\ell}_{\mathrm{g}}$ can be upgraded to $\equiv^{(\ell)}_{q,n}$, for any levels q and n, classically as well as in restriction to finite transition systems.*

Proof. The proof is analogous to the sequence of upgradings in Lemmas 34, 35 and 37. Let $\mathfrak{A}, a \sim^{\ell}_{\mathrm{g}} \mathfrak{B}, b$. Combining the construction from Lemma 34 with the construction of faithful $(2\ell+2)$-acyclic guarded covers we find $\hat{\mathfrak{A}} \sim_{\mathrm{g}} \mathfrak{A}$ and $\hat{\mathfrak{B}} \sim_{\mathrm{g}} \mathfrak{B}$ such that $\hat{\mathfrak{A}} \equiv^{(\ell)}_{q,1} \hat{\mathfrak{B}}$.

This can further be boosted to $\equiv^{(\ell)}_{q,n}$ for any given n, if we pass to n-fold sums of disjoint copies: $n \cdot \hat{\mathfrak{A}} \equiv^{(\ell)}_{q,n} n \cdot \hat{\mathfrak{B}}$. □

6.3 The guarded characterisation theorem

To finish the argument for Theorem 6 we follow the pattern of Observation 13 and Lemma 22 and finally prove the following.

Proposition 48. *Both classically and in the sense of finite model theory: if $\varphi(x) \in \mathrm{FO}$ is invariant under guarded bisimulation $\sim_{\mathrm{g}}$ then φ is invariant under $\sim^{\ell}_{\mathrm{g}}$, where ℓ is the locality rank of φ.*

Proof. By upgrading: either we upgrade $\sim^{\ell}_{\mathrm{g}}$ directly to $\equiv^{(\ell)}_{q,n}$ where q and n are the local quantifier rank and scattering rank of φ in Gaifman form with locality rank ℓ, or we appeal to Proposition 19 and use the fact that φ can be expressed with scattering rank 1 so that an upgrading of $\sim_{\mathrm{g}}^{\ell}$ to $\equiv_{q,1}^{(\ell)}$ is in fact sufficient. Either way, for $n = 1$ or any desired value of n, the following diagram shows that φ is indeed invariant under $\sim_{\mathrm{g}}^{\ell}$ (overall or in restriction to finite models). $\square$

$$\begin{array}{ccc} \mathfrak{A}, a & \overset{\sim_{\mathrm{g}}^{\ell}}{\rule{3cm}{0.4pt}} & \mathfrak{B}, b \\ \Big| \sim_{\mathrm{g}} & & \Big| \sim_{\mathrm{g}} \\ \hat{\mathfrak{A}}, \hat{a} & \overset{\equiv_{q,n}^{(\ell)}}{\rule{3cm}{0.4pt}} & \hat{\mathfrak{B}}, \hat{b} \end{array}$$

Note that the status in finite model theory of the full characterisation result of Andréka, van Benthem, and Németi—Theorem 12 above—remains open, as the present techniques only deal with vocabularies of width two.

Further remarks

Among other related open issues ranks prominently the question whether the characterisation theorem of Janin and Walukiewicz—that the modal μ-calculus precisely captures the bisimulation invariant fragment of monadic second-order logic—is valid also in the sense of finite model theory. The techniques employed here seem to shed no immediate light on this matter.

Other ramifications in the modal domain do seem to be amenable to the techniques developed here. In particular, we mention characterisation theorems in the presence of other natural restrictions, apart from finiteness. Classical and other natural frame conditions can be considered. For the class of connected frames, for instance, preliminary results have been obtained in unpublished communication with A. Dawar. Graded bisimulation and modal logics with graded modalities, incorporating number restrictions similar in spirit to those encountered with our q-back-and-forth requirements in section 4.1, would seem to provide another interesting test case for the present techniques.

Acknowledgement I am grateful for comments from the anonymous referees, which helped to improve the presentation.

References

[1] N. Alon, *Tools from higher algebra*, in Handbook of Combinatorics, R. Graham et al., ed., vol. II, North-Holland, 1995, pp. 1749–1783.

[2] H. Andréka, J. van Benthem, and I. Németi, *Modal languages and bounded fragments of predicate logic*, Journal of Philosophical Logic, 27 (1998), pp. 217–274.

[3] J. Barwise and J. van Benthem, *Interpolation, preservation, and pebble games*, Journal of Symbolic Logic, 64 (1999), pp. 881–903.

[4] P. Blackburn, M. de Rijke, and Y. Venema, *Modal Logic*, Cambridge Tracts in Theoretical Computer Science, Cambridge University Press, 2001.

[5] C. C. Chang and H. J. Keisler, *Model Theory*, North-Holland, 1990.

[6] K. Compton, *Some useful preservation theorems*, Journal of Symbolic Logic, 48 (1983), pp. 427–440.

[7] H.-D. Ebbinghaus and J. Flum, *Finite Model Theory*, Springer, 2nd ed., 1999.

[8] H.-D. Ebbinghaus, J. Flum, and W. Thomas, *Mathematical Logic*, Springer, 1994.

[9] H. Gaifman, *On local and nonlocal properties*, in Logic Colloquium '81, J. Stern, ed., North Holland, 1982, pp. 105–135.

[10] E. Grädel, *On the restraining power of guards*, Journal of Symbolic Logic, 64 (1999), pp. 1719–1742.

[11] E. Grädel, C. Hirsch, and M. Otto, *Back and forth between guarded and modal logics*, ACM Transactions on Computational Logics, 3 (2002), pp. 418 – 463.

[12] I. Hodkinson and M. Otto, *Finite conformal hypergraph covers and Gaifman cliques in finite structures*, Bulletin of Symbolic Logic, 9 (2003). (to appear).

[13] M. Otto, *Bounded Variable Logics and Counting*, Lecture Notes in Logic, Springer, 1997.

[14] ——, *Bisimulation invariance and finite models*, in Colloquium Logicum 2002, Lecture Notes in Logic, ASL, 2002. in preparation.

[15] ——, *Modal and guarded characterisation theorems over finite transition systems*, in Proceedings of 17th Annual IEEE Symposium on Logic in Computer Science LICS '02, 2002, pp. 371–380.

[16] B. Poizat, *A Course in Model Theory*, Springer-Verlag, 2000.

[17] E. Rosen, *Modal logic over finite structures*, Journal of Logic, Language and Information, 6 (1997), pp. 427–439.

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

[19] ——, *Modal Logic and Classical Logic*, Bibliopolis, Napoli, 1983.