Expressive Completeness through Logically Tractable Models

Martin Otto
Department of Mathematics
Technische Universität Darmstadt

November 13, 2012

Abstract

How can we prove that some fragment of a given logic has the power to define precisely all structural properties that satisfy some characteristic semantic preservation condition? This issue is a fundamental one for classical model theory and applications in non-classical settings alike. While methods differ greatly, and while the classical methods can usually not be matched for instance in the setting of finite model theory, this note surveys some interesting commonality revolving around the use and availability of tractable representatives in the relevant model classes. The construction of models in which simple invariants like partial types based on some weak fragment control all the relevant structural properties, may be seen at the heart of such questions. We highlight some constructions involving degrees of acyclicity and saturation that can be achieved in finite model constructions, and discuss their uses towards expressive completeness w.r.t. bisimulation based equivalences in hypergraphs and relational structures. The emphasis is on the combinatorial challenges in such more constructive approaches that work in non-classical settings and especially in finite model theory. One new result concerns expressive completeness w.r.t. guarded negation bisimulation, a back-and-forth equivalence involving local homomorphisms.

Contents

1 Introduction

Model theory explores the logical definability of structural properties, i.e., the expressiveness of logics. Structural properties are here the material against which the semantics of logics under consideration is measured.

Classes of structural properties can be specified, for instance, in terms of other logics, but also in terms of a priori extra-logical considerations like algebraic or recursion-theoretic and algorithmic criteria. In a universal-algebraic tradition, closure conditions are a key example; in relation to the semantics of logical formalisms they are usually described as preservation conditions. In connection with algorithmic criteria, levels of computational complexity carve out natural classes of structural properties.

Logics are specified in terms of syntax and semantics. A logic may or may not provide the expressiveness to deal with certain classes of structural properties. The logics under consideration can be very concrete fragments of established systems, e.g., certain syntactic fragments of first-order logic if one is looking for the expressive means to capture just those first-order properties that satisfy some natural closure property. In different context, e.g., if one is looking for the expressive means to capture all structural properties of a given computational complexity, one may want to cast the net for candidate logics as wide as possible and restrict them by no more than the most rudimentary criteria concerning the manner in which their syntax and semantics are presented.

In any of these situations, precise matches between some class of structural properties and the expressiveness of some logic are particularly attractive. For the logic involved in such a match, one obtains a model-theoretic characterisation of its expressiveness, an answer to the question:

- Which properties precisely can be expressed?

For the class of structural properties involved, one obtains a descriptive characterisation:

- Syntax that captures these, and just these, properties.

Both aspects are of interest for the model-theoretic study of logical semantics. But typically they also have applications beyond the quintessentially model-theoretic interest. In particular, the availability of a logic that precisely captures a given class of properties provides a language for the specification and manipulation of just these properties, which is rich enough to express all intended properties and safeguarded against breaching the underlying semantic constraint. It may in addition provide a tool for surveying and for analysing this class of properties, including for instance the possibility to determine whether a given property is of this kind.

This paper concentrates on one technical aspect shared by various expressive completeness results, from typical classical examples to more recent explorations in finite model theory. The obvious differences have often been stressed: classical expressive completeness results for fragments of first-order logic are typically compactness-based; alternative, more constructive and combinatorial arguments are required in those cases where expressive completeness results can be established in finite model theory. Meanwhile the growing number of qualified expressive completeness results in finite model theory [19, 15, 16, 17, 20, 10, 3, 4, 18] calls for a re-assessment of the earlier primarily negative view that focused on "failures in the finite" in comparison to the well-known classical preservation theorems. Also the major methodological differences may have hidden some interesting commonality that does prevail more often than had first been appreciated.

In this paper I attempt to describe one such common aspect that seems to link techniques across the divide of classical versus finite model theory, which I see in the use of logically tractable models. As a first approximation think of logical tractability w.r.t. some weaker logic L, whose expressive power is to be established, as a criterion that guarantees that, in certain well-behaved structures, L is unusually strong in the sense that L-descriptions of configurations (L-types) determine the behaviour of these configurations w.r.t. some stronger logic or up to some more powerful notion of equivalence than is usually associated with L.

Structures with saturation properties provide striking classical examples of this kind. Consider ω-saturated τ-structures $\mathfrak{A}$ and $\mathfrak{B}$ in some fixed finite relational vocabulary τ. If tuples $\mathbf{a} \in \mathfrak{A}$ and $\mathbf{b} \in \mathfrak{B}$ satisfy exactly the same first-order formulae $\varphi(\mathbf{x}) \in \mathrm{FO}[\tau]$, then, due to ω-saturation, $\mathbf{a}$ and $\mathbf{b}$ are linked by a back-and-forth system of local isomorphisms that establishes that $\mathfrak{A}, \mathbf{a}$ and $\mathfrak{B}, \mathbf{b}$ are related by *partial isomorphy*, $\mathfrak{A}, \mathbf{a} \simeq_{\mathrm{part}} \mathfrak{B}, \mathbf{b}$, and, by Karp's theorem, satisfy exactly the same formulae in the infinitary logic $\mathrm{FO}_\infty[\tau]$.[1] So the class of ω-saturated structures reduces FO_∞-equivalence (partial isomorphy) to FO-equivalence (elementary equivalence), and we may regard ω-saturated models as particularly tractable representatives of their elementary equivalence class, because here the expressive power of FO reaches beyond the usual limits and determines the nature of tuples up to partial isomorphy. Of course, over count-

[1] Here FO_∞ denotes the logic which allows for conjunctions and disjunctions over arbitrary sets of formulae; classical notation is $L_{\infty\omega}$ for FO_∞, in line with $L_{\omega\omega}$ for FO.

able structures, partial isomorphy coincides with actual isomorphy; but while countable, saturated models might look even more desirable, only very special elementary classes possess such representatives.

Outline. After some preliminaries and first examples in this section, we look at forms of acyclicity and finitary saturation in hypergraphs in Sections 2 and 3 before analysing tractability in suitable guarded covers for relational structures in Section 4. The expressive completeness results in Section 4 concern the guarded fragment of first-order logic as well as certain guarded negation fragments, for which the status in finite model theory was unknown. In the final section, Section 5, we outline an expressive completeness result from [5] in the context of descriptive complexity.

1.1 Preliminaries

For later use we fix some terminology and recall the basic notions related to back-and-forth equivalences and types and preservation properties.

In this paper we deal with structures over finite relational vocabularies τ; τ-structures are denoted as in $\mathfrak{A} = (A, (R^{\mathfrak{A}})_{R\in\tau})$, where A is the domain of $\mathfrak{A}$. For n-tuples we write $\mathbf{a} = (a_1, \dots, a_n)$ and often indicate membership of the components a_i in A in suggestive notation like $\mathbf{a} \in A$ or $\mathbf{a} \in \mathfrak{A}$. FO denotes first-order logic, FO$[\tau]$ if we want to indicate the vocabulary; the collection of FO-formulae of quantifier rank up to $q \in \mathbb{N}$ is denoted FO_q. Those variables that may occur free in a formula are highlighted in the notation $\varphi(\mathbf{x})$, which indicates that at most those variable symbols listed in the displayed tuple $\mathbf{x}$ occur free in φ. We write $\mathfrak{A} \models \varphi[\mathbf{a}]$ or $\mathfrak{A}, \mathbf{a} \models \varphi$ to state that φ is satisfied by the assignment of $\mathbf{a}$ to $\mathbf{x}$ in $\mathfrak{A}$. Two formulae are called logically equivalent, $\varphi \equiv \psi$, if they are satisfied by exactly the same structures and assignments. The restricted notion of logical equivalence over some class $\mathcal{C}$ of structures, $\varphi \equiv_{\mathcal{C}} \psi$, only admits structures $\mathfrak{A} \in \mathcal{C}$ to the test whether $\mathfrak{A} \models \varphi[\mathbf{a}]$ iff $\mathfrak{A} \models \psi[\mathbf{a}]$.

Degrees of indistinguishability between structures are formalised as usual: for instance, elementary equivalence of structures, denoted $\mathfrak{A} \equiv \mathfrak{B}$, means that $\mathfrak{A}$ and $\mathfrak{B}$ satisfy exactly the same FO-sentences (formulae without free variables). Degrees of indistinguishability between tuples in structures, or of structures with parameters as assignments to free variables, can be formalised in terms of types. A type is just a collection of formulae that can simultaneously be realised by tuples. For instance, the complete FO-type of a tuple $\mathbf{a} \in \mathfrak{A}$ is $\{\varphi(\mathbf{x}) \in \mathrm{FO} \colon \mathfrak{A}, \mathbf{a} \models \varphi\}$, its FO_q-type $\{\varphi(\mathbf{x}) \in \mathrm{FO}_q \colon \mathfrak{A}, \mathbf{a} \models \varphi\}$. Then $\mathfrak{A}, \mathbf{a} \equiv \mathfrak{B}, \mathbf{b}$ means that $\mathbf{a}$ in $\mathfrak{A}$ and $\mathbf{b}$ in $\mathfrak{B}$ satisfy the same complete FO-type. Approximations to full elementary equivalence are provided by levels $\equiv^q$ of agreement w.r.t. all FO-formulae up to quantifier rank q, i.e., by equality of FO_q-types. Since we assume τ to be finite and relational, the equivalences $\equiv^q$ are of finite index for every fixed quantifier rank level q and fixed width of the parameter tuples in $\mathfrak{A}, \mathbf{a} \equiv^q \mathfrak{B}, \mathbf{b}$. All these notions have natural variants for other logics to be considered below.

Many relevant equivalences between structures with parameter tuples can be cast in terms of model theoretic back-and-forth games. Levels of first-order indistinguishability are captured by the familiar Ehrenfeucht–Fraïssé game or by back-and-forth systems of local isomorphisms. Finite local isomorphisms are here denoted as $\rho\colon \mathbf{a} \mapsto \mathbf{b}$ where tuples $\mathbf{a} \in \mathfrak{A}$ and $\mathbf{b} \in \mathfrak{B}$ enumerate the domain and range of p and $p\colon \mathfrak{A}\restriction \mathbf{a} \simeq \mathfrak{B}\restriction \mathbf{b}$ says that p is a local isomorphism, i.e., an isomorphisms of the induced substructures. Back-and-forth systems of local isomorphisms are characterised by appropriate back-and-forth conditions. For the familiar case of the q-round Ehrenfeucht–Fraïssé game, we deal with a system $(I_m)_{m\leqslant q}$ of sets of local isomorphisms and the back-and-forth conditions require that any $\rho \in I_{m+1}$ has extensions $\rho' \in I_m$ with $b \in \mathrm{image}(\rho')$ (respectively with $a \in \mathrm{dom}(\rho')$), for any choice of $b \in B$ (respectively of $a \in A$). If there is such a system with $(\rho\colon \mathbf{a} \mapsto \mathbf{b}) \in I_q$, then $\mathfrak{A}, \mathbf{a}$ and $\mathfrak{B}, \mathbf{b}$ are called q-isomorphic, $\mathfrak{A}, \mathbf{a} \simeq^q \mathfrak{B}, \mathbf{b}$. By the Ehrenfeucht–Fraïssé theorem, the following are equivalent for any two structures in a finite relational vocabulary τ with parameter tuples:

(i) $\mathfrak{A}, \mathbf{a} \simeq^q \mathfrak{B}, \mathbf{b}$ (q-isomorphism);

(ii) the existence of a winning strategy for the second player, for q rounds played from position $\mathfrak{A}, \mathbf{a}; \mathfrak{B}, \mathbf{b}$, in the familiar Ehrenfeucht–Fraïssé game;

(iii) $\mathfrak{A}, \mathbf{a} \equiv^q \mathfrak{B}, \mathbf{b}$ (indistinguishability in FO_q).

The common refinement $\simeq^\omega$ of all finite levels $(\simeq^q)_{q\in\mathbb{N}}$ correspondingly captures full elementary equivalence, $\mathfrak{A}, \mathbf{a} \simeq^\omega \mathfrak{B}, \mathbf{b}$ if, and only if, $\mathfrak{A}, \mathbf{a} \equiv \mathfrak{B}, \mathbf{b}$. In terms of the game, this corresponds to the existence of a winning strategy for the second player for every chosen finite number of rounds. The stronger notion of partial isomorphy, $\mathfrak{A}, \mathbf{a} \simeq_{\mathrm{part}} \mathfrak{B}, \mathbf{b}$, corresponds to the existence of a winning strategy for the second player in the infinite game, or to the existence of a single set I of local isomorphisms that is closed w.r.t. the back-and-forth conditions, and with $(\rho\colon \mathbf{a} \mapsto \mathbf{b}) \in I$. Karp's theorem may be regarded as the upgrading of the Ehrenfeucht–Fraïssé equivalence from the finite round game to the infinite game, from the limit $\simeq^\omega$ of the finite levels of q-isomorphy $\simeq^q$ to partial isomorphy, and from indistinguishability in FO to indistinguishability in FO_∞. The following are equivalent:

(i) $\mathfrak{A}, \mathbf{a} \simeq_{\mathrm{part}} \mathfrak{B}, \mathbf{b}$ (partial isomorphy);

(ii) the existence of a winning strategy for the second player in the infinite game from position $\mathfrak{A}, \mathbf{a}; \mathfrak{B}, \mathbf{b}$;

(iii) $\mathfrak{A}, \mathbf{a} \equiv^\infty \mathfrak{B}, \mathbf{b}$ (indistinguishability in FO_∞).

The point of ω-saturation in the brief remark above is precisely that, for ω-saturated $\mathfrak{A}$ and $\mathfrak{B}$, the system I of all finite local isomorphisms $\rho\colon \mathbf{a} \mapsto \mathbf{b}$ with $\mathfrak{A}, \mathbf{a} \equiv \mathfrak{B}, \mathbf{b}$ (finite elementary maps) is closed w.r.t. back-and-forth conditions and hence $\mathfrak{A}, \mathbf{a} \simeq_{\mathrm{part}} \mathfrak{B}, \mathbf{b}$ whenever $\mathfrak{A}, \mathbf{a} \equiv \mathfrak{B}, \mathbf{b}$.

We shall encounter typical variants for fragments and other logics below.

Preservation properties for formulae of some logic or syntactic fragment formulate a semantic criterion in terms of structural equivalences or transformations. We give the definition for some arbitrary notion of structural equivalence $\mathfrak{A}, \mathbf{a} \approx \mathfrak{B}, \mathbf{b}$ between τ-structures with parameters. This is the case we shall be concerned with later, but the concept obviously extends also to non-symmetric relationships like the homomorphism relation $\mathfrak{A}, \mathbf{a} \xrightarrow{\text{hom}} \mathfrak{B}, \mathbf{b}$ which is defined in terms of the existence of a homomorphism from $\mathfrak{A}$ to $\mathfrak{B}$ that maps $\mathbf{a}$ to $\mathbf{b}$. Instead of a preservation property of a formula φ one may always think of a closure property of the underlying class of models of φ. For instance, preservation of φ under homomorphisms is equivalent to closure of the model class of φ under homomorphisms.

Definition 1.1. For any relation between structures, like a notion of equivalence $\mathfrak{A}, \mathbf{a} \approx \mathfrak{B}, \mathbf{b}$ between τ-structures with parameters, we say that a formula $\varphi(\mathbf{x})$ is *preserved under the relation* $\approx$ if the following implication is satisfied for all $\mathfrak{A}, \mathbf{a}$ and $\mathfrak{B}, \mathbf{b}$

$$\mathfrak{A}, \mathbf{a} \approx \mathfrak{B}, \mathbf{b} \text{ and } \mathfrak{A} \models \varphi[\mathbf{a}] \quad \Longrightarrow \quad \mathfrak{B} \models \varphi[\mathbf{b}].$$

If in restriction to a class of structures $\mathcal{C}$ we merely admit structures $\mathfrak{A}, \mathfrak{B}$ from $\mathcal{C}$ to the test, we say that φ is preserved under $\approx$ over $\mathcal{C}$.

Note that preservation under $\approx$ over $\mathcal{C}$ will typically be a strictly weaker condition than preservation under $\approx$ throughout.

The following definition of *expressive completeness* refers to some class of properties of structures or parameter tuples in structures – the kind of properties that may or may not be definable with sentences or formulae in a logic. Formally, we may identify a "property" with an isomorphism closed class of structures (with parameters), and a class of properties with a class of such classes.

Definition 1.2. A logic (or fragment of a logic) L is said to be *expressively complete* for a class $\mathcal{P}$ if every property in $\mathcal{P}$ is defined by some L-formula.

If also every model class of a formula in L belongs to $\mathcal{P}$, we say that L *captures* $\mathcal{P}$.

Classical model-theoretic preservation theorems are instances of this: they state that some syntactic fragment of FO satisfies some preservation property and that this fragment is expressively complete for the class of all FO-definable properties that satisfy this preservation condition. The real content is usually in the expressive completeness part.

1.2 Two examples

Homomorphism preservation. It is well-known and easy to check that all first-order formulae generated from atomic formulae by just conjunction, disjunction and existential quantification (i.e., without negation and universal quantification) are preserved under homomorphisms. This *existential positive fragment* of FO is here denoted ∃posFO. The question is immediate whether this is a precise match. The classical Lyndon–Tarski theorem is a typical representative of a whole family of corresponding "preservation theorems."

Theorem 1.3 (Lyndon–Tarski)**.** *The existential positive fragment of* FO *captures the class of all* FO*-definable properties that are closed under homomorphisms. I.e., the following are equivalent for* $\varphi(\mathbf{x}) \in \mathrm{FO}[\tau]$*:*

(i) φ *is preserved under homomorphisms:*
whenever $\mathfrak{A}, \mathbf{a} \xrightarrow{\text{hom}} \mathfrak{B}, \mathbf{b}$*, then* $\mathfrak{A} \models \varphi[\mathbf{a}]$ *implies* $\mathfrak{B} \models \varphi[\mathbf{b}]$*;*

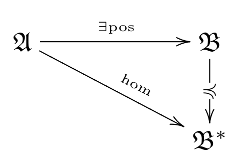
(ii) φ *is logically equivalent to some existential positive formula:*
there is some $\varphi' \in \exists\text{posFO}[\tau]$ *such that* $\varphi \equiv \varphi'$*.*

Note that, as a corollary, we get effective syntax for a class of first-order properties of interest. The class of all first-order formulae that are preserved under homomorphisms, on the other hand, is easily seen to be undecidable.

The interesting direction in the above equivalence is the claim of *expressive completeness* of ∃posFO for the class of all first-order definable properties that are preserved under homomorphisms.

Let us consider this classical expressive completeness assertion also with a view to illustrate an essential proof technique that can be described as an *upgrading* of relations between structures. For simplicity we drop parameters and speak of FO-sentences, which is essentially no loss of generality. The task is to show that a sentence $\varphi \in \text{FO}$ that is preserved under homomorphisms is equivalently expressible in ∃posFO. A standard compactness-based preparation (cf. Lemma 3.2.1. in [9]) reduces this claim to the assertion that such φ is preserved under the relation of *positive existential transfer*, $\mathfrak{A} \overset{\exists\text{pos}}{\Longrightarrow} \mathfrak{B}$ defined by the condition that for every $\psi \in \exists\text{posFO}$, $\mathfrak{A} \models \psi$ implies $\mathfrak{B} \models \psi$. Unlike $\xrightarrow{\text{hom}}$, the relation $\overset{\exists\text{pos}}{\Longrightarrow}$ is defined in terms of logic. If we also invoke the Löwenheim–Skolem property of FO, it suffices to establish preservation of φ under $\overset{\exists\text{pos}}{\Longrightarrow}$ in restriction to countable structures.

By preservation of ∃posFO under homomorphisms, $\xrightarrow{\text{hom}}$ implies $\overset{\exists\text{pos}}{\Longrightarrow}$. It would be trivial to show that preservation under homomorphisms implies preservation under ∃posFO transfer, if conversely $\mathfrak{A} \overset{\exists\text{pos}}{\Longrightarrow} \mathfrak{B}$ generally implied $\mathfrak{A} \xrightarrow{\text{hom}} \mathfrak{B}$. This is clearly not the case: $\xrightarrow{\text{hom}}$ is strictly stronger than $\overset{\exists\text{pos}}{\Longrightarrow}$. But this difference disappears for a countable model $\mathfrak{A}$ and for an ω-saturated $\mathfrak{B}$. If we replace the given $\mathfrak{B}$ in $\mathfrak{A} \overset{\exists\text{pos}}{\Longrightarrow} \mathfrak{B}$ by some ω-saturated elementary extension $\mathfrak{B}^*$ (not necessarily countable), for which in particular $\mathfrak{B}^* \equiv \mathfrak{B}$, then $\mathfrak{A} \overset{\exists\text{pos}}{\Longrightarrow} \mathfrak{B}$ automatically implies $\mathfrak{A} \overset{\exists\text{pos}}{\Longrightarrow} \mathfrak{B}^*$, and thus gets upgraded to $\mathfrak{A} \xrightarrow{\text{hom}} \mathfrak{B}^*$. Now $\mathfrak{A} \models \varphi$ implies that $\mathfrak{B}^* \models \varphi$ since φ is preserved under homomorphisms; and this implies that also $\mathfrak{B} \models \varphi$, because $\mathfrak{B} \equiv \mathfrak{B}^*$.



So the detour through $\mathfrak{B}^*$ establishes that φ is indeed preserved under $\overset{\exists\text{pos}}{\Longrightarrow}$, and the expressive completeness claim follows. What we use is the fact that the class of ω-saturated models

– is rich enough to provide representatives of every elementary class
(we need a representative in the $\equiv$-class of the given $\mathfrak{B}$)

– is well-behaved or tractable for the intended upgrading: every ∃posFO-preserving finite partial map into an ω-saturated target structure can be extended in an ∃posFO-preserving manner.[2]

The second assertion is due to the fact that an ω-saturated structure realises every satisfiable ∃posFO-type over finitely many parameters, and reflects the same phenomenon that was mentioned in connection with elementary equivalence and partial isomorphy above.

The Lyndon–Tarski theorem is one of those classical expressive completeness results that do hold true also in the sense of finite model theory. This long open problem was solved by Benjamin Rossman [20], with very sophisticated finitary saturation methods. These also fall in the category of constructive finitary counterparts of classical saturation arguments that we are interested in.

We shall use the following notion to describe how, over certain classes of well-behaved or tractable structures, a weaker logic suffices to determine the behaviour of finite configurations w.r.t. a stronger logic. The same notion can naturally be applied to induced equivalences between configurations.

Definition 1.4. For a class of structures $\mathcal{C}$ and two logics[3] L_0 and L, we say that L_0 *controls* L *over* $\mathcal{C}$, in shorthand notation: $\mathrm{L}_0 \rhd \mathrm{L}$ over $\mathcal{C}$, if L_0-types determine L-types over $\mathcal{C}$ in the sense that, for $\mathfrak{A} \in \mathcal{C}$ and tuples $\mathbf{a}, \mathbf{a}' \in \mathfrak{A}$, always $\mathfrak{A}, \mathbf{a} \equiv_{\mathrm{L}_0} \mathfrak{A}, \mathbf{a}' \;\Rightarrow\; \mathfrak{A}, \mathbf{a} \equiv_{\mathrm{L}} \mathfrak{A}, \mathbf{a}'$. We also say that L_0-equivalence determines L-equivalence: $\equiv_{\mathrm{L}_0} \rhd \equiv_{\mathrm{L}}$ over $\mathcal{C}$.

The version for logical equivalences will be the one to apply in upgrading steps towards expressive completeness arguments. In finite model theory arguments, compactness is not available to pass to a smooth and uniform limit. We shall use the finite index approximations of the logical equivalences, which also lead to more concrete constructive assertions. For instance,

$$\mathrm{L}_{f(q)} \rhd \mathrm{FO}_q \text{ over } \mathcal{C}$$

is used to say that, in order to determine the first-order behaviour of configurations up to quantifier rank q, it suffices to control their nesting depth $f(q)$ types in the fragment L.

Bisimulation preservation. This example illustrates these ideas in a less classical setting, although not one of finite model theory. It involves preservation under *bisimulation* and logics of a modal character. It deals with a non-elementary class of *tree models*, which here arise as tractable representatives obtained in a natural process of tree unfolding. The example takes us closer to the technical concerns in the main part of the paper, insofar it involves acyclicity criteria for the relevant class of well-behaved models.

[2] In case of a countable source structure $\mathfrak{A}$, a countable chain of extension steps based on finite partial homomorphisms generates an actual homomorphism in the limit.

[3] We do not assume any closure properties for logics; otherwise we should refer to fragments of logics here.

We look at finite relational vocabularies τ of width 2, i.e., with monadic and binary relations. A τ-structure is an edge- and vertex-coloured graph, or a labelled transition system. Bisimulation equivalence between labelled transition systems is the natural equivalence based on the infinite version of the modal back-and-forth game. A single pebble for each structure is moved along labelled transitions and monadic relations need to be respected by pebble positions. A bisimulation relation between $\mathfrak{A}$ and $\mathfrak{B}$ describes a (not necessarily deterministic) winning strategy for the second player. Formally, a bisimulation relation is formalised as a non-empty relation $Z \subseteq A \times B$ satisfying the following.

(a) *local isomorphy* w.r.t. all monadic $P \in \tau$:
for all $(a, b) \in Z$, $a \in P^{\mathfrak{A}} \Leftrightarrow b \in P^{\mathfrak{B}}$;

(b) *back-and-forth* conditions for each binary $E \in \tau$:
(*back*) for all $(a, b) \in Z$ and for every b' with $(b, b') \in E^{\mathfrak{B}}$ there is some a' with $(a, a') \in E^{\mathfrak{A}}$ such that $(a', b') \in Z$;
(*forth*) for all $(a, b) \in Z$ and for every a' with $(a, a') \in E^{\mathfrak{A}}$ there is some b' with $(b, b') \in E^{\mathfrak{B}}$ such that $(a', b') \in Z$.

We write $\mathfrak{A}, a \sim \mathfrak{B}, b$ to assert the existence of a bisimulation relation Z as above with $(a, b) \in Z$, which encodes a winning strategy for the second player in the underlying infinite bisimulation game from position $\mathfrak{A}, a; \mathfrak{B}, b$. The existence of a winning strategy in the q-round game, which would be formalised by a back-and-forth system of the form $(Z_m)_{m \leqslant q}$ in the usual manner, is denoted as $\mathfrak{A}, a \sim^q \mathfrak{B}, b$. Note that, by design, the finite levels $\sim^q$, their common refinement $\sim^\omega$, and full bisimilarity $\sim$, are in exactly the same relationship as are $\simeq^q$, $\simeq^\omega$ and $\simeq_{\text{part}}$ in the classical first-order setting.

	q-round game	finite-round game	infinite game
calssical:	$\simeq^q$	$\simeq^\omega$	$\simeq_{\text{part}}$
modal:	$\sim^q$	$\sim^\omega$	$\sim$

Conditions (a) and (b) characterise the Ehrenfeucht–Fraïssé rules that match the expressive power of basic modal logic ML. This fragment of first-order logic has formulae with a single free variable; its quantifier-free formulae make assertions just about the monadic relations, so that (a) reflects quantifier-free equivalence; and quantifications are of the relativised form $\forall y(Exy \rightarrow \varphi(y))$ and $\exists y(Exy \wedge \varphi(y))$, which is captured in the back-and-forth conditions (b). We write ML_q for the fragment of modal logic of quantifier rank up to q, and $\mathfrak{A}, a \equiv^q_{\text{ML}} \mathfrak{B}, b$ for indistinguishability in ML_q. The modal variants of the Ehrenfeucht–Fraïssé and Karp theorems are immediate.

The following are equivalent:

(i) $\mathfrak{A}, a \sim^q \mathfrak{B}, b$ (q-bisimilarity);

(ii) the existence of a winning strategy for the second player in the q-round bisimulation game from position $\mathfrak{A}, a; \mathfrak{B}, b$;

(iii) $\mathfrak{A}, a \equiv^q_{\text{ML}} \mathfrak{B}, b$ (indistinguishability in ML_q).

Considering the infinitary variant ML_∞ of ML with conjunctions and disjunctions over arbitrary sets of formulae, the following are equivalent:

(i) $\mathfrak{A}, a \sim \mathfrak{B}, b$ (bisimilarity);

(ii) the existence of a winning strategy for the second player in the infinite bisimulation game from position $\mathfrak{A}, a; \mathfrak{B}, b$;

(iii) $\mathfrak{A}, a \equiv^{\infty}_{\mathrm{ML}} \mathfrak{B}, b$ (indistinguishability in ML_∞).

Countable tree models in which all relevant multiplicities are boosted to be infinite, are tractable models for our purposes: in these, the bisimulation behaviour even determines the isomorphism type. Such well-behaved models can be obtained as representatives within every bisimulation class, through the simple process of ω-tree-unfolding.

Let $\mathfrak{A} = (A, (E^{\mathfrak{A}})_{E \in \tau(2)}, (P^{\mathfrak{A}})_{P \in \tau(1)})$ be a τ-structure, where we split the finite vocabulary τ into $\tau(1)$ and $\tau(2)$ according to arities. For any chosen parameter $a \in \mathfrak{A}$ as the distinguished root vertex, let the ω-tree-unfolding of $\mathfrak{A}$ from a be the following tree-like τ-structure $\mathfrak{A}^{\omega *}_a$. The domain of $\mathfrak{A}^{\omega *}_a$ is the set of all $(\tau(2) \times \mathbb{N})$-edge-labelled paths from $a_0 := a$ in $\mathfrak{A}$:

$$\rho = a_0(E_1, m_1)a_1(E_2, m_2)a_2 \dots (E_n, m_n)a_n$$

where $n \in \mathbb{N}$ is the length of this path, $m_j \in \mathbb{N}$ for $1 \leqslant j \leqslant n$, and the label components $E_j \in \tau(2)$ are such that $(a_{j-1}, a_j) \in E^{\mathfrak{A}}_j$ for $1 \leqslant j \leqslant n$. The binary relations $E \in \tau(2)$ are interpreted naturally as the sets of all pairs (ρ, ρ') of such paths where ρ' extends ρ by one (E, m)-step according to $\rho' = \rho(E, m)a'$. The unary $P \in \tau(1)$ are interpreted such that the projection to the target vertex

$$\pi \colon a_0(E_1, m_1)a_1 \dots (E_n, m_n)a_n \mapsto a_n$$

is a homomorphism. We identify the path of length 0 from a with a itself, which thus becomes the distinguished root vertex also of the resulting tree structure. So $\pi \colon \mathfrak{A}^{\omega *}_a, a \xrightarrow{\text{hom}} \mathfrak{A}, a$, and it is easy to check that the graph of the projection π is a bisimulation relation. It follows that in particular $\mathfrak{A}, a \sim \mathfrak{A}^{\omega *}, a$.

Observation 1.5. *If $\mathfrak{A}$ and $\mathfrak{B}$ are countably branching, then $\mathfrak{A}, a \sim \mathfrak{B}, b$ implies $\mathfrak{A}^{\omega *}_a, a \simeq \mathfrak{B}^{\omega *}_b, b$. Generally also $(\mathfrak{A}^{\omega *}_a)^{\omega *}_a \simeq \mathfrak{A}^{\omega *}_a$.*

The model theory of tree structures is a classical example of a model theory closely linked to algorithmic methods. Key results like Rabin's theorem on the decidability of monadic second-order logic MSO over the full binary tree, and, as a consequence, decidability of the satisfiability problem for MSO over tree structures, are driven by game arguments, (de-)composition techniques and automata theoretic methods. (MSO is the extension of first-order logic by second-order quantification over subsets of the domain; over tree structures, MSO provides the expressive power to speak about subtrees, initial segments, reachability and paths, infinite paths, well-foundedness, etc.)

An interesting fragment of MSO (over trees as well as over arbitrary transition systems) is the *modal μ-calculus* L_μ; it extends basic modal logic ML by least and greatest fixed point operators that provide a natural mechanism for monadic monotone recursion and inductive definitions. L_μ has the expressive means, for instance, to make reachability assertions, well-foundedness assertions, etc. But L_μ is also a logic of modal character in the sense of being preserved under bisimulation equivalence – a property clearly not shared by MSO or even FO (over trees), since both can count finite multiplicities and, for instance, make assertions about finite bounds on the branching degree in a tree.

Over tree models, the expressive powers of both MSO and L_μ can be matched to the computational powers of suitable kinds of tree automata. Using such matches, Janin and Walukiewicz showed in their fundamental paper [14] that MSO and L_μ are equally expressive over the class of ω-tree-unfoldings of arbitrary transition systems. Intuitively: the one obvious difference in expressive power is related to the ability to control finite multiplicities; this is essentially the only advantage that MSO has over L_μ in tree models; and this advantage is systematically denied in tree structures in which all multiplicities are infinite. Because ω-tree-unfoldings represent any transition system up to bisimulation equivalence, this collapse (indeed, a tractability phenomenon) entails a celebrated expressive completeness result.

Theorem 1.6 (Janin–Walukiewicz)**.** *The modal μ-calculus captures the class of all* MSO*-definable properties (of elements in transition systems) that are closed under bisimulation equivalence. In other words, the following are equivalent for any formula $\varphi(x) \in \mathrm{MSO}(\tau)$:*

(i) *$\varphi(x)$ is preserved under bisimulation equivalence: whenever $\mathfrak{A}, a \sim \mathfrak{B}, b$, then $\mathfrak{A} \models \varphi[a]$ implies $\mathfrak{B} \models \varphi[b]$;*

(ii) *$\varphi(x)$ is logically equivalent to some $\mathrm{L}_\mu[\tau]$-formula.*

We briefly discuss those generic aspects of the proof that are relevant for our tractability considerations. It is interesting to note that compactness is not available for the logics at hand, which means that neither can classical saturation arguments be brought to bear, nor even the usual preparation that would relate expressibility in L_μ to some transfer property for L_μ-formulae (cf. the above discussion of the Lyndon–Tarski theorem).

Let $\mathcal{T}^\omega[\tau]$ stand for the class of all ω-tree-unfoldings of τ-transition systems. Essentially, the proof of Janin and Walukiewicz uses the automata theoretic reformulation to show that, for some suitable function f,

$$(*) \qquad \mathrm{L}_\mu^{f(q)} \triangleright \mathrm{MSO}^q \quad \text{over } \mathcal{T}^\omega[\tau].$$

We check that this implies the desired expressive completeness of L_μ for MSO-definable properties that are preserved under $\sim$.

Let $\varphi(x) \in \mathrm{MSO}[\tau]^q$ be preserved under $\sim$. Then $(*)$ implies that φ is equivalent to some formula of $\mathrm{L}_\mu^{f(q)}$ in restriction to the class $\mathcal{T}^\omega[\tau]$: $\varphi \equiv_{\mathcal{T}^\omega} \varphi' \in \mathrm{L}_\mu^{f(q)}$. This is just because indistinguishability w.r.t. L_μ^m is an equivalence relation of finite index over the class of all τ-structures (recall that τ is finite), whose equivalence classes are therefore L_μ^m-definable. But equivalence over $\mathcal{T}^\omega[\tau]$ implies equivalence throughout, by $\sim$-preservation and because – up to bisimulation equivalence – any τ-structure is represented by its ω-tree-unfolding in $\mathcal{T}^\omega[\tau]$:

$$\begin{array}{rrl}
 & \mathfrak{A} \models \varphi[a] & \\
\Leftrightarrow & \mathfrak{A}_a^{\omega *} \models \varphi[a] & \text{since } \varphi \text{ is preserved in } \mathfrak{A}, a \sim \mathfrak{A}_a^{\omega *}, a; \\
\Leftrightarrow & \mathfrak{A}_a^{\omega *} \models \varphi'[a] & \text{since } \mathfrak{A}_a^{\omega *} \in \mathcal{T}^\omega[\tau]; \\
\Leftrightarrow & \mathfrak{A} \models \varphi'[a] & \text{since } \varphi' \in \mathrm{L}_\mu \text{ is preserved in } \mathfrak{A}_a^{\omega *}, a \sim \mathfrak{A}, a.
\end{array}$$

For L_μ-definability we seem to need the more fine-grained control in $(*)$, not just a blanket $\mathrm{L}_\mu \rhd \mathrm{MSO}$. This is because, in the absence of compactness, definability in L does not follow from preservation under $\equiv_{\mathrm{L}}$. We may also gain some more constructive information from the knowledge of the function f, which gives an upper bound on the increase in quantifier complexity in the passage from $\varphi \in \mathrm{MSO}$ to its equivalent in L_μ. These features are characteristic of several finite model theory adaptations of classical expressive completeness proofs.

Interestingly, the status of the Janin–Walukiewicz expressive completeness result in finite model theory is still open. This is different for the counterpart at the level of FO rather than MSO.

Theorem 1.7 (van Benthem–Rosen). *Both, in the sense of ordinary and of finite model theory, basic modal logic* ML *captures the class of all* FO*-definable properties of transition systems that are closed under bisimulation equivalence. In other words, the following are equivalent for any formula* $\varphi(x) \in \mathrm{FO}(\tau)$*:*

(i) $\varphi(x)$ *is preserved under bisimulation equivalence (among finite structures): whenever* $\mathfrak{A}, a \sim \mathfrak{B}, b$ *(for finite* $\mathfrak{A}$, $\mathfrak{B}$*), then* $\mathfrak{A} \models \varphi[a]$ *implies* $\mathfrak{B} \models \varphi[b]$*;*

(ii) $\varphi(x)$ *is logically equivalent (over finite structures) to some* $\varphi' \in \mathrm{ML}[\tau]$.

Different accounts of proofs that offer various generalisations and also allow us to extract sharp exponential bounds on the increase in quantifier rank in the passage form $\varphi \in \mathrm{FO}$ to its equivalent $\varphi' \in \mathrm{ML}$, can be found in [17, 16, 10].

$$\begin{array}{ccc}
\mathfrak{A}, a & \text{———} \sim^{f(q)} \text{———} & \mathfrak{B}, b \\
| & & | \\
\sim & & \sim \\
| & & | \\
\mathfrak{A}^*, a & \text{———} \equiv^{q} \text{———} & \mathfrak{B}^*, b \quad \in \mathcal{C}
\end{array}$$

One salient feature of those accounts is the use of locally acyclic (tree-like) *finite* unfoldings of finite transition systems, accompanied by a finite boost in multiplicities. Over a corresponding class $\mathcal{C}$ of (finite) target structures we have

$$\left.\begin{array}{rl}
\mathrm{ML}_{f(q)} \rhd & \mathrm{FO}_q \\
\sim^{f(q)} \rhd & \equiv^q
\end{array}\right\} \quad \text{over } \mathcal{C},$$

where $f(q) = 2^q - 1$. Similar to $(*)$, this serves to establish expressive completeness. Note that usable representatives in $\mathcal{C}$ *must* avoid distinctions both w.r.t. *short cycles* and w.r.t. *small multiplicities*. Neither feature is determined by the bisimulation type of a given structure, but corresponding distinctions are expressible in FO. For instance, no degree of bisimulation equivalence between $\mathfrak{A}, a$ and $\mathfrak{B}, b$ would rule out that a is part of a $(q+1)$-cycle in $\mathfrak{A}$ or has out-degree $\geqslant q$, while b is not part of such a cycle or has out-degree $< q$. But in these situations $\mathfrak{A}, a$ and $\mathfrak{B}, b$ would be distinguishable in FO_q; the second player would win the q-round Ehrenfeucht–Fraïssé game from position $\mathfrak{A}, a; \mathfrak{B}, b$.

See Example 2.17 below for the elimination of short cycles. Degrees of acyclicity and of saturation w.r.t. multiplicities will also concern us in the following sections, where we deal with preservation under a bisimulation-like equivalence between more complex relational structures than transition systems. At the superficial level, this generalisation takes us from relational structures of width 2 (edge labels and vertex colours in graph-like structures) to structures with finitely many relations of arbitrary arities. Non-trivial overlaps between tuples in those relations replace the much simpler notion of incidence between edges and vertices in graphs. It turns out that hypergraphs provide a suitable level of abstraction for the formulation and analysis of acyclicity and saturation criteria in these settings. The next section introduces the relevant concepts and terminology.

2 Degrees of acyclicity in hypergraphs

2.1 Hypergraphs

With a τ-structure $\mathfrak{A} = (A, (R^{\mathfrak{A}})_{R \in \tau})$ in some fixed finite relational vocabulary τ we associate a hypergraph $\mathrm{H}(\mathfrak{A})$ on A whose hyperedges are the *guarded subsets* of $\mathfrak{A}$.

To facilitate the passage from a tuple to its set of components we use the notation $[\mathbf{a}] := \{a_i \colon 1 \leqslant i \leqslant r\}$ if $\mathbf{a} = (a_1, \dots, a_r)$ is an r-tuple.

Definition 2.1. A subset $s \subseteq A$ is *guarded* in $\mathfrak{A}$ if $|s| \leqslant 1$ or if $s \subseteq [\mathbf{a}]$ for some tuple $\mathbf{a} \in R^{\mathfrak{A}}$ for some $R \in \tau$.

The *hypergraph of guarded subsets* of $\mathfrak{A}$ is the hypergraph $\mathrm{H}(\mathfrak{A}) := (A, S[\mathfrak{A}])$ where $S[\mathfrak{A}] := \{s \subseteq A \colon s \text{ guarded in } \mathfrak{A}\}$ is the set of all guarded subsets of $\mathfrak{A}$.

Generally, a *hypergraph* $\mathrm{H} = (A, S)$ consists of a domain A and some collection S of subsets of this domain as the set of hyperedges. Its *width* is the maximum of the cardinalities of the hyperedges (if it exists).

Proviso 2.2. *For our treatment we always assume that* $\bigcup S \supseteq A$ *and that* S *is closed under subsets:* $s' \subseteq s \in S$ *implies* $s' \in S$.

This is clearly the case for the hypergraphs $\mathrm{H}(\mathfrak{A})$ of guarded subsets of τ-structures $\mathfrak{A}$, by definition. Also by definition, the width of $\mathrm{H}(\mathfrak{A})$ is bounded by the width of τ, i.e., by the maximal arity of relations in τ.

The induced sub-hypergraph on some subset A' of the domain A of a hypergraph $\mathrm{H} = (A, S)$ is obtained through the natural restriction:

$$\mathrm{H} \restriction A' = (A', S') \quad \text{where} \quad S' := \{s \cap A' \colon s \in S\}.$$

Together with any hypergraph H we consider two graphs: its *Gaifman graph* $\mathfrak{G}(\mathrm{H})$ and its intersection graph $\mathfrak{I}(\mathrm{H})$:

Definition 2.3. For a hypergraph $\mathrm{H} = (A, S)$ we define

(a) the associated *Gaifman graph* to be the undirected graph

$$\mathfrak{G}(\mathrm{H}) := (A, E) \quad \text{where} \quad E := \{(a, a') \in A^2 \colon a \neq a' \text{ and } \{a, a'\} \in S\}.^4$$

(b) the *intersection graph* to be the undirected graph

$$\mathfrak{I}(\mathrm{H}) := (S, \Delta) \quad \text{where} \quad \Delta := \{(s, s') \in S^2 \colon s \neq s' \text{ and } s \cap s' \neq \emptyset\}.$$

The reason for the terminology in (a) is that the usual Gaifman graph $\mathfrak{G}(\mathfrak{A})$ of a relational structure $\mathfrak{A}$ is exactly what we here obtain as $\mathfrak{G}(\mathrm{H}(\mathfrak{A}))$.

Note that passage to the hypergraph of guarded subsets or to the Gaifman graph is not generally compatible with the passage to induced substructures, i.e., it may be that $\mathrm{H}(\mathfrak{A} \restriction A') \neq \mathrm{H}(\mathfrak{A}) \restriction A'$ and $\mathfrak{G}(\mathfrak{A} \restriction A') \neq \mathfrak{G}(\mathfrak{A}) \restriction A'$; on the other hand always $\mathfrak{G}(\mathrm{H} \restriction A') = (\mathfrak{G}(\mathrm{H})) \restriction A'$.

The associated graph $\mathfrak{G}(\mathrm{H})$ can be pictured as the superposition of cliques, one for every hyperedge of H. Since not every clique in the associated graph needs to be induced by a single hyperedge, though, H cannot in general be retrieved from $\mathfrak{G}(\mathrm{H})$ and in this sense holds strictly more information. Interestingly, this discrepancy in information content between $\mathfrak{G}(\mathrm{H})$ and H exactly corresponds to the failure of *conformality* – one of the two constituents of hypergraph acyclicity to be reviewed next.

As an intuitive approximation, which will be made technically precise later, hypergraph acyclicity calls for the minimisation of coincidental overlaps between hyperedges that is compatible with the transition pattern laid down in the intersection graph.

2.2 Hypergraph acyclicity

Of the several notions of hypergraph acyclicity in the literature we are here interested in the strongest, called α-acyclicity or just *acyclicity*, see [7, 8]. This notion has a number of equivalent characterisations, which we briefly review.

For the first of these characterisations (Graham's decomposition algorithm) we need to waive the proviso that the collection of hyperedges must be closed under subsets. This proviso will be in place everywhere else, and this first characterisation will only serve as background and context. Technically we shall rely exclusively on the other, equivalent characterisations, especially the third, see Definition 2.6 in combination with Lemma 2.7.

Definition 2.4. A finite hypergraph $\mathrm{H} = (A, S)$ is called *tree-decomposable* if it can be transformed into the empty hypergraph by some sequence of applications of the following two reduction steps:

[4] This simple definition is adequate because S is assumed to be closed under subsets.

(i) deletion of $a \in A$ for a vertex a contained in at most one hyperedge. Formally, if $|\{s \in S \colon a \in s\}| \leqslant 1$, then we may replace $\mathrm{H} = (A, S)$ by the induced sub-hypergraph $\mathrm{H}' := \mathrm{H} \restriction (A \setminus \{a\})$.

(ii) deletion of $s' \in S$ for some $s' \subseteq s \in S$.

By extension, an infinite hypergraph is tree-decomposable if all its finite induced sub-hypergraphs are tree-decomposable.

The above characterisation is easily seen to be equivalent to the existence of a hypergraph tree-decomposition in the following sense. Its relationship with the usual tree-decompositions of graphs and relational structures, which is an important tractability criterion for many algorithmic issues, is close but subtly restrictive: a hypergraph tree-decomposition may only use the existing hyperedges as bags.

Definition 2.5. A *tree decomposition* of a hypergraph $\mathrm{H} = (A, S)$ is an acyclic graph $\mathfrak{T} = (V, E)$ together with a surjective map $\lambda \colon T \to S$ such that for every $a \in A$, the vertex set $\{v \in V : a \in \lambda(v)\}$ is connected in $\mathfrak{T}$.

Less obvious is the equivalence of hypergraph acyclicity with the combination of the following two local conditions. Recall from graph theory that a *cycle* of length n in a graph $\mathfrak{G}$ is a homomorphism from the standard cycle of length n, $\mathfrak{C}_n = (\mathbb{Z}_n, \{(i, i \pm 1) \colon i \in \mathbb{Z}_n\})$[5] to $\mathfrak{G}$. If $h \colon \mathfrak{C}_n \to \mathfrak{G}$ is a cycle of some length $n > 3$, then the pair (i, j) is a *chord* of this cycle if $j \neq i \pm 1$ and $\{h(i), h(j)\}$ is guarded (in our liberal formalisation this may be either because $h(i) = h(j)$ or because $(h(i), h(j))$ is an edge of $\mathfrak{G}$). The cycle $h \colon \mathfrak{C}_n \to \mathfrak{G}$ is *chordless* if it does not have any chords; one checks that this means that the homomorphism h is an isomorphism between $\mathfrak{C}_n$ and the induced subgraph $\mathfrak{G} \restriction h(\mathbb{Z}_n)$. A *clique* of size n in $\mathfrak{G}$ is an induced subgraph that is isomorphic to the complete graph on $\mathbb{Z}_n$, $\mathfrak{K}_n = (\mathbb{Z}_n, \{(i, j) \colon i \neq j\})$.

Definition 2.6. A hypergraph $\mathrm{H} = (A, S)$ is called

(a) *conformal* if every clique in its Gaifman graph $\mathfrak{G}(\mathrm{H})$ is fully contained in some hyperedge $s \in S$.

(b) *chordal* if its Gaifman graph $\mathfrak{G}(\mathrm{H})$ has *no* chordless cycles of lengths greater than 3.

Note that chordality is entirely phrased in terms of the associated graph $\mathfrak{G}(\mathrm{H})$ while conformality addresses the potential discrepancy between H and $\mathfrak{G}(\mathrm{H})$. It is precisely for conformal hypergraphs that H can be uniquely reconstructed from $\mathfrak{G}(\mathrm{H})$. See [7] for a proof of the following.

Lemma 2.7. *A hypergraph is tree-decomposable in the sense of Definition 2.4 (admits a tree decomposition in the sense of Definition 2.5) if, and only if, it is conformal and chordal.*

[5]All arithmetic in $\mathbb{Z}_n$ is to be understood modulo n.

2.3 Hypergraph bisimulation

The idea of hypergraph bisimulation is a natural extension of the notion of bisimulation to a back-and-forth equivalence between hypergraphs. Where bisimulation of transition systems ensures local similarity between vertices in terms of the pattern of available transitions to other vertices, hypergraph bisimulation ensures local similarity between hyperedges in terms of the overlap pattern with other hyperedges. Correspondingly, the game formulation would involve moves between hyperedges that need to respect overlaps. The back-and-forth systems that capture the existence of a winning strategy for the second player in the infinite game consist of collections of local bijections between hyperedges with back-and-forth conditions reflecting overlap respecting moves.

Where we consider local bijections between sets A and B, as in $\rho\colon \mathrm{dom}(\rho) \to \mathrm{image}(\rho)$, we often indicate their domain and range and/or tuples that enumerate these sets and indicate the mapping. This means that we often blur the distinction between the view of $\mathrm{dom}(\pi) \subseteq A$ and $\mathrm{image}(\pi) \subseteq B$ as subsets, or as tuples enumerating these subsets, or as hyperedges of hypergraphs over domains A and B; notation like $\rho\colon \mathbf{a} \mapsto \mathbf{b}$ and $\rho\colon s \mapsto t$ may indicate those shifts in perspective.

Definition 2.8. A *hypergraph bisimulation* between hypergraphs $\mathrm{H} = (A, S^{\mathrm{H}})$ and $\mathrm{K} = (B, S^{\mathrm{K}})$ is a non-empty back-and-forth system Z of local bijections between A and B with the following properties

(a) every $\rho \in Z$ is a bijection between some $s = \mathrm{dom}(\rho) \in S^{\mathrm{H}}$ and some $t = \mathrm{image}(\rho) \in S^{\mathrm{K}}$.

(b) back-and-forth conditions:
(*back*) for all $(\rho\colon s \mapsto t) \in Z$ and for every $t' \in S^{\mathrm{K}}$ there is some $(\rho'\colon s' \mapsto t') \in Z$ such that $\rho'^{-1}\restriction(t \cap t') = \rho^{-1}\restriction(t \cap t')$;
(*forth*) for all $(\rho\colon s \mapsto t) \in Z$ and for every $s' \in S^{\mathrm{H}}$ there is some $(\rho'\colon s' \mapsto t') \in Z$ such that $\rho'\restriction(s \cap s') = \rho\restriction(s \cap s')$.

We write $\mathrm{H}, s \sim \mathrm{K}, t$ and say that H, s and K, t are bisimilar if there is a hypergraph bisimulation Z between H and K with some $(\rho\colon s \mapsto t) \in Z$.

It is important to note that the back-and-forth conditions do not even indirectly require any bijectivity properties for combinations of the local bijections ρ. For instance, with $\rho\colon s \mapsto t$ and $\rho'\colon s' \mapsto t'$ as in the *forth* property, it is admissible that $t' = t$ even though $s' \neq s$.

The following definition of a *hypergraph cover* captures the special case in which a hypergraph bisimulation is induced by a homomorphism from the first to the second structure. A (hypergraph) homomorphism from $\mathrm{H} = (A, S^{\mathrm{H}})$ to $\mathrm{K} = (B, S^{\mathrm{K}})$ is a global map $\rho\colon A \to B$ whose restriction to every hyperedge of H is a bijection onto some hyperedge of K; the collection of its restrictions to the hyperedges of H thus automatically satisfy condition (a) and the *forth* condition in (b). What may be missing is the *back* condition. In the usual geometric imagery of covers, this *back* condition is responsible for the availability of lifts of paths of overlapping hyperedges to the covering hypergraph H.

Definition 2.9. A homomorphism $\pi\colon \hat{\mathrm{H}} \to \mathrm{H}$ between hypergraphs $\hat{\mathrm{H}} = (\hat{A}, \hat{S})$ and $\mathrm{H} = (A, S)$ is a *hypergraph cover* if $Z = \{\pi \restriction \hat{s} \colon \hat{s} \in \hat{S}\}$ is a hypergraph bisimulation. Shorthand notation: $\pi\colon \hat{\mathrm{H}} \xrightarrow{\sim} \mathrm{H}$. We also say that $\hat{\mathrm{H}}$ *covers* H.

An alternative view of both hypergraph bisimulation and covers can be based on corresponding notions of bisimulations between transition systems (based on suitable decorations of the intersection graphs). This has been expounded elsewhere, for instance in [12].

2.4 Examples

Tree unfoldings. Just as for transition systems, there is a natural construction of tree unfoldings for hypergraphs that yields covers by acyclic hypergraphs. As for graphs and transition systems, these covers are typically infinite even if we start from a finite hypergraph. We here discuss ω-tree-unfoldings, which simultaneously boost all multiplicities, because they will be useful later; to achieve just acyclicity in a cover, a plain tree unfolding would suffice. Let $\mathrm{H} = (A, S)$ be a hypergraph with intersection graph $\mathfrak{I} = (S, \Delta)$. Considered as a transition system, $\mathfrak{I}$ admits an ω-tree-unfolding, from any choice of $s \in S$ as a root, into a tree structure $\mathfrak{I}_s^{\omega *}$, which is based on the domain of all $\mathbb{N}$-edge-labelled paths from s in $\mathfrak{I}$ (cf. Section 1.2). Let $\mathfrak{T}$ be the disjoint union of countably many isomorphic copies of every such $\mathfrak{I}_s^{\omega *}$ for every $s \in S$, all joined by edges to a single new root vertex, which we symbolically label $\emptyset$.

We look at the natural projection $\pi\colon \mathfrak{T} \setminus \{\emptyset\} \to \mathfrak{I}$ as a cover $\pi\colon \mathfrak{T} \setminus \{\emptyset\} \xrightarrow{\sim} \mathfrak{I}$ in the natural manner. A hypergraph cover $\pi\colon \mathrm{H}^{\omega *} \xrightarrow{\sim} \mathrm{H}$ is obtained as follows.

The idea is to insert actual hyperedges for the elements of $\mathfrak{T}$, with appropriate overlaps according to Δ. The domain $A^{\omega *}$ of $\mathrm{H}^{\omega *}$ will be a suitable quotient of the disjoint union $\bigcup_{\rho \in \mathfrak{T} \setminus \{\emptyset\}} (\pi(\rho) \times \{\rho\})$, where the vertices ρ of $\mathfrak{T} \setminus \{\emptyset\}$ are labelled paths in $\mathfrak{I}$ from some s_0 to some $s = \pi(\rho) \subseteq A$ and ρ is used as a tag to make the union disjoint. For next neighbours $\rho = s_0 \ldots s$ and $\rho' = s_0 \ldots s s'$ in some copy of $\mathfrak{I}_{s_0}^{\omega *}$ in $\mathfrak{T}$ we need to identify elements of $\pi(\rho) \times \{\rho\} = s \times \{\rho\}$ and $\pi(\rho') \times \{\rho'\} = s' \times \{\rho'\}$ that come from the intersection $s \cap s'$. This identification needs to be extended by transitivity along paths in $\mathfrak{T}$. For $(a_1, \rho_1) \in \pi(\rho_1) \times \{\rho_1\}$ and $(a_2, \rho_2) \in \pi(\rho_2) \times \{\rho_2\}$ put:

$$(a_1, \rho_1) \approx (a_2, \rho_2) \quad \text{if} \quad \begin{array}{l} a_1 = a_2 \quad \text{and } \rho_1 \text{ and } \rho_2 \\ \text{are connected in } \mathfrak{T} \restriction \{\rho \colon a_1 \in \pi(\rho)\}. \end{array}$$

Let

$$A^{\omega *} := \Bigl(\bigcup_{\rho \in \mathfrak{T}} \pi(\rho) \times \{\rho\} \Bigr) \,/ \approx$$

and declare just the subsets represented by sets $\pi(\rho) \times \{\rho\}$ to be the hyperedges of the desired $\mathrm{H}^{\omega *} = (A^{\omega *}, S^{\omega *})$. The projection π naturally induces a projection π that sends the equivalence class of some $(a, \rho) \in \pi(\rho) \times \{\rho\}$ to a.

Then π induces a hypergraph bisimulation and $\pi\colon \mathrm{H}^{\omega *} \xrightarrow{\sim} \mathrm{H}$ is a hypergraph cover, which we refer to as the *ω-tree-unfolding of* H.

The underlying tree $\mathfrak{T}$ naturally yields a hypergraph tree decomposition according to Definition 2.5 whence $\mathrm{H}^{\omega *}$ is seen to be acyclic.

Definition 2.10. The *ω-tree-unfolding* of the hypergraph H is the bisimilar cover $\pi \colon \mathrm{H}^{\omega *} \xrightarrow{\sim} \mathrm{H}$ by the acyclic hypergraph $\mathrm{H}^{\omega *}$ as constructed above.

So every hypergraph admits a cover by an acyclic hypergraph. We shall return to this simple infinitary cover construction in comparison with finite approximations to acyclicity *and* unbounded multiplicities in Section 3; for the choice of ω-tree-unfoldings rather than plain tree unfoldings we point to Observation 3.2 in particular.

From Observation 1.5 we find that also here bisimilarity between countable base hypergraphs implies isomorphy of their ω-tree-unfoldings.

Finite conformal covers. While we shall see below that finite hypergraphs need not admit any cover by a finite acyclic hypergraph, covers by finite conformal hypergraphs can always be obtained. We meanwhile know two different constructions, a combinatorially very simple one from [13] and a more sophisticated one, with exponentially better size bounds, from [5]. The latter one is really a corollary to the construction of weakly N-acyclic finite covers which is at the centre of [5] (see Section 2.5 below).

Limits for chordality in finite covers. Maybe the simplest hypergraph that is not a graph and fails to be acyclic is formed by the facets of the 3-simplex (the faces of the tetrahedron), i.e., the full width 3 hypergraph on 4 vertices:

$$\mathrm{H}_4^3 := \big(\{1,2,3,4\}, \{\{1,2,3,4\} \setminus \{i\} \colon i = 1,2,3,4\}\big).$$

Clearly H_4^3 is chordal, but fails to be conformal because its Gaifman graph is the 4-clique and the whole domain is not a hyperedge. It is also not hard to see that this hypergraph admits *no* cover by an acyclic hypergraph that would be even just locally finite. In fact, one checks that the 1-neighbourhood of every vertex $\hat{v}$ in any acyclic cover $\pi \colon \hat{\mathrm{H}} \xrightarrow{\sim} \mathrm{H}_4^3$ must be infinite: it must cover (acyclically, in the graph sense) the cycle formed by the three edges not incident on $v = \pi(\hat{v})$ in $\mathfrak{G}(\mathrm{H}_4^3) \simeq \mathfrak{K}_4$, which are the far ends of the 3 hyperedges incident on v in H_4^3.

This is in marked contrast to the situation for finite graphs where a simple product with a finite Cayley graph of large girth always produces covers by finite N-locally acyclic graphs, see e.g. [16] and Example 2.17 below.

On the other hand, a simple hexagonal grid pattern as in Figure 1, which doubly unfolds each of these 3-cycles, yields a natural cover by an infinite conformal hypergraph whose shortest chordless cycles have length 6. A geometric identification of opposite borders in any hexagon of even radius $r = 2m$ in this infinite grid remains compatible with the natural projection and the *back* requirements, hence yields a cover of H_4^3 by a finite hypergraph $\hat{\mathrm{H}}_{2m}$, which is still conformal and, for $m \geqslant 2$ still has no chordless cycles of length less than 6. It is interesting to note that there are two kinds of chordless cycles in these

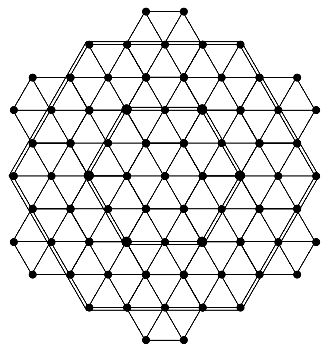


Figure 1: hexagonal grid pattern for covers of H_4^3

covers: 'local cycles' that arise as local covers of the 3-cycles in the neighbourhood of any vertex $v \in H_4^3$, and those that arise through identifications along the boundary of the fundamental domain. (For $m = 1$, $\hat{H}_2$ has chordless cycles of the latter kind of length 4, since this is the diameter of its hexagonal base.)

2.5 Degrees of acyclicity in finite covers

As the example of H_4^3 suggests, finite cover constructions can at best aim for degrees of acyclicity that render small sub-configurations acyclic.

Definition 2.11. A hypergraph H is called *N-acyclic* for some $N \in \mathbb{N}$ if every induced sub-hypergraph on up to N vertices of H is acyclic. *N-chordality* and *N-conformality* are similarly defined.

The following theorem from [18] is based on an inductive construction, by induction on the width of H. It proceeds through

(a) the construction of locally finite conformal and N-chordal covers of finite conformal hypergraphs of width w using the existence of finite conformal and N-chordal covers for hypergraph of width $w - 1$.

(b) the construction of finite conformal and N-chordal covers from bounded pieces of locally finite conformal and N-chordal covers of sufficiently large diameter; this involves identifications in a non-trivial glueing process and mending of defects w.r.t. *back* requirements where these are violated along the boundary.

Width $w-1$ hypergraphs arise in (a) as localisations in the 1-neighbourhoods of vertices of H after removal of the central vertex v, similar to the triangle graph resulting from the deletion of any one vertex from H_4^3; in the general situation, conformality ensures that a cover of this punctured 1-neighbourhood can be completed in a canonical manner by re-insertion of a cover $\hat{v}$ above the central vertex v. Step (b) uses a combinatorial group construction which we shall present in another useful role in Section 2.6 below.

Theorem 2.12. *Every finite hypergraph* H *admits, for every* $N \in \mathbb{N}$*, a cover* $\pi\colon \hat{\mathrm{H}} \xrightarrow{\sim} \mathrm{H}$ *by some finite, conformal and* N*-chordal hypergraph* $\hat{\mathrm{H}}$.

A further relaxation of acyclicity requirements in finite covers $\pi\colon \hat{\mathrm{H}} \xrightarrow{\sim} \mathrm{H}$ is motivated by the desire to avoid not necessarily all small cyclic configurations in the covering hypergraph, but to achieve acyclicity of small configurations at least 'projectively', under π. In other words, we are content with a state of affairs in which small cyclic homomorphic images in the covering hypergraph $\hat{\mathrm{H}}$ can be decomposed in projection to the base hypergraph H. This condition does not rule out, for instance, a chordless cycle of length 4 in $\mathfrak{G}(\hat{\mathrm{H}})$ provided it projects onto a chordal cycle in $\mathfrak{G}(\mathrm{H})$, e.g., onto a path of length 2 through identification of two opposite vertices along the cycle. Technically we go via tree decompositions.

In light of Definition 2.5, N-acyclicity of the covering hypergraph in $\pi\colon \hat{\mathrm{H}} \xrightarrow{\sim} \mathrm{H}$ requires every subset $B \subseteq \hat{A}$ of up to N vertices in $\hat{\mathrm{H}}$ to admit a tree decomposition $\lambda\colon T \to \hat{S}$ of the following kind.

(i) every induced hyperedge $\hat{s} \cap B$ is contained in some bag $\lambda(v)$: $\hat{s} \cap B \subseteq \lambda(v)$ for some v, and

(ii) for every $b \in B$, the vertex set $\{v \in V : b \in \lambda(v)\}$ is connected in $\mathfrak{T}$.

In other words, the bags $\lambda(v)$ of this tree decomposition must be hyperedges of $\hat{\mathrm{H}}$. For a weakly N-acyclic cover we relax this and only require the projections of the bags to be (subsets of) hyperedges in the base hypergraph H.

Definition 2.13. A hypergraph cover $\pi\colon \hat{\mathrm{H}} \xrightarrow{\sim} \mathrm{H}$ is called *weakly* N*-acyclic* for some $N \in \mathbb{N}$ if every subset $B \subseteq \hat{A}$ of up to N vertices in $\hat{\mathrm{H}}$ admits a tree decomposition $\lambda\colon T \to \mathcal{P}(B)$ with bags $\lambda(v)$ such that $\pi(\lambda(v)) \in S$.

The following theorem is a core result from [5]. The highly uniform and regular construction there involves a suitable quotient of a term-based hypergraph, which is organised so as to provide the witnesses of all *back*-requirements. Due to the uniformity of the construction, these covers preserve all symmetries of the base hypergraph; they are also of feasible size and allow meaningful and important applications of a more algorithmic nature – but they fail to come close to full N-acyclicity in the sense of Definition 2.11.

Theorem 2.14. *Every finite hypergraph* H *admits, for every* $N \in \mathbb{N}$*, a weakly* N*-acyclic cover* $\pi\colon \hat{\mathrm{H}} \xrightarrow{\sim} \mathrm{H}$ *by some finite, conformal hypergraph* $\hat{\mathrm{H}}$.

For diverse applications of these weakly N-acyclic covers we refer to [5]. In the context of this paper we encounter them in connection with Ptime canonisation and an abstract capturing result in the sense of descriptive complexity in Section 5. An application from [6] towards the finite model property of the guarded negation fragment GNF is indicated in Section 4.4, where we use it to establish a new expressive completeness result for that logic.

2.6 Highly acyclic Cayley groups

The following construction of Cayley graphs and groups from [18] owes its basic idea to an elegant and simple idea for the construction of Cayley groups of large girth, as presented in [1]. We review the basic details, and sketch the extension of this basic idea from [18]. This construction uses amalgamated chains of smaller Cayley graphs to form the nuclei of larger ones – similar to the manner in which uniform trees form the nuclei of Cayley groups of large girth in the standard construction, which we review first.

Cayley groups and graphs. In the following, a *Cayley group* and its *Cayley graph* consist of an abstract group G and a specified set E of non-trivial involutive group elements that generate G. I.e., $e \neq 1$ and $e^2 = 1$ for every $e \in E$, and every group element $g \in G$ is the product of some reduced sequence of generators: $g = \prod_{i \leqslant k} e_i$ with $k \in \mathbb{N}$, $e_i \in E$ for $i \leqslant k$ and $e_i \neq e_{i+1}$ (reduced) for $i < k$.[6] Every generator $e \in E$ induces an edge relation on G consisting of all the pairs of the form (g, ge). This edge relation is symmetric, since $(ge, g) = (ge, (ge)e)$. We refer to both, G as a group with generators $e \in E$ and G as an E-edge-coloured graph, just as G. Each generator $e \in G$ also acts as a permutation π_e on the E-coloured graph G, where $\pi_e \colon G \to G$ maps every g to ge. This map is a permutation, in fact an involution, since e is involutive; in terms of the E-coloured graph G, it swaps every pair of e-related vertices (g, ge). One checks that the abstract group G (with generators $e \in E$) is isomorphic to the subgroup of the permutation group of the vertex set G generated by these permutations (with generators π_e for $e \in E$).

In this sense, G as a group generated by E and G as an E-coloured graph contain exactly the same structural information up to isomorphism.[7]

In an arbitrary E-coloured graph $(V, (R_e)_{e \in E})$, each $e \in E$ similarly induces an involutive permutation of the vertex set V provided that every vertex $v \in V$ is incident with *at most one e-edge*. In the following we want to use the term *E-coloured graph* in this restricted meaning.

Definition 2.15. An *E-coloured graph* is a structure $(V, (R_e)_{e \in E})$ with disjoint, non-empty, irreflexive and symmetric edge relations R_e for $e \in E$ such that, for each $e \in E$, every vertex is incident with at most one edge in R_e.

Every E-coloured graph $\mathfrak{A} = (V, (R_e)_{e \in E})$ in the sense of this definition gives rise to a Cayley group G, with non-trivial involutive generators $e \in E$. This Cayley group is obtained as a subgroup of the permutation group $\mathrm{Sym}(V)$ according to the above recipe. We denote this Cayley group and its Cayley graph as $\mathrm{sym}(\mathfrak{A})$:

$$\mathrm{sym}(\mathfrak{A}) := \langle \pi_e \colon e \in E \rangle^{\mathrm{Sym}(V)}.$$

Finite trees and large girth. The girth of a graph is the minimal length of a non-trivial cycle (and undefined or infinite if the graph is acyclic). The *girth of a Cayley group* is the girth of the Cayley graph; equivalently, it is the minimal

[6] We use multiplicative notation and write $1 \in G$ for the neutral element; precedence in group products is to the left, i.e., $gh\ell = (gh)\ell$.

[7] Just up to isomorphism because the passage from the graph to the group involves an arbitrary choice of a root vertex to become the unit element of the group.

positive length of a reduced generator sequence $(e_i)_{i\leqslant k}$ such that $\prod_{i\leqslant k} e_i = 1$. Finite Cayley graphs clearly must have finite girth. Since Cayley graphs are such generic and highly symmetric graph objects, it is interesting to see which degree of acyclicity can be achieved. As discussed in [1], the following is sub-optimal in terms of the relationship between degree (number of generators), girth (length of shortest cycles) and size that can be realised in Cayley graphs; but it stands out as a particularly elegant and transparent construction.

Observation 2.16. *For the regularly E-coloured tree of depth n, $\mathcal{T}_E^n$, whose vertices are the reduced sequences $\sigma = (e_i)_{i\leqslant k}$ of generators $e_i \in E$ of lengths $k \leqslant n$, with e-edges between σ and σe, the Cayley group* $\mathrm{sym}(\mathcal{T}_E^n)$ *has girth greater than n (in fact, greater than $4n+1$).*

In its weak form the observation is immediate if we consider just the effect that a reduced sequence of permutations π_e has on the root vertex of $\mathcal{T}_E^n$: each new permutation π_e moves this vertex one step further away from the root, increasing its depth in the tree by 1, up until it may reach depth n (if the sequence has length n). The stronger claim can similarly be checked by focusing on a clever choice of a leaf node instead.

As one example for the usefulness of such groups consider the following construction of N-locally acyclic bisimilar covers for hypergraphs of width 2. This simple construction can be used as the basis for several expressive completeness proofs for modal logics, as discussed in [17, 16] and in Section 1.2 above, including substantial further variations on Theorem 1.7 e.g. in [10].

Example 2.17. Let $\mathrm{H} = (A, E)$ be a finite undirected graph considered as a hypergraph of width 2, and G a Cayley group with generator set E. Then the following natural product of H and G, with the natural projection, provides a cover:

$$\hat{H} = (A \times G, \hat{E}) \quad \text{where} \quad \hat{E} = \big\{\{(a,g),(a',ge)\} \colon e = \{a,a'\} \in E\big\}.$$

If the girth of G is greater than $(2\ell+1)$, then $\hat{H}$ has no non-trivial cycles of lengths up to $2\ell+1$, which means that this cover is acyclic in the ℓ-neighbourhood of any one of its vertices.

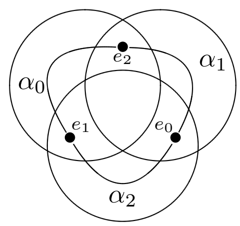
More than large girth. In a Cayley group G with generator set E, consider subgroups $G[\alpha]$ generated by subsets $\alpha \subseteq E$ and their cosets $gG[\alpha] \subseteq G$. In the Cayley graph, the coset $gG[\alpha]$ is the connected component of g in the reduct that just retains the edge relations R_e for $e \in \alpha$. The lattice of subsets $\alpha \subseteq E$ induces a lattice of equivalence relations $\sim_\alpha$ on the graph G, where $g \sim_\alpha h$ if $g^{-1}h \in G[\alpha]$. The finest of these is equality in the form of $\sim_\emptyset$, and the coarsest is the full binary relation on G in the form of $\sim_E$. For the Cayley groups to be considered here, we will always have compatibility with intersections in the sense that $G[\alpha_1 \cap \alpha_2] = G[\alpha_1] \cap G[\alpha_2]$. As shown in [18], this property is a consequence of a more fundamental compatibility condition, viz. of *compatibility* *of the subgraphs* $G[\alpha]$ with G, which requires that cycles in G induce cycles in the Cayley graphs of all its subgroups $G[\alpha]$:

$$(\dagger) \qquad \prod e_i = 1 \text{ in } G \quad \Rightarrow \quad \prod \pi_{e_i} = \mathrm{id} \text{ in } \mathrm{sym}(G[\alpha])$$

for every sequence of generators (e_i) and every $\alpha \subseteq E$. This condition can be guaranteed in the group constructions presented in [18].

We now focus on short cycles in G w.r.t. to the induced equivalence relations $\sim_\alpha$, rather than w.r.t. the basic edge relations R_e. A $\sim_\alpha$-edge is witnessed by a path w.r.t. edges in $\bigcup_{e\in\alpha} R_e$, but the length of such a path is only bounded by the size of $G[\alpha]$. Large girth of G does not rule out short cycles w.r.t. $\sim_\alpha$-edges for a succession of distinct $\alpha \subseteq E$ in this far more general sense.

In a sense G will always have many short cycles, as an almost trivial example may illustrate.



Let $\alpha_i \subseteq E$ and $e_i \in E$ for $i = 0, 1, 2$ be such that $e_i \in (\alpha_{i+1} \cap \alpha_{i+2}) \setminus \alpha_i$ (cyclic indexing in $\mathbb{Z}_3$). Then e_{i+1} and e_{i+2} are linked by an $\sim_{\alpha_i}$-edge, but not by an edge of the other two equivalence relations $\sim_\alpha$ for $\alpha = \alpha_{i+1}, \alpha_{i+2}$.

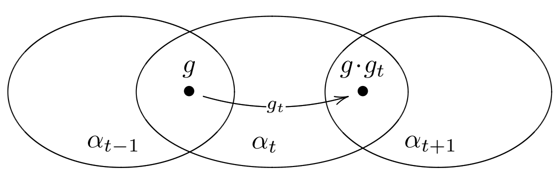
The following condition of just *local* non-degeneracy avoids this trivialisation of the matter.

Definition 2.18. Let G be a Cayley group with generators E as above, whose action on the Cayley graphs of its subgroups $G[\alpha]$ for $\alpha \subseteq E$ is compatible with the group structure of G according to $(\dagger)$.

(a) A cyclically indexed n-tuple $(g_t)_{t\in\mathbb{Z}_n}$ in G forms a *non-trivial cycle* of length n w.r.t. the *colouring* $\alpha\colon t \mapsto \alpha_t \subseteq E$ if
   (i) $\prod_{t\in\mathbb{Z}_n} g_t = 1$,
   (ii) $g_t \in G[\alpha_t]$, but $g \cdot G[\alpha_{t-1} \cap \alpha_t] \cap g \cdot g_t \cdot G[\alpha_t \cap \alpha_{t+1}] = \emptyset$.

(b) G is *N-acyclic* if it has no non-trivial coloured cycles of lengths $n \leqslant N$.

Condition (a) (ii) may be seen a stipulation for non-trivial coset cycles, ruling out local shortcuts that would avoid the segment $g_t \in G[\alpha_t]$ altogether.



Observation 2.19. *The intersection condition on subgroups,* $G[\alpha] \cap G[\beta] = G[\alpha \cap \beta]$ *(itself a consequence of condition (†)) says that* G *does not have non-trivially coloured* 2*-cycles in the sense of (a).*

Cayley groups of this kind are obtained in [18] in an inductive construction that eliminates short non-trivial coloured cycles within subgroups $G[\alpha]$ for subsets $\alpha \subseteq E$ of increasing size. To eliminate undesirable cycles in $G[\alpha]$, the inductive step uses amalgamated chains of Cayley graphs $G[\alpha']$ for $\alpha' \subsetneq \alpha$, as would occur as a pattern of overlapping cosets along a non-trivially coloured cycle in $G[\alpha]$. These chains are used as components of E-coloured graphs $\mathcal{K}$ from which G is obtained as $G = \mathrm{sym}(\mathcal{K})$. Some care needs to be taken to maintain compatibility of the action of G with the Cayley graphs of the already settled smaller $G[\alpha']$.

Proposition 2.20. *For every finite set* E *and every* $N \in \mathbb{N}$ *there is an* N*-acyclic Cayley group with generator set* E*.*

Interestingly, N-acyclicity of Cayley groups as defined above is directly related to the N-acyclicity of an associated hypergraph of cosets. Let $\mathrm{H}(G)$ be the hypergraph whose vertex set consists of all cosets $gG[\alpha]$ for $g \in G$ and $\alpha \subseteq E$, with hyperedges induced by group elements g according to

$$[g] := \{gG[\alpha] \colon \alpha \subseteq E\}.$$

Observation 2.21. *For any Cayley group* G *satisfying the compatibility condition (†), the coset hypergraph* $\mathrm{H}(G)$ *is* N*-acyclic if, and only if,* G *is* N*-acyclic.*

Proof. See Observation 3.2 in [18] for N-acyclicity of $\mathrm{H}(G)$, given N-acyclicity of G. The converse implication is not discussed in [18], and may serve here to illustrate the point of condition (a),(ii) of Definition 2.18.

In preparation consider $g_i \in G[\alpha_i]$ for $i = t-1, t, t+1$. Then, for any $h \in G$, $hG[\alpha_i] = hg_iG[\alpha_i]$, and therefore in $\mathrm{H}(G)$:

– $g_{t-1}G[\alpha_{t-1}]$ and $g_{t-1}G[\alpha_t] = g_{t-1}g_tG[\alpha_t]$ are linked by hyperedge $[g_{t-1}]$,

– $g_{t-1}G[\alpha_t] = g_{t-1}g_tG[\alpha_t]$ and $g_{t-1}g_tG[\alpha_{t+1}]$ are linked by hyperedge $[g_{t-1}g_t]$.

Then the condition $g_t \notin G[\alpha_{t-1} \cap \alpha_t] \cdot G[\alpha_t \cap \alpha_{t+1}]$ precisely corresponds to the condition that no hyperedge of $\mathrm{H}(G)$ directly links $g_{t-1}G[\alpha_{t-1}]$ to $g_{t-1}g_tG[\alpha_{t+1}]$.

Now a cycle $(g_t)_{t \in \mathbb{Z}_n}$ in G that is non-trivially coloured by $\alpha \colon t \mapsto \alpha_t$ would give rise in $\mathrm{H}(G)$ to cyclic tuples $(h_tG[\alpha_t])_{t \in \mathbb{Z}_n}$ with $h_{t+1} = h_tg_{t+1}$. In these cyclic tuples, next neighbours are linked by hyperedges, but none of the next neighbour triples $h_{t-1}G[\alpha_{t-1}], h_tG[\alpha_t], h_{t+1}G[\alpha_{t+1}]$ is covered by a common hyperedge. For $n = 3$ this violates 3-conformality; for $3 < n \leqslant N$ it either violates N-chordality or, again, 3-conformality. □

3 Finite saturation w.r.t. multiplicities

Apart from short cycles, distinctions between small multiplicities are an obvious obstacle to controlling finite levels $\equiv^q$ of elementary equivalence by finite levels of bisimulation equivalence $\sim^{f(q)}$. This is true in the hypergraph setting (of relational structures of arbitrary width) as in the graph setting. In the graph setting, multiplicities can just be boosted to infinity in ω-tree-unfoldings or, in a finitary setting, multiplied by a factor of q in a simple product with the q-clique $\mathcal{K}_q$. As with acyclicity, the corresponding task is more challenging for hypergraphs. It turns out to be convenient to replace the concept of high branching degrees and multiplicities by a notion of *free* branching between hyperedges, because hyperedges can overlap in more than just a single vertex. We review the notion of *freeness* from [18].

Definition 3.1. (a) Two subsets of vertices $B_1, B_2 \subseteq A$ are called *n-free* in the hypergraph $\mathrm{H} = (A, S)$, if the graph distance between the two vertex sets $B_i \setminus (B_1 \cap B_2)$ in the Gaifman graph of $\mathrm{H}\restriction(A \setminus (B_1 \cap B_2))$ is greater than n.

(b) $\mathrm{H} = (A, S)$ is *(n, K)-free* if, for any subset $B \subseteq A$ of up to K vertices, every $s \in S$ has some companion $s' \in S$ for which
  - $\mathrm{H}, s' \sim \mathrm{H}, s$ and $B \cap s' = B \cap s$;
  - s' and B are n-free.

(c) A cover $\pi\colon \hat{\mathrm{H}} = (\hat{A}, \hat{S}) \xrightarrow{\sim} \mathrm{H}$ is *(n, K)-free* if, for any subset $\hat{B} \subseteq \hat{A}$ of up to K vertices, every $\hat{s} \in \hat{S}$ has some companion $\hat{s}' \in \hat{S}$ for which
  - $\pi(\hat{s}') = \pi(\hat{s})$ and $\hat{B} \cap \hat{s}' = \hat{B} \cap \hat{s}$;
  - $\hat{s}'$ and $\hat{B}$ are n-free.

Note that freeness of a cover according to (c) is a stronger notion than freeness of the covering hypergraph, since the condition $\pi(\hat{s}') = \pi(\hat{s})$ implies $\hat{\mathrm{H}}, \hat{s}' \sim \hat{\mathrm{H}}, \hat{s}$, but not vice versa.

For the following compare the construction of ω-tree-unfoldings of hypergraphs encountered in Section 2.4, based on the ω-tree-unfolding of the underlying intersection graph $\mathfrak{I}(\mathrm{H})$, and especially Definition 2.10.

Observation 3.2. *The ω-tree-unfolding $\mathrm{H}^{\omega*}$ of H is not only acyclic but also (n, K)-free as a cover $\pi\colon \mathrm{H}^{\omega*} \xrightarrow{\sim} \mathrm{H}$ w.r.t. to the natural projection, for all $n, K \in \mathbb{N}$.*

Proof. Any guarded subset s of $\mathrm{H}^{\omega*}$ is represented at some node v of the underlying tree $\mathcal{T}$, which gives rise to the hypergraph tree decomposition of $\mathrm{H}^{\omega*}$. Due to the connectivity condition for tree decompositions (cf. Definition 2.5), hyperedges that are represented in siblings subtrees rooted at distinct successors of v are in distinct connected components of $\mathrm{H}^{\omega*} \setminus s$. As every type of sibling subtree arises with infinite multiplicity in $\mathcal{T}$, freeness follows. NB: we assume that the set of hyperedges is closed under passage to subsets, according to Proviso 2.2; this is essential for the direct manner in which we achieve freeness here. ☐

N-acyclic Cayley groups can be used to achieve any desired finite degree of freeness in covers obtained as reduced products of the given hypergraph with the group.

3.1 Reduced products with groups

As a simple example, consider a hypergraph $\mathrm{H} = (A, S)$ and a Cayley group with generator set S. A natural reduced product of H and G can be obtained as a quotient of the plain product of the hypergraph H with the set of group elements G. We think of the plain product as a G-indexed stack of isomorphic copies of H. We now identify subsets corresponding to the same $s \subseteq S$ in different layers g and g' whenever g and g' are related by an s-edge in G (the edges in the Cayley graph of G are labelled by hyperedges of H, which are the generators in this case).

A slight generalisation of this construction turns out to be useful. We admit an arbitrary set E of generators for G, and associate the generators $e \in E$ with subsets $\rho(e) \subseteq A$ of the domain of H in which corresponding layers are to be identified.

Definition 3.3. For a Cayley group G generated by E, a hypergraph $\mathrm{H} = (A, S)$, and a map $\rho \colon E \to \mathcal{P}(A)$, let $\mathrm{H} \times_\rho G$ be the hypergraph with domain $\hat{A} := (A \times G)/{\approx}$ where

$$(a, g) \approx (a, g') \quad \Leftrightarrow \quad g^{-1} \circ g' \in G[E_a] \ \text{ for } \ E_a = \{e \in E \colon a \in \rho(e)\}$$

and hyperedges $[s, g] := \{[a, g] \colon a \in s, g \in G\}$, where we write $[a, g]$ for the $\approx$-class of $(a, g) \in A \times G$.

By these definitions, $[a, g] \in [s, h]$ iff $a \in s$ and $g^{-1} \circ h \in G[E_a]$. The natural projection $\pi \colon [a, g] \mapsto a$ turns this construction into a cover $\pi \colon \mathrm{H} \times_\rho G \xrightarrow{\sim} \mathrm{H}$.

It is shown in [18] that reduced products with Cayley groups preserve N-acyclicity of hypergraphs, provided

- ρ maps generators to (subsets of)[8] hyperedges of H, and
- G is N-acyclic.

To enrich a hypergraph in the sense of freeness, we may pick an N-acyclic Cayley group with generator set $S \times \{1, \dots, m\}$, and the natural association of generator (s, i) to the hyperedge s. Then the subgroups $G[E_a]$ that identify layers in the fibre above a are generated by $E_a = \{(s, i) \colon a \in s, 1 \leqslant i \leqslant m\}$, and every hyperedge $[s, g]$ in the reduced product is part of at least every one of the m many layers $\mathrm{H} \times \{g(s, i)\}$ for $1 \leqslant i \leqslant m$. As is also shown in [18], this simple reduced product construction, with an N-acyclic Cayley group generated by $S \times \{1, \dots, m\}$, provides a cover that is (n, K)-free provided m and N are large enough in relation to n and K, while preserving N-acyclicity. This gives the following strengthening of Theorem 2.12.

Corollary 3.4. *Every finite hypergraph* H *admits, for all* $n, K, N \in \mathbb{N}$, *covers* $\pi \colon \hat{\mathrm{H}} \xrightarrow{\sim} \mathrm{H}$ *by finite, conformal,* (n, K)*-free and* N*-chordal hypergraphs* $\hat{\mathrm{H}}$.

[8] Recall that we here assume here that S is closed under subsets.

4 Tractability of rich N-acyclic finite covers

We return to the tractability idea, and to the intention to control the FO_q-type in suitable relational structures $\mathfrak{A}$ by (essentially) the level $f(q)$ bisimulation type in the associated hypergraph of guarded subsets $\mathrm{H}(\mathfrak{A})$ (plus the basic local relational information corresponding to quantifier-free types). The role of modal logic is here played by the guarded fragment of first-order logic, $\mathrm{GF} \subseteq \mathrm{FO}$, which has been put to great uses in algorithmic model theory since its introduction by Andréka, van Benthem and Németi in [2]. The associated back-and-forth equivalence of *guarded bisimulation* may be looked at as both,

(a) the Ehrenfeucht–Fraïssé notion capturing the restricted pattern of quantification in GF, which is always relativised to guarded tuples;

(b) the strengthening of hypergraph bisimulation between the hypergraphs of guarded sets by the requirement that the local relational information (quantifier-free type) is respected.

4.1 Guarded bisimulation and GF

Guarded bisimulation. We fix a finite relational vocabulary τ as before and first discuss guarded bisimulation equivalence under the preferred angle of (b) above.

Definition 4.1. A *guarded bisimulation* between τ-structures $\mathfrak{A}$ and $\mathfrak{B}$ is a non-empty back-and-forth system Z that

(i) is a hypergraph bisimulation between the associated hypergraphs of guarded subsets $\mathrm{H}(\mathfrak{A})$ and $\mathrm{H}(\mathfrak{B})$, and

(ii) consists of local isomorphisms ρ between the substructures $\mathfrak{A}{\restriction}s$ and $\mathfrak{B}{\restriction}t$ for the guarded subsets $s = \mathrm{dom}(\rho) \subseteq A$ and $t = \mathrm{image}(\rho) \subseteq B$.

$\mathfrak{A}, \mathbf{a}$ and $\mathfrak{B}, \mathbf{b}$ are guarded bisimilar, written $\mathfrak{A}, \mathbf{a} \sim_{\mathrm{g}} \mathfrak{B}, \mathbf{b}$, if $\rho \colon \mathbf{a} \mapsto \mathbf{b}$ is a local isomorphism and there is a guarded bisimulation Z such that all restrictions of ρ to guarded subsets (in $\mathfrak{A}$ or $\mathfrak{B}$) are in Z.

The idea for the corresponding *guarded bisimulation game* is a combination of the game protocol for the hypergraph bisimulation game with the prescription that the second player loses as soon as the current position does not correspond to an isomorphism between the induced substructures. A guarded bisimulation then witnesses the existence of a winning strategy for the second player in the infinite game.

The guarded fragment. The guarded fragment of first-order logic restricts first-order quantifications to guarded tuples through explicit relativisation by relational atoms. We call a tuple $\mathbf{a}$ *strictly guarded* in the τ-structure $\mathfrak{A}$ if the set of its components $[\mathbf{a}]$ is a singleton set or agrees with the set of components $[\mathbf{a}']$ of some tuple $\mathbf{a}' \in R^{\mathfrak{A}}$ in one of the relations of $\mathfrak{A}$. In fact, GF restricts quantification to strictly guarded tuples.

More specifically, GF[τ] admits just *guarded quantification* of the form

$$\exists \mathbf{y}(\alpha(\mathbf{x}) \wedge \varphi(\mathbf{x})) \quad \text{and} \quad \forall \mathbf{y}(\alpha(\mathbf{x}) \rightarrow \varphi(\mathbf{x}))$$

where $\alpha(\mathbf{x})$ is an atomic τ-formula (a relational atom or an equality) in which every free variable of φ must occur, and $\mathbf{y} \subseteq \mathbf{x}$ is a tuple of variables among those in $\mathbf{x}$ (the free variables of φ being among those in $\mathbf{x}$ by our notational conventions). A GF[τ]-formula is called *strictly guarded* if it explicitly guards all its free variables in the syntactic form of $\alpha(\mathbf{x}) \wedge \varphi(\mathbf{x})$ where, as above, α is an atomic τ-formula in which every variable of $\mathbf{x}$ occurs.

The nesting depth of a formula of GF is the depth w.r.t. guarded quantification steps, bounded by but typically smaller than the first-order quantifier rank. We denote as $\mathrm{GF}_q[\tau]$ the set of GF[τ]-formulae of guarded nesting depth up to q, and use $\equiv^q_{\mathrm{GF}}$ to denote indistinguishability in GF_q.

With a view to GF and its Ehrenfeucht–Fraïssé analysis, the q-round guarded bisimulation game is defined so as to match a single guarded quantification step in a single round.[9]

Consider a game position $\mathfrak{A}, \mathbf{a}; \mathfrak{B}, \mathbf{b}$ where the association between tuples $\mathbf{a}$ and $\mathbf{b}$ can be thought of as the partial map $\rho \colon \mathbf{a} \mapsto \mathbf{b}$ or as a pebble placement. The second player has lost the game if ρ fails to be a local isomorphism. Otherwise the game may continue for a further round as follows.

The first player passes to a restriction $(\rho_0 \colon \mathbf{a}_0 \mapsto \mathbf{b}_0) \subseteq \rho$ (lifting some pebble pairs off the board), and, in one of $\mathfrak{A}$ or $\mathfrak{B}$, completes the residual tuple, $\mathbf{a}_0$ or $\mathbf{b}_0$, to some strictly guarded tuple in that structure (placing pebbles in one of the structures); the second player must extend the other tuple in the opposite structure accordingly (placing matching pebbles in the opposite structure).

The guarded variant of the Ehrenfeucht–Fraïssé theorem states the equivalence of the following:

(i) $\mathfrak{A}, \mathbf{a} \sim^q_{\mathrm{g}} \mathfrak{B}, \mathbf{b}$ (guarded q-bisimilarity);

(ii) the existence of a winning strategy for the second player in the q-round guarded bisimulation game from position $\mathfrak{A}, \mathbf{a}; \mathfrak{B}, \mathbf{b}$;

(iii) $\mathfrak{A}, \mathbf{a} \equiv^q_{\mathrm{GF}} \mathfrak{B}, \mathbf{b}$ (indistinguishability in GF_q).

Again, the existence of a winning strategy for the second player in the infinite game, i.e., full guarded bisimulation equivalence, corresponds to indistinguishability in the infinitary variant of GF (a guarded Karp theorem).

In particular, every formula of GF is preserved under guarded bisimulation, and by natural variants of the classical proof ideas, using one of

$$\equiv_{\mathrm{GF}} \rhd \equiv \qquad \text{over the classes of } \begin{cases} \omega\text{-saturated } \tau\text{-structures, or} \\ \text{guarded } \omega\text{-tree-unfoldings,} \end{cases}$$

[9]The count of rounds inherited from the natural notion in the hypergraph bisimulation game would be up to two times the desired count for GF; the difference goes unnoticed in the infinite game. One could moreover admit moves to guarded rather than strictly guarded configurations without changing the resulting equivalence.

one shows expressive completeness of GF for the class of all FO-definable properties of relational structures that are closed under guarded bisimulation equivalence. The restriction to the classical setting here comes from the passage through infinite tractable companions (besides the use of compactness in the first variant).

Theorem 4.2 (Andréka–van Benthem–Németi)**.** *In the sense of classical model theory, the guarded fragment* GF *captures the class of all first-order definable properties (of guarded tuples in relational structures) that are closed under guarded bisimulation equivalence. In other words, the following are equivalent for any* $\varphi \in \mathrm{FO}(\tau)$*:*

(i) *φ is preserved under guarded bisimulation:*
whenever $\mathfrak{A}, \mathbf{a} \sim_{\mathrm{g}} \mathfrak{B}, \mathbf{b}$*, then* $\mathfrak{A}, \mathbf{a} \models \varphi$ *implies* $\mathfrak{B}, \mathbf{b} \models \varphi$*;*

(ii) *φ is logically equivalent to some* $\varphi' \in \mathrm{GF}[\tau]$.

Towards an expressive completeness argument that neither uses compactness nor passage through infinite structures, we need to find (finite) representatives in the $\sim_{\mathrm{g}}$-classes of (finite) structures that support logical control of the form

$$\equiv_{\mathrm{GF}}^{f(q)} \rhd \equiv^{q} \qquad \text{over } \mathcal{C}.$$

Classes of sufficiently free and sufficiently acyclic guarded covers can serve this purpose. These are derived from the corresponding cover constructions for hypergraphs.

4.2 Tractability in guarded covers

Guarded covers The following is just the analogue of Definition 2.9, w.r.t. guarded bisimulation as a strengthening of hypergraph bisimulation (cf. Definition 4.1).

Definition 4.3. (a) A map $\pi \colon \hat{\mathfrak{A}} \to \mathfrak{A}$ between τ-structures $\hat{\mathfrak{A}}$ and $\mathfrak{A}$ is a *guarded cover*, denoted $\pi \colon \hat{\mathfrak{A}} \xrightarrow{\sim_{\mathrm{g}}} \mathfrak{A}$, if $Z = \{\pi \restriction \hat{s} \colon \hat{s} \text{ guarded in } \hat{\mathfrak{A}}\}$ is a guarded bisimulation.

(b) A guarded cover is called (n, K)-free if it satisfies the analogue of condition (c) in Definition 3.1: for any subset $\hat{B} \subseteq \hat{A}$ of size up to K, every guarded $\hat{s}$ in $\hat{\mathfrak{A}}$ has some guarded companion $\hat{s}'$ in $\hat{\mathfrak{A}}$ for which $\pi(\hat{s}') = \pi(\hat{s})$, $\hat{s}' \cap \hat{B} = \hat{s} \cap \hat{B}$, and the distance between $\hat{s}' \setminus \hat{B}$ and $\hat{B} \setminus \hat{s}'$ in $\hat{\mathfrak{A}} \setminus (\hat{s}' \cap \hat{B})$ is greater than n.

Other properties of τ-structures, like conformality, N-acyclicity, N-chordality are all naturally defined in terms of the associated hypergraphs of guarded subsets. Covers are called conformal, N-acyclic, N-chordal if the covering structure (i.e., its hypergraph of guarded subsets) has this property.

Definition 4.4. A τ-structure $\mathfrak{A}$ is *(n, K)-free* if it satisfies the analogue of condition (b) in Definition 3.1 w.r.t. guarded bisimulation equivalence $\sim_{\mathrm{g}}$: for any subset $B \subseteq A$ of size up to K, every guarded s in $\mathfrak{A}$ has some guarded bisimilar companion s' for which $\mathfrak{A}, s' \sim_{\mathrm{g}} \mathfrak{A}, s$, $s' \cap B = s \cap B$, and the distance between $s' \setminus B$ and $B \setminus s'$ in $\mathfrak{A} \setminus (s' \cap B)$ is greater than n.

As discussed in connection with Definition 3.1 above, (n, K)-freeness of a (guarded) cover may be strictly stronger than (n, K)-freeness of the covering structure.

The projection π in a hypergraph cover $\pi \colon \hat{\mathrm{H}} \xrightarrow{\sim} \mathrm{H}(\mathfrak{A})$ is injective in restriction to every hyperedge of $\hat{\mathrm{H}}$. This allows us to expand any cover of the hypergraph of guarded subsets associated with $\mathfrak{A}$ to become a guarded cover $\pi \colon \hat{\mathfrak{A}} \xrightarrow{\sim_{\mathrm{g}}} \mathfrak{A}$. In fact, there is a unique pull-back of the local relational structure in the guarded subsets of $\mathfrak{A}$ to $\hat{\mathrm{H}}$ that turns π into a guarded cover and $\hat{\mathrm{H}}$ into the hypergraph of guarded subsets of the covering structure $\hat{\mathfrak{A}}$. Writing $\hat{S}$ for the set of hyperedges of $\hat{\mathrm{H}} = (\hat{A}, \hat{S})$, which are to become the guarded subsets of $\hat{\mathfrak{A}}$, the local isomorphism condition on guarded covers forces the interpretations of the relations in $\hat{\mathfrak{A}} := (\hat{A}, (R^{\hat{\mathfrak{A}}})_{R \in \tau})$ to be

$$R^{\hat{\mathfrak{A}}} := \{\hat{\mathbf{a}} \colon [\hat{\mathbf{a}}] \in \hat{S}, \pi(\hat{\mathbf{a}}) \in R^{\mathfrak{A}}\};$$

and with these, $\pi \colon \hat{\mathfrak{A}} \xrightarrow{\sim_{\mathrm{g}}} \mathfrak{A}$ is indeed a guarded cover.

Moreover, (n, K)-freeness of the hypergraph cover directly translates into (n, K)-freeness of the guarded cover. Corollary 3.4 therefore implies that every finite τ-structure $\mathfrak{A}$ admits (n, K)-free guarded covers by finite conformal and N-chordal structures $\hat{\mathfrak{A}}$, for any desired thresholds $n, K, N \in \mathbb{N}$.

By the same token, covers by ω-tree-unfoldings of $\mathrm{H}(\mathfrak{A})$, according to Observation 3.2, induce infinite covers by ω-tree-unfoldings $\mathfrak{A}^{\omega *}$ that are (n, K)-free covers for all $n, K \in \mathbb{N}$ simultaneously, besides being fully acyclic, i.e., tree-decomposable with guarded bags. And again, ω-tree-unfoldings inherit from Observation 1.5 the nice feature that guarded bisimilarity between countable τ-structures implies isomorphy of their ω-tree-unfoldings.

Observation 4.5. *The following classes of τ-structures are representative of the class of all τ-structures up to $\sim_{\mathrm{g}}$:*

(a) the class of ω-tree-unfoldings $\{\mathfrak{A}^{\omega *} \colon \mathfrak{A}$ a τ-structure $\}$;

(b) $\mathcal{C}_{n,K,N}[\tau]$, the class of conformal, N-chordal and (n, K)-free τ-structures, for any choice of parameters $n, K, N \in \mathbb{N}$.

Moreover, those classes in (b) remain similarly representative in restriction to the class of just finite τ-structures.

We next discuss the right choices of parameters n, K, N in relation to q and the width of τ to provide the crucial control

$$(**) \qquad \equiv_{\mathrm{GF}}^{f(q)} \rhd \equiv^{q} \qquad \text{over } \mathcal{C}_{n,K,N}[\tau].$$

The classes $\mathcal{C}_{n,K,N}[\tau]$ thus prove to be the right counterparts for the class of ω-tree-unfoldings for the purposes of a finite model theory analogue of the expressive completeness claim in Theorem 4.2.

A closure operation. One ingredient in the structural analysis towards $(**)$ involves a closure operation, whose properties make it interesting for the study of N-acyclic τ-structures (for sufficiently large N). Similar to convex closures, the idea here is to close some small subset $D \subseteq A$ in $\mathfrak{A}$ under *short chordless paths* (in the Gaifman graph $\mathfrak{G}(\mathfrak{A})$), where short refers to some locality or distance parameter $n \in \mathbb{N}$. We call a subset $A' \subseteq A$ *n-closed* in $\mathfrak{A}$ if every chordless path of length up to n in $\mathfrak{G}(\mathfrak{A})$ that links two nodes in A' fully runs within A'. An *n-closed substructure* of $\mathfrak{A}$ is an induced substructure $\mathfrak{A}' \subseteq \mathfrak{A}$ whose domain A' is n-closed in $\mathfrak{A}$.

In general a τ-structure $\mathfrak{A}$ could have no n-closed subsets other than the trivial ones, $\emptyset$ and A. The *n-closure* of $D \subseteq A$ in $\mathfrak{A}$ is defined to be the $\subseteq$-minimal closed superset of D, which always exists, since any intersection of n-closed sets is n-closed

$$\mathrm{cl}_n(D) := \bigcap\{A' \subseteq A\colon D \subseteq A', A' \text{ } n\text{-closed in } \mathfrak{A}\}.$$

The following is established in [18] and turns n-closures and n-closed substructures into useful building blocks for back-and-forth arguments in structures that are N-acyclic for sufficiently large N. I should stress that the arguments in [18] are far from yielding useful insights into the actual bounds.

Lemma 4.6. *If N is sufficiently large in relation to m, n and the width of τ, then there is a uniform bound on the size of the n-closure of any subset of up to m elements in any N-acyclic τ-structure $\mathfrak{A}$.*

By choosing N large enough to bound even the uniform bound of the lemma, we therefore get the following.

Corollary 4.7. *If N is sufficiently large in relation to m, n and the width of τ, then the n-closure of up to m elements in any N-acyclic τ-structure $\mathfrak{A}$ induces an n-closed substructure of bounded size that is tree-decomposable with guarded bags (within $\mathfrak{A}$).*

The isomorphism type of an induced substructure that admits a tree-decomposition with guarded bags in $\mathfrak{A}$ is determined up to multiplicities by its guarded bisimulation type. A bound on its size moreover entails a bound on the guarded nesting depth of the characteristic GF-formula. In this situation even the GF_d-type for sufficiently large d suffices for the complete description up to multiplicities. Freeness conditions guarantee that for small substructures also the multiplicities can be matched, and the availability of further extensions. These ingredients, augmented by suitable composition arguments and fast growing sequences of bounds that allow for q iterations, can be put together to give the necessary back-and-forth systems. These systems consist of local isomorphisms between closed substructures of bounded size of τ-structures in $\mathcal{C}_{n,K,N}[\tau]$. Parameters n, K, N need to be sufficiently large to guarantee the required back-and-forth extensions. For suitable choices, these back-and-forth systems then establish $(**)$.

For τ-structures $\mathfrak{A}$ and $\mathfrak{B}$, consider finite local isomorphisms ρ from $\mathfrak{A}$ to $\mathfrak{B}$ whose domain $A' \subseteq A$ and image $B' \subseteq B$ are regarded as the domains of finite induced substructures $\mathfrak{A}' \subseteq \mathfrak{A}$ and $\mathfrak{B}' \subseteq \mathfrak{B}$. Let us indicate this situation as $\rho\colon \mathfrak{A}' \mapsto \mathfrak{B}'$, and further as

$$\rho\colon \mathfrak{A}' \longmapsto^{q}_{\mathrm{GF}} \mathfrak{B}'$$

if ρ is compatible with $\mathrm{GF}_q[\tau]$-types: if $\mathbf{a} \in \mathrm{dom}(\rho)$ is guarded in $\mathfrak{A}$, then $\mathfrak{A}, \mathbf{a} \equiv^{q}_{\mathrm{GF}} \mathfrak{B}, \rho(\mathbf{a})$. Then, for suitable parameter sequences m_k, q_k and n_k, the system $(I_k)_{k \leqslant q}$ with

$$I_k := \big\{(\rho\colon \mathfrak{A}' \longmapsto^{q_k}_{\mathrm{GF}} \mathfrak{B}')\colon |\rho| \leqslant m_k, \mathfrak{A}' \subseteq \mathfrak{A} \text{ and } \mathfrak{B}' \subseteq \mathfrak{B}\ n_k\text{-closed}\ \big\}$$

satisfies the usual back-and-forth conditions, provided $\mathfrak{A}, \mathfrak{B} \in \mathcal{C}_{n,K,N}[\tau]$ for sufficiently large n, K, N.

Corollary 4.8. *For sufficiently large choices of $n, K, N \in \mathbb{N}$, and for some sufficiently fast growing function f, the class $\mathcal{C}_{n,K,N}[\tau]$ of τ-structures that are conformal, N-chordal and (n, K)-free, is representative of all τ-structures up to guarded bisimulation $\sim_{\mathrm{g}}$ both generally and among finite structures and such that $\sim_{\mathrm{g}}^{f(q)} \rhd \equiv^q$ over $\mathcal{C}_{n,K,N}[\tau]$.*

The Ehrenfeucht–Fraïssé argument underlying this $\rhd$-assertion over $\mathcal{C}_{n,K,N}[\tau]$ does not rely on Gaifman locality, but provides an alternative approach to a winning strategy in the q-round Ehrenfeucht–Fraïssé game over a very special class of structures, via specifically adapted hierarchical tree-like decompositions. Because we are not working in tree-like or even locally tree-like structures, however, the configurations that are being matched by these strategies are not full neighbourhoods of some bounded radius, but size-bounded small tree-decomposable configurations that arise as induced substructures and will typically be strictly contained in the Gaifman-local neighbourhoods in question. The relevant closures are closures under short chordless paths, rather than under short distances. To appreciate the difference, it may be instructive to recall that the whole of the structural complexity that arises in these game arguments, may arise in the 1-neighbourhood of single points. Towards an extreme case, let us wrap up the given structures $\mathfrak{A}$ and $\mathfrak{B}$ in structures that add a single extra pivot vertex to each structure and replace every relation of arity r in either structure by its conical extension to an $(r+1)$-arity relation that uniformly joins the new vertex as a new last component to every tuple in the original relation. This process preserves every level of guarded bisimulation as well as of first-order equivalence. It is also essentially compatible with the structural criteria that govern our technique, but trivialises Gaifman locality.

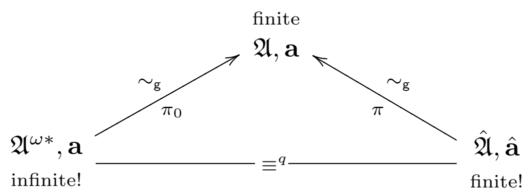
The argument for the control of $\equiv^q$ by $\sim_{\mathrm{g}}^{f(q)}$ also points to an interesting relationship between the finite cover constructions that take us into $\mathcal{C}_{n,K,N}[\tau]$ and the more straightforward but infinite ω-tree-unfoldings that are available only in the classical setting.[10] Note that the latter also take us to $\mathcal{C}_{n,K,N}[\tau]$, by Observation 3.2. So we obtain the following.

[10] I am indebted to Balder ten Cate for discussions that brought out this point. This is important for a new application in Section 4.4.

Corollary 4.9. *For each* $q \in \mathbb{N}$, *every finite* τ*-structure* $\mathfrak{A}$ *admits a* finite *guarded cover* $\pi\colon \hat{\mathfrak{A}} \xrightarrow{\sim_{\mathrm{g}}} \mathfrak{A}$ *which is indistinguishable in* FO_q *from its (infinite)* ω*-tree-unfolding* $\mathfrak{A}^{\omega *}$.

Note that, a posteriori, the ω-tree-unfoldings of $\mathfrak{A}$ and $\hat{\mathfrak{A}}$ are seen to be isomorphic, by the analogue of Observation 1.5 for ω-tree-unfoldings of τ-structures.

In a sense, the first-order theory of the very uniform and necessarily infinite ω-tree unfoldings can be approximated to any desired degree by the first-order theories of finite covers.

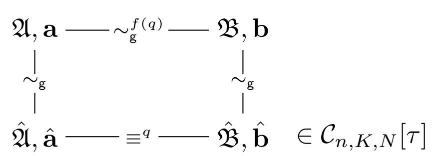


4.3 Expressive completeness of GF

On the basis of the above, it is straightforward to establish the expressive completeness of GF for the class of all FO-definable properties that are closed under guarded bisimulation equivalence, Theorem 4.10 below, which is proved in [18].

Every $\equiv_{\mathrm{GF}}^{f(q)}$ is of finite index, and each $\equiv_{\mathrm{GF}}^{f(q)}$-class $\mathrm{GF}_{f(q)}$-definable. It therefore suffices to show that a given $\varphi \in \mathrm{FO}_q$ that is preserved under $\sim_{\mathrm{g}}$ (over all finite τ-structures) is in fact preserved under $\equiv_{\mathrm{GF}}^{f(q)}$ (over all finite τ-structures), for some suitable choice of $f(q)$.

In the case relevant for finite model theory this is immediate from Corollary 4.8. If $\varphi(\mathbf{x}) \in \mathrm{FO}_q$ and $\mathfrak{A}, \mathbf{a} \sim_{\mathrm{g}}^{f(q)} \mathfrak{B}, \mathbf{b}$ for finite τ-structures $\mathfrak{A}$ and $\mathfrak{B}$ with guarded tuples $\mathbf{a}$ and $\mathbf{b}$, we pass to finite $\hat{\mathfrak{A}}, \hat{\mathbf{a}} \sim_{\mathrm{g}} \mathfrak{A}, \mathbf{a}$ and $\hat{\mathfrak{B}}, \hat{\mathbf{b}} \sim_{\mathrm{g}} \mathfrak{B}, \mathbf{b}$ in the appropriate $\mathcal{C}_{n,K,N}[\tau]$, which guarantees that $\hat{\mathfrak{A}}, \hat{\mathbf{a}} \sim_{\mathrm{g}}^{f(q)} \hat{\mathfrak{B}}, \hat{\mathbf{b}}$ implies $\hat{\mathfrak{A}}, \hat{\mathbf{a}} \equiv^q \hat{\mathfrak{B}}, \hat{\mathbf{b}}$.



Now $\hat{\mathfrak{A}}, \hat{\mathbf{a}} \sim_{\mathrm{g}}^{f(q)} \hat{\mathfrak{B}}, \hat{\mathbf{b}}$ follows from $\hat{\mathfrak{A}}, \hat{\mathbf{a}} \sim_{\mathrm{g}} \mathfrak{A}, \mathbf{a} \sim_{\mathrm{g}}^{f(q)} \mathfrak{B}, \mathbf{b} \sim_{\mathrm{g}} \hat{\mathfrak{B}}, \hat{\mathbf{b}}$, and then $\hat{\mathfrak{A}}, \hat{\mathbf{a}} \equiv^q \hat{\mathfrak{B}}, \hat{\mathbf{b}}$ further implies that $\hat{\mathfrak{A}}, \hat{\mathbf{a}} \models \varphi$ iff $\hat{\mathfrak{B}}, \hat{\mathbf{b}} \models \varphi$ as $\varphi \in \mathrm{FO}_q$. By $\sim_{\mathrm{g}}$-preservation of φ over finite structures this translates into $\mathfrak{A}, \mathbf{a} \models \varphi$ iff $\mathfrak{B}, \mathbf{b} \models \varphi$, since $\mathfrak{A}, \mathbf{a} \sim_{\mathrm{g}} \hat{\mathfrak{A}}, \hat{\mathbf{a}}$ and $\mathfrak{B}, \mathbf{b} \sim_{\mathrm{g}} \hat{\mathfrak{B}}, \hat{\mathbf{b}}$.

Theorem 4.10. *Also in the sense of finite model theory, the guarded fragment* GF *captures the class of all first-order definable properties (of guarded tuples in relational structures) that are closed under guarded bisimulation equivalence. The following are equivalent for any* $\varphi \in \mathrm{FO}(\tau)$:

(i) φ *is preserved under guarded bisimulation among finite* τ*-structures;*
(ii) φ *is logically equivalent over all finite* τ*-structures to some* $\varphi' \in \mathrm{GF}[\tau]$.

4.4 Expressive completeness of kGNF

In this section we carry the above analysis further to achieve a similar expressive completeness result, also in finite model theory, for a newly prominent extension of the guarded fragment. This *guarded negation fragment* of [6] stems from the ongoing search for ever more expressive fragments of first-order logic that share some of the important model theoretic and algorithmic benefits of modal logic. The guarded fragment itself belongs in this tradition, with its very natural generalisation of modal quantification, and with properties like the finite model property and decidability, good bounds on small models, good model checking and decidability complexities, preservation and expressive completeness, etc. The recent proposal of the *unary negation fragment* in [21], which similarly shares good properties with ML, has shifted the focus from restricted quantification patterns to restricted negation. This fragment is built on (unrestricted) existential quantification (and no universal quantification as a basic), and just formulae with at most one free variable can be negated. It may thus be viewed as the extension of the positive existential fragment $\exists$posFO by a negation operation that only applies to 'unary' formulae. The *guarded negation fragment* of [6] goes further in allowing a negation operation on formulae whose tuple of free variables is explicitly guarded. Both fragments therefore extend $\exists$posFO by restricted forms of negation, which may be nested. Neither preservation under homomorphisms nor under guarded bisimulation applies; instead, the characteristic structural equivalences turn out to be interesting convolutions of (local) homomorphisms with bisimulation-like (back-and-forth) equivalences.

In the context of this paper, an expressive completeness argument for a variant of GNF w.r.t. to its characteristic preservation property (preservation under *guarded negation bisimulation*), which works in finite model theory, may serve as a good further test for the methods developed.

Guarded negation bisimulation. We present a bisimulation game that, besides local isomorphisms between strictly guarded tuples also involves local homomorphisms that respect strictly guarded tuples.

Definition 4.11. Call a homomorphism $h\colon \mathfrak{A} \xrightarrow{\text{hom}} \mathfrak{B}$ between τ-structures *guarded* if it is injective in restriction to every strictly guarded tuple of its domain; in shorthand notation, $h\colon \mathfrak{A} \xrightarrow{\text{ghom}} \mathfrak{B}$.

Note that a guarded homomorphism is a local isomorphism in restriction to the induced substructures on strictly guarded tuples of its domain. The homomorphic projections π in guarded covers $\pi\colon \hat{\mathfrak{A}} \xrightarrow{\sim_{\mathrm{g}}} \mathfrak{A}$ are special examples of guarded homomorphism.

Valid positions in the guarded negation bisimulation game over τ-structures $\mathfrak{A}$ and $\mathfrak{B}$ are local isomorphisms between strictly guarded tuples, $\rho\colon \mathbf{a} \mapsto \mathbf{b}$, just as in the guarded bisimulation game. But a single round involves a more intricate protocol based on finite guarded homomorphisms.

In position $\mathfrak{A}, \mathbf{a}; \mathfrak{B}, \mathbf{b}$, where $\rho\colon \mathbf{a} \mapsto \mathbf{b}$ is a local isomorphism between the strictly guarded substructures $\mathfrak{A} \restriction [\mathbf{a}]$ and $\mathfrak{B} \restriction [\mathbf{b}]$, the first player chooses some finite subset $A_0 \subseteq A$ or some finite $B_0 \subseteq B$. The second player needs to choose a guarded homomorphism h from the induced substructure on this subset to the opposite structure that respects ρ in the overlaps $[\mathbf{a}] \cap A_0$ or $[\mathbf{b}] \cap B_0$, as the case may be. (These overlaps may be empty.) Finally the first player chooses any strictly guarded tuple in the domain of h, which is paired with its h-image in the opposite structure to give the new position.

Consider, for example, a round in which the first player chooses a subset $A_0 \subseteq A$ on the $\mathfrak{A}$-side; in response, the second player must come up with some $h\colon \mathfrak{A} \restriction A_0 \stackrel{\text{ghom}}{\longrightarrow} \mathfrak{B}$ that maps $\mathbf{a} \restriction A_0$ to $\rho(\mathbf{a} \restriction A_0)$; then the first player chooses some strictly guarded tuple $\mathbf{a}'$ in A_0, and the new position is determined as $\mathfrak{A}, \mathbf{a}'; \mathfrak{B}, h(\mathbf{a}')$.

Again, the natural liberalisation that allows for guarded rather than strictly guarded configurations in game positions results in essentially the same notion of equivalence.

We are interested in the restriction of this game protocol that imposes a bound k on the size of the subsets A_0 or B_0 that the first player is allowed to choose. The existence of a strategy for the second player in the infinite or in the q-round version of the game with size bound k, in position $\mathfrak{A}, \mathbf{a}; \mathfrak{B}, \mathbf{b}$, is denoted as $\mathfrak{A}, \mathbf{a} \sim_{\mathsf{gn}[k]} \mathfrak{B}, \mathbf{b}$ or $\mathfrak{A}, \mathbf{a} \sim^q_{\mathsf{gn}[k]} \mathfrak{B}, \mathbf{b}$, respectively. A definition of these equivalences in terms of appropriate back-and-forth systems would be straightforward; the outer format would be the same as for the guarded bisimulation equivalences, only with the more complex back-and-forth requirements to reflect the above rules for a single round.

Note that, as seen from the first player, this game is a strengthening of the guarded bisimulation game, provided k is at least the width of τ: the first player may choose subsets A_0/B_0 of the form $[\mathbf{a}']/[\mathbf{b}']$, to which the second player is forced to respond with a local isomorphism that determines a strictly guarded tuple in the opposite structure, which must also respect the overlap.

The guarded negation fragments kGNF. We define the *k-bounded guarded negation fragment* $k\mathrm{GNF}[\tau] \subseteq \mathrm{FO}[\tau]$ as follows. The formulae we are really interested in are the strictly guarded ones. As with GF it does not hurt, however, to include outer boolean combinations. These fragments are slight variants of the k-bounded guarded negation fragments of [6]; they are mere syntactic variants as far as their strictly guarded formulae are concerned, but differ in their expressiveness for formulae with unguarded tuples of free variables.

Let $k\mathrm{GNF}[\tau]$ stand for the syntactic fragment of $\mathrm{FO}[\tau]$ generated from atomic τ-formulae $\alpha(\mathbf{x})$ by closure under

- the boolean connectives $\wedge, \vee, \neg$, and
- existential quantifications over conjunctions of strictly guarded formulae,

of the form

$$\alpha(\mathbf{x}) \wedge \exists \mathbf{z} \bigwedge_i (\alpha_i(\mathbf{z}^i) \wedge \varphi_i(\mathbf{z}^i)),$$

where the $\mathbf{z}^i$ are tuples of variables among $\mathbf{x}$ and $\mathbf{z}$, the tuples $\mathbf{z}^i$ together consist of up to k distinct variables, α and the α_i are atomic τ-formulae in which all the displayed variables occur (and, by general convention, the free variables of φ_i are among those displayed).

This definition of $k\mathrm{GNF}[\tau]$ is designed to match the rules of k-bounded guarded negation bisimulation, with the expected Ehrenfeucht–Fraïssé and Karp theorems. For τ-structures $\mathfrak{A}, \mathbf{a}$ and $\mathfrak{B}, \mathbf{b}$ with strictly guarded tuples $\mathbf{a}$ and $\mathbf{b}$, the following are equivalent:

(i) $\mathfrak{A}, \mathbf{a} \sim^q_{\mathsf{gn}[k]} \mathfrak{B}, \mathbf{b}$;

(ii) $\mathfrak{A}, \mathbf{a} \equiv^q_{k\mathrm{GNF}} \mathfrak{B}, \mathbf{b}$.

Letting $\equiv^\infty_{k\mathrm{GNF}}$ denote equivalence in the infinitary variant of kGNF, the following are also equivalent:

(i) $\mathfrak{A}, \mathbf{a} \sim_{\mathsf{gn}[k]} \mathfrak{B}, \mathbf{b}$;

(ii) $\mathfrak{A}, \mathbf{a} \equiv^\infty_{k\mathrm{GNF}} \mathfrak{B}, \mathbf{b}$.

A corresponding capturing result for the classical context is given in [6]. This classical variant of expressive completeness of kGNF for FO-definable properties (of strictly guarded tuples in relational structures) that are closed under $\sim_{\mathsf{gn}[k]}$ can be proved using suitable tree unfoldings and/or ω-saturated models and compactness. We shall return to the issue below.

Observation 4.12. *Every guarded cover* $\pi \colon \mathfrak{A} \times_\rho G \xrightarrow{\sim_{\mathrm{g}}} \mathfrak{A}$ *by a reduced product as in Definition 3.3 is also a guarded negation cover in the sense that the set of all restrictions of* π *to strictly guarded tuples of* $\mathfrak{A} \times_\rho G$ *is a back-and-forth system for* $\sim_{\mathsf{gn}[k]}$ *(for all* k*, and even without any size bound).*

It seems that essential model-theoretic arguments for kGNF can be reduced to related issues for GF. The basic idea, which we shall also encounter in our study below, is the following.

Existential quantifications of the kind allowed in $k\mathrm{GNF}[\tau]$ assert the existence, at some strictly guarded tuple $\mathbf{a}$, of guarded homomorphic images of certain τ-structures of size up to $k + \mathrm{width}(\tau)$, whose strictly guarded image tuples individually need to satisfy certain further $k\mathrm{GNF}[\tau]$ properties. If each of these required homomorphic images is guarded, as a whole, by means of a new relation R of arity k in an expanded vocabulary $\tau_R := \{R\}$, then the existence of these homomorphic images becomes expressible in $\mathrm{GF}[\tau_R]$.

The non-existence of a guarded homomorphic image of some τ-structure at a strictly guarded tuple $\mathbf{a}$, on the other hand, is only very partially expressible in $\mathrm{GF}[\tau_R]$ – after all, it is *not* preserved under guarded bisimulation even of the expanded structures. Just acyclic, guarded tree-decomposable images can be forbidden in GF. But then all other – viz., cyclically embedded – configurations of bounded size can be ruled out by passage to suitable N-acyclic guarded covers. In fact, in this context the weakly N-acyclic covers of [5] come into their own (cf. Theorem 2.14 here). These covers were primarily designed to show that formulae of GF that have infinite models without homomorphic images of some finite collection of finite configurations also have finite models that avoid these homomorphisms. In other words, GF is shown to have the finite model property in restriction to every class of τ-structures that is defined in terms of a finite number of forbidden finite homomorphic images.

In particular, this reduction idea has been used in [6] to establish the finite model property for kGNF.

Theorem 4.13. *[Barany–ten Cate–Segoufin] The logics kGNF have the finite model property.*

Here we want to explore expressive completeness of kGNF for FO properties that are closed under $\sim_{\mathsf{gn}[k]}$ with methods that work for finite model theory. More specifically, we seek a construction of finite representatives in the $\sim_{\mathsf{gn}[k]}$-classes of finite τ-structures over which kGNF controls FO in the sense of

$$(\dagger\dagger) \qquad \sim_{\mathsf{gn}[k]}^{f(q)} \rhd \equiv^{q} .$$

In the following we always assume that $k \geqslant w$ where w is the width of the vocabulary τ. This is simply to ensure that guarded back-and-forth extensions can be captured in the framework of kGNF.

Tree unfoldings that are k-rich. Recall that an induced substructure relationship $\mathfrak{A}_0 \subseteq \mathfrak{A}$ does not imply an induced substructure relationship between the Gaifman graphs or the hypergraphs of guarded subsets, because a non-trivial intersection $[\mathbf{a}] \cap A_0$ for strictly guarded tuples $\mathbf{a}$ of $\mathfrak{A}$ need not be guarded in $\mathfrak{A}_0$. This problem, and that of just weak rather than induced substructure relationships between a guarded homomorphic image and its surrounding structure, need to be overcome in reductions that seek to capture levels of $\sim_{\mathsf{gn}[k]}$-equivalence of τ-structures in terms of levels of $\sim_{\mathsf{g}}$-equivalence of expansions, which are introduced to render the relevant homomorphic images guarded.

Definition 4.14. We say that $\mathfrak{A}_0 \subseteq \mathfrak{A}$ is a *strict substructure* relationship, denoted $\mathfrak{A}_0 \subseteq_{\mathsf{s}} \mathfrak{A}$, if $\mathrm{H}(\mathfrak{A}_0) = \mathrm{H}(\mathfrak{A}) \restriction A_0$. For a strictly guarded parameter tuple $\mathbf{a}$ in $\mathfrak{A}$, we say that $\mathfrak{A}_0$ is a *strict substructure over* $\mathbf{a}$, denoted $\mathfrak{A}_0 \subseteq_{\mathsf{s}} \mathfrak{A}/\mathbf{a}$, if the extension of $\mathrm{H}(\mathfrak{A}_0)$ by all subsets of $A_0 \cap [\mathbf{a}]$ is equal to the induced sub-hypergraph $\mathrm{H}(\mathfrak{A}) \restriction A_0$.

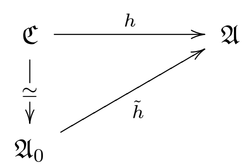
If $\mathfrak{A}_0 \subseteq_{\mathsf{s}} \mathfrak{A}$, then the intersection of any guarded subset of $\mathfrak{A}$ with $\mathfrak{A}_0$ is contained in a subset $[\mathbf{a}_0]$ for some strictly guarded tuple $\mathbf{a}_0$ of $\mathfrak{A}_0$. For $\mathfrak{A}_0 \subseteq_{\mathsf{s}} \mathfrak{A}/\mathbf{a}$ this condition is relaxed to allow containment in some such $[\mathbf{a}_0]$ or in $[\mathbf{a}]$.

In preparation for a reduction from kGNF to GF we look at another notion of richness, that captures a form of (finitary) saturation w.r.t. homomorphic images of bounded size. For convenience we introduce the ad-hoc short notation

$$h \colon \mathfrak{C}, \mathbf{c} \stackrel{\text{ghom}}{\longrightarrow} \mathfrak{A}/\mathbf{a}$$

to say that h is a guarded homomorphism from $\mathfrak{C}$ to $\mathfrak{A}$ which maps the tuple $\mathbf{c}$ injectively into $[\mathbf{a}]$. We use this in connection with strictly guarded $\mathbf{a}$, but $\mathbf{c}$ is not required to be guarded in $\mathfrak{C}$.

Definition 4.15. A τ-structure $\mathfrak{A}$ is called *k-rich* if, for every strictly guarded tuple $\mathbf{a}$ in $\mathfrak{A}$ and every guarded homomorphism $h\colon \mathfrak{C}, \mathbf{c} \stackrel{\text{ghom}}{\longrightarrow} \mathfrak{A}/\mathbf{a}$ from a τ-structure $\mathfrak{C}$ of size up to k, there exists a strict substructure $\mathfrak{A}_0 \subseteq_{\mathsf{s}} \mathfrak{A}/\mathbf{a}$ over $\mathbf{a}$ and an isomorphism $\rho\colon \mathfrak{C} \simeq \mathfrak{A}_0$ such that $\rho(\mathbf{c}) = h(\mathbf{c})$, $[h(\mathbf{c})] = [\mathbf{a}] \cap A_0$, and $\tilde{h} := h \circ \rho^{-1}$ is compatible with $\sim_{\mathsf{gn}[k]}$ in the sense that for every strictly guarded tuple $\mathbf{a}'$ of $\mathfrak{A}_0$: $\mathfrak{A}, \mathbf{a}' \sim_{\mathsf{gn}[k]} \mathfrak{A}, \tilde{h}(\mathbf{a}')$.



Note that, being an isomorphic copy of the guarded homomorphism h, the map $\tilde{h} = h \circ \rho^{-1}\colon \mathfrak{A}_0 \longrightarrow \mathfrak{A}$ also is a guarded homomorphism.

It may be worth pointing out that k-richness typically postulates small cyclic configurations. For instance an undirected graph $\mathfrak{A}$ that has paths of length 2 must also have chordless 4-cycles if it is to be 4-rich.

It would be nice to have a direct finitary construction that yields k-rich $\sim_{\mathsf{gn}[k]}$-covers, for instance via reduced products. This remains open for now, and instead we here make a detour through infinite tree unfoldings that are k-rich, together with an appeal to a finite model property.

Proposition 4.16. *Every finite τ-structure $\mathfrak{A}$ admits an infinite tree-like $\sim_{\mathsf{gn}[k]}$-cover $\pi\colon \mathfrak{A}^* \stackrel{\sim_{\mathsf{gn}[k]}}{\longrightarrow} \mathfrak{A}$ by some k-rich structure $\mathfrak{A}^*$.*

Proof. Fix a finite τ-structure $\mathfrak{A}$. We describe the construction of a k-rich unfolding of $\mathfrak{A}$ from some strictly guarded root tuple $\mathbf{a}_0$, as a $\sim_{\mathsf{gn}[k]}$-cover

$$\pi\colon \mathfrak{A}^* \stackrel{\sim_{\mathsf{gn}[k]}}{\longrightarrow} \mathfrak{A}.$$

Instead of the τ-structure $\mathfrak{A}^*$ itself we construct an expansion that interprets one new k-ary relation R. For $\tau_R = \tau \cup \{R\}$ we describe a tree-like τ_R-structure $\mathfrak{T}(\mathfrak{A})$ whose τ-reduct will be the desired $\mathfrak{A}^*$.

$\mathfrak{T}$ is obtained as the limit of a regular process of stepwise extensions of finite initial segments of $\mathfrak{T}$ in their leaf tuples. These leaf tuples of intermediate stages are strictly guarded tuples of $\mathfrak{T}$, whose π-images are strictly guarded tuples of $\mathfrak{A}$. We identify the root tuple $\mathbf{a}^*$ for the construction in $\mathfrak{T}$ with the designated tuple $\mathbf{a}_0$ in $\mathfrak{A}$.

As further structural building blocks for $\mathfrak{T}$ we use a finite family of guarded homomorphisms $h\colon \mathfrak{C}, \mathbf{c} \stackrel{\text{ghom}}{\longrightarrow} \mathfrak{A}, /\mathbf{a}$ from τ-structures $\mathfrak{C}$ of size up to k, where $\mathbf{c}$ is mapped injectively into some strictly guarded $[\mathbf{a}]$ in $\mathfrak{A}$. Let this family, enumerated as $(h_i, \mathfrak{C}_i, \mathbf{c}_i)_{i \leqslant I}$, represent every such object up to isomorphism. By $\tilde{\mathfrak{C}}_i$ we denote the expansion of $\mathfrak{C}_i$ to a τ_R-structure that interprets R as the full k-ary relation over $\mathfrak{C}_i$.

An isomorphic copy of $\mathfrak{A} \restriction \mathbf{a}_0$ forms the root configuration and initial stage in the construction of $\mathfrak{T}$; this is also the leaf tuple at this stage. Inductively, we extend in each new stage every leaf tuple $\mathbf{a}^*$ above $\mathbf{a} = \pi(\mathbf{a}^*)$ as follows. For every $i \in I$ such that h_i maps $\mathbf{c}_i$ injectively into $[\mathbf{a}]$, we introduce a fresh isomorphic copy of $\tilde{\mathfrak{C}}_i$ and identify the components of $\mathbf{c}_i$ in this new copy of $\tilde{\mathfrak{C}}_i$ with components of $\mathbf{a}^*$ according to h_i. The projection π is extended accordingly, and all strictly guarded tuples in the attached copy of $\mathfrak{C}_i$ that are not fully contained in $[\mathbf{a}^*]$ become leaf tuples for the next stage. In this manner it is guaranteed that the copy of $\mathfrak{C}_i$ becomes a strict substructure over $\mathbf{a}^*$ in the τ-reduct of $\mathfrak{T}$.

We let $\mathfrak{T}$ be the countably infinite τ_R-structure that is reached in the limit, $\mathfrak{A}^*$ its τ-reduct.

Clearly $\pi \colon \mathfrak{A}^* \xrightarrow{\text{ghom}} \mathfrak{A}$ by construction. Moreover, every strictly guarded tuple $\mathbf{a}^*$ in $\mathfrak{A}^*$ above $\mathbf{a} = \pi(\mathbf{a}^*)$ has undergone the extension step described above. It is therefore associated with the required isomorphic embeddings of structures $\mathfrak{C}, \mathbf{c}$ as strict substructures of $\mathfrak{A}^*$ over $\mathbf{a}^*$. For every $h \colon \mathfrak{C}, \mathbf{c} \xrightarrow{\text{ghom}} \mathfrak{T}/\mathbf{a}^*$ of size up to k the isomorphism type of $\pi \circ h \colon \mathfrak{C}, \mathbf{c} \xrightarrow{\text{ghom}} \mathfrak{A}/\pi(\mathbf{a}^*)$ agrees with one of the $h_i, \mathfrak{C}_i, \mathbf{c}_i$.

Finally, π is a $\sim_{\mathsf{gn}[k]}$-covering: its restrictions to the strictly guarded tuples of $\mathfrak{A}^*$ satisfy the back-and-forth conditions for $\sim_{\mathsf{gn}[k]}$ by construction. $\square$

We note that $\mathfrak{T} = \mathfrak{T}(\mathfrak{A})$ in the proof can be replaced by its ω-tree unfolding $\mathfrak{T}^{\omega *}$ without affecting the claims. Crucially, the k-richness of these unfoldings serves to lift $\sim_{\mathsf{gn}[k]}$-equivalence to $\sim_{\mathsf{g}}$-equivalence, level by level.

Observation 4.17. *If $\mathfrak{A}, \mathbf{a} \sim^m_{\mathsf{gn}[k]} \mathfrak{B}, \mathbf{b}$, then the above construction yields tree-decomposable τ_R-structures $\mathfrak{T}(\mathfrak{A})$ and $\mathfrak{T}(\mathfrak{B})$ such that $\mathfrak{T}(\mathfrak{A}), \mathbf{a} \sim^m_{\mathsf{g}} \mathfrak{T}(\mathfrak{B}), \mathbf{b}$.*

Proof. Consider strictly guarded tuples $\mathbf{a}$ and $\mathbf{b}$ such that $\mathfrak{A}, \mathbf{a} \sim^m_{\mathsf{gn}[k]} \mathfrak{B}, \mathbf{b}$ for $m \geqslant 1$. In view of the construction of $\mathfrak{T}(\mathfrak{A})$ and $\mathfrak{T}(\mathfrak{B})$ it is essential to show that the k-bounded guarded homomorphisms $h \colon \mathfrak{C}, \mathbf{c} \xrightarrow{\text{ghom}} \mathfrak{A}/\mathbf{a}$ and those at $\mathfrak{B}/\mathbf{b}$ are the same, up to isomorphism.

This follows from the fact that the first player may, for instance, propose the image set $A_0 := h(C)$ as a challenge in the $\sim_{\mathsf{gn}[k]}$-game: the second player has to respond with a guarded homomorphism $h' \colon h(\mathfrak{C}) \xrightarrow{\text{ghom}} \mathfrak{B}$ that maps $\mathbf{c}$ injectively into $[\mathbf{b}]$, because h' respects the bijection $\rho \colon \mathbf{a} \mapsto \mathbf{b}$. Then $h' \circ h$ is the desired match for h at $\mathfrak{B}/\mathbf{b}$. The remainder of this round in the $\sim_{\mathsf{gn}[k]}$-game also guarantees that every extension step in the $\sim_{\mathsf{g}}$-game on $\mathfrak{T}(\mathfrak{A})$ and $\mathfrak{T}(\mathfrak{B})$ involving some strictly guarded tuple $\mathbf{a}'$ (not necessarily corresponding to the whole of the isomorphic copy of $\tilde{\mathfrak{C}}$ in $\mathfrak{T}(\mathfrak{A})$, but maybe just one of the strictly guarded tuples in the τ-structure) can be matched by translation via h'.

It is important for this argument that the building blocks $\mathfrak{C}$ used in the generation of $\mathfrak{T}(\mathfrak{A})$ and $\mathfrak{T}(\mathfrak{B})$ at leaves above $\mathbf{a}$ and $\mathbf{b}$ arise in isomorphic pairs whenever $\mathfrak{A}, \mathbf{a} \sim^1_{\mathsf{gn}[k]} \mathfrak{B}, \mathbf{b}$ and that they are embedded into $\mathfrak{T}(\mathfrak{A})$ and $\mathfrak{T}(\mathfrak{B})$ as strict substructures over these parameter tuples. □

If there is any finite τ_R-structure $\tilde{\mathfrak{A}}_0$ such that $\tilde{\mathfrak{A}}_0 \sim_{\mathsf{g}} \mathfrak{T} = \mathfrak{T}(\mathfrak{A})$ then its ω-tree-unfolding must be isomorphic to $\mathfrak{T}^{\omega *}$ (cf. the remarks after Corollary 4.9). Existence of a finite τ_R-structure of this kind follows from the finite model property of GF, because, due to its regularity, $\mathfrak{T}$ realises only finitely many GF-types. If instead we apply the finite model property for kGNF$[\tau_R]$ from [6], see Theorem 4.13, then by the same token we even find a finite τ_R-structure $\tilde{\mathfrak{A}}_0$ for which simultaneously

$$(\#) \qquad \begin{array}{rcl} \tilde{\mathfrak{A}}_0|_\tau & \sim_{\mathsf{gn}[k]} & \mathfrak{T}(\mathfrak{A})|_\tau \;\sim_{\mathsf{gn}[k]}\; (\mathfrak{T}(\mathfrak{A}))^{\omega *}|_\tau \;\sim_{\mathsf{gn}[k]}\; \mathfrak{A}, \\ \tilde{\mathfrak{A}}_0 & \sim_{\mathsf{g}} & \mathfrak{T}(\mathfrak{A}), \text{ and} \\ (\tilde{\mathfrak{A}}_0)^{\omega *} & \simeq & (\mathfrak{T}(\mathfrak{A}))^{\omega *}. \end{array}$$

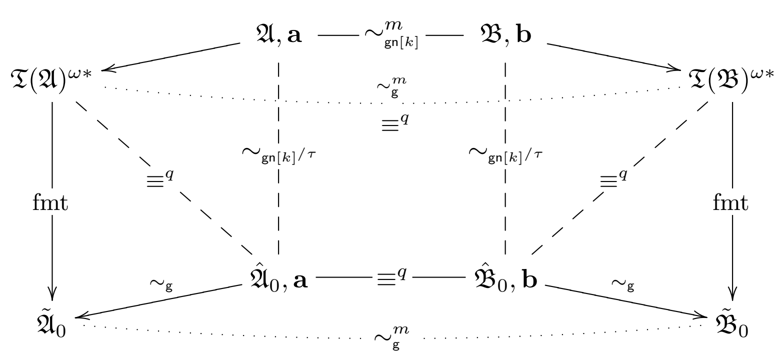
We may moreover replace any such $\tilde{\mathfrak{A}}_0$ by a finite (n, K)-free, N-acyclic cover $\pi \colon \hat{\mathfrak{A}}_0 \xrightarrow{\sim_{\mathsf{g}}} \tilde{\mathfrak{A}}_0$, for any choice of n, K, N, without changing its $\sim_{\mathsf{g}}$-type or any of the above three conditions. With these preparations, the desired expressive completeness proof is straightforward.

Theorem 4.18. *Both classically and in the sense of finite model theory,* kGNF *captures the class of all first-order definable properties (of strictly guarded tuples in relational structures) that are closed under* $\sim_{\mathsf{gn}[k]}$*-equivalence. I.e., the following are equivalent for any* $\varphi \in \mathrm{FO}(\tau)$*:*

- *(i)* φ *is preserved under* $\sim_{\mathsf{gn}[k]}$ *(among finite* τ*-structures);*
- *(ii)* φ *is logically equivalent (over all finite* τ*-structures) to some* $\varphi' \in k\mathrm{GNF}[\tau]$*.*

Proof. We focus on the finite model theory version. The classical version can be proved with the usual methods, or along the same lines as in this proof.

It suffices to show that any $\varphi \in \mathrm{FO}_q[\tau]$ that is preserved under $\sim_{\mathsf{gn}[k]}$ among finite τ-structures is preserved under $\sim^m_{\mathsf{gn}[k]}$ for some sufficiently large m.



Towards this end, we may pass from any two τ-structures with strictly guarded parameter tuples, $\mathfrak{A}, \mathbf{a}$ and $\mathfrak{B}, \mathbf{b}$ to finite companions $\tilde{\mathfrak{A}}_0$ and $\tilde{\mathfrak{B}}_0$ and coverings $\hat{\mathfrak{A}}_0$ and $\hat{\mathfrak{B}}_0$ according to (#) that are (n, K)-free and N-acyclic for any desired values of n, K, N, and such that

$$\begin{array}{lcl} \hat{\mathfrak{A}}_0|_\tau, \mathbf{a} \sim_{\mathsf{gn}[k]} \tilde{\mathfrak{A}}_0|_\tau, \mathbf{a} \sim_{\mathsf{gn}[k]} \mathfrak{A}, \mathbf{a} & \text{and} & \hat{\mathfrak{A}}_0, \mathbf{a} \sim_{\mathsf{g}} \tilde{\mathfrak{A}}_0, \mathbf{a} \sim_{\mathsf{g}} \mathfrak{T}(\mathfrak{A}), \mathbf{a}; \\ \hat{\mathfrak{B}}_0|_\tau, \mathbf{b} \sim_{\mathsf{gn}[k]} \tilde{\mathfrak{B}}_0|_\tau, \mathbf{b} \sim_{\mathsf{gn}[k]} \mathfrak{B}, \mathbf{b} & \text{and} & \hat{\mathfrak{B}}_0, \mathbf{b} \sim_{\mathsf{g}} \tilde{\mathfrak{B}}_0, \mathbf{b} \sim_{\mathsf{g}} \mathfrak{T}(\mathfrak{B}), \mathbf{b}. \end{array}$$

For $\mathfrak{A}, \mathbf{a} \sim^m_{\mathsf{gn}[k]} \mathfrak{B}, \mathbf{b}$, these also imply, by Observation 4.17 , that

$$\mathfrak{T}(\mathfrak{A}), \mathbf{a} \sim^m_{\mathsf{g}} \mathfrak{T}(\mathfrak{B}), \mathbf{b} \quad \text{and therefore also} \quad \mathfrak{T}(\mathfrak{A})^{\omega *}, \mathbf{a} \sim^m_{\mathsf{g}} \mathfrak{T}(\mathfrak{B})^{\omega *}, \mathbf{b}.$$

From (#), also $\mathfrak{T}(\mathfrak{A})^{\omega *}, \mathbf{a} \simeq (\tilde{\mathfrak{A}}_0)^{\omega *}, \mathbf{a} \simeq (\hat{\mathfrak{A}}_0)^{\omega *}, \mathbf{a}$ and similarly for $\mathfrak{B}$. For suitably large m, this implies

$$\mathfrak{T}(\mathfrak{A})^{\omega *}, \mathbf{a} \equiv^q \mathfrak{T}(\mathfrak{B})^{\omega *}, \mathbf{b}$$

because these ω-tree-unfoldings are (n, K)-free and fully acyclic (cf. Observation 3.2). For suitably large choices of n, K, N and m, we are moreover in the situation of Corollary 4.9 and find that

$$\mathfrak{T}(\mathfrak{A})^{\omega *}, \mathbf{a} \simeq (\tilde{\mathfrak{A}}_0)^{\omega *}, \mathbf{a} \equiv^q \hat{\mathfrak{A}}_0, \mathbf{a} \quad \text{and} \quad \mathfrak{T}(\mathfrak{A})^{\omega *}, \mathbf{a} \simeq (\tilde{\mathfrak{B}}_0)^{\omega *}, \mathbf{b} \equiv^q \hat{\mathfrak{B}}_0, \mathbf{b}.$$

It follows that $\hat{\mathfrak{A}}_0, \mathbf{a} \equiv^q \hat{\mathfrak{B}}_0, \mathbf{b}$, so that in particular, for the τ-reducts,

$$\hat{\mathfrak{A}}_0|_\tau, \mathbf{a} \equiv^q \hat{\mathfrak{B}}_0|_\tau, \mathbf{b}.$$

Overall, we have thus found finite companions for finite $\mathfrak{A}$ and $\mathfrak{B}$ such that

$$\mathfrak{A}, \mathbf{a} \sim_{\mathsf{gn}[k]} \hat{\mathfrak{A}}_0|_\tau, \mathbf{a} \equiv^q \hat{\mathfrak{B}}_0|_\tau, \mathbf{b} \sim_{\mathsf{gn}[k]} \mathfrak{B}, \mathbf{b},$$

provided just $\mathfrak{A}, \mathbf{a} \sim^m_{\mathsf{gn}[k]} \mathfrak{B}, \mathbf{b}$.

This shows that φ is preserved under $\sim^m_{\mathsf{gn}[k]}$ among finite τ-structures. □

5 The guarded fragment of Ptime

As remarked above, the weakly N-acyclic covers of Theorem 2.14 can be used to prove the finite model property for GF under constraints involving forbidden homomorphic images [5], and, through that, to prove the finite model property also for GNF in [6]. Several other applications are presented in [5]. One of these is Ptime canonisation w.r.t. guarded bisimulation equivalence. This is particularly interesting in the present context, because it entails a capturing result in descriptive complexity. We briefly present the main ingredients and discuss this application from [5] in this final section. This treatment, too, is a generalisation of a much simpler result from [15] in the world of graph-like structures, to the world of relational structures of any width.

The modal analogue. For a finite transition system $\mathfrak{A} = (A, (E^{\mathfrak{A}}), (P^{\mathfrak{A}}))$ in a fixed finite vocabulary τ (with some binary relations E and some unary predicates P) one may consider bisimulation equivalence as an equivalence relation on A, which we denote as $\sim^{\mathfrak{A}}$. Then there is a natural quotient of $\mathfrak{A}$ w.r.t. $\sim^{\mathfrak{A}}$,

$$\begin{array}{rcl} \mathfrak{A}/_{\sim} & := & \big(A/_{\sim^{\mathfrak{A}}}, (E^{\mathfrak{A}/\sim}), (P^{\mathfrak{A}/\sim})\big), \\ \text{where} \quad E^{\mathfrak{A}/\sim} & = & \{([a],[a'])\colon (a,a') \in E^{\mathfrak{A}}\}, \\ P^{\mathfrak{A}/\sim} & = & P^{\mathfrak{A}}/_{\sim^{\mathfrak{A}}}. \end{array}$$

One checks that $\mathfrak{A}/_{\sim}, [a] \sim \mathfrak{A}, a$. Moreover, the bisimulation quotients $\mathfrak{A}/_{\sim}$ carry a uniformly definable and therefore isomorphism-invariant linear ordering. Enumerating the elements of $\mathfrak{A}/_{\sim}$ w.r.t. to this order, we obtain an ordered representation of the isomorphism type of $\mathfrak{A}/_{\sim}$ over a set $\{1, \dots, n\}$ of the right size. This representation is uniquely determined by the $\sim$-type of $\mathfrak{A}$ (uniquely determined as a structure, not just up to isomorphism (!)). Denote the image structure resulting in this manner as $I_{\sim}(\mathfrak{A})$; and as $I_{\sim}(\mathfrak{A}, a)$ if we include a distinguished element, which is mapped to the element representing its equivalence class $[a]$ in $\mathfrak{A}/_{\sim}$. Then the map $\mathfrak{A} \mapsto I_{\sim}(\mathfrak{A})$ provides *canonisation* (for finite transition systems with a distinguished element) w.r.t. bisimulation equivalence:

- for all finite $\mathfrak{A}, a$: $I_{\sim}(\mathfrak{A}, a) \sim \mathfrak{A}, a$;
- for all finite τ-structures $\mathfrak{A}, a$ and $\mathfrak{A}', a'$:
  $\mathfrak{A}, a \sim \mathfrak{A}', a'$ iff $I_{\sim}(\mathfrak{A}, a) = I_{\sim}(\mathfrak{A}', a')$.

Moreover, this map is polynomial time computable, and hence provides *Ptime canonisation*.

Now a property (of elements in finite τ-transition systems) is closed under bisimulation equivalence if, for all $\mathfrak{A}, a$, it is true of $\mathfrak{A}, a$ iff it is true of $\mathfrak{A}/_{\sim}, [a]$ iff it is true of $I_{\sim}(\mathfrak{A}, a)$. Therefore any polynomial time computable property of elements in finite τ-transition systems that is closed under bisimulation equivalence may alternatively be evaluated after passage to the canonical representative $I_{\sim}(\mathfrak{A}, a)$. Pre-processing with $I_{\sim}$ acts like a filter that enforces closure under bisimulation equivalence: the composition of an arbitrary Ptime decision algorithm with the Ptime algorithm for $I_{\sim}$ will always result in a Ptime decision procedure for some Ptime property that is closed under bisimulation equivalence. The class of Ptime computable properties that are closed under bisimulation, *bisimulation invariant Ptime*, Ptime/$\sim$, therefore admits an effective syntactic representation. It is fully and precisely represented by the compositions of arbitrary polynomially clocked algorithms with a fixed Ptime algorithm for the computation of $I_{\sim}$. Though the argument is simple, the result is remarkable in that it provides effective syntax for the class of Ptime properties that are $\sim$-invariant. The class of all Ptime algorithms that decide such properties is undecidable as a subclass of all Ptime decision algorithms. We note the similarity of this phenomenon with expressive completeness of an effective syntactic fragment (like $\exists$posFO) for a class of FO-properties defined in terms of an undecidable semantic constraint (like preservation under homomorphisms).

In fact we can do even better here as well, and provide an actual logic that is expressively complete for Ptime/$\sim$, viz. a generalised variant of the modal μ-calculus L_μ. See [11] and the original reference [15].

Weakly N-acyclic covers and canonisation. An extension of the above idea from graph-like structures to arbitrary relational structures and guarded bisimulation has to deal with an interesting problem: there is no natural quotient τ-structure $\mathfrak{A}/_{\sim_\mathrm{g}}$ for finite τ-structures $\mathfrak{A}$ w.r.t. guarded bisimulation equivalence. A natural concise quotient representation of the $\sim_\mathrm{g}$-class of $\mathfrak{A}$ can be derived from the game graph of the $\sim_\mathrm{g}$-game over $\mathfrak{A}$. Up to bisimulation, this game graph can be represented as a transition system whose states are the $\sim_\mathrm{g}$-classes of guarded tuples over $\mathfrak{A}$. This abstraction does provide Ptime computable *complete invariants* $I_{\sim_\mathrm{g}}(\mathfrak{A})$ or $I_{\sim_\mathrm{g}}(\mathfrak{A}, \mathbf{a})$ w.r.t. guarded bisimulation equivalence:

– for all finite τ-structures $\mathfrak{A}, \mathbf{a}$ and $\mathfrak{A}', \mathbf{a}'$:
  $\mathfrak{A}, \mathbf{a} \sim \mathfrak{A}', \mathbf{a}'$ iff $I_{\sim_\mathrm{g}}(\mathfrak{A}, \mathbf{a}) = I_{\sim_\mathrm{g}}(\mathfrak{A}', \mathbf{a}')$.

But the images under $I_{\sim_\mathrm{g}}$ are not τ-structures, and therefore cannot provide canonisation. One can try to read $I_{\sim_\mathrm{g}}(\mathfrak{A})$ as a transition system that specifies the local overlap structure between guarded tuples and their quantifier-free types, similar to the use made of the intersection graph of $\mathrm{H}(\mathfrak{A})$ towards the construction of guarded tree unfoldings. The infinite tree unfolding of $I_{\sim_\mathrm{g}}(\mathfrak{A})$ can be used in this manner, and its ω-tree-unfolding would lead to $\mathfrak{A}^{\omega *}$. However, $I_{\sim_\mathrm{g}}(\mathfrak{A})$ itself cannot serve as the building plan for a finite canonical representative of $\mathfrak{A}$ modulo $\sim_\mathrm{g}$ in this naive manner. The following trivial example may illustrate this point.

Example 5.1. Let $\mathfrak{A}$ be a directed 3-cycle w.r.t. the binary relation E. Then $\mathfrak{A}$ realises exactly two quantifier-free types of guarded tuples, the type of a directed E-edge, and the type of a singleton vertex (without reflexive E-loop). Disregarding the 1-type, the overlap pattern w.r.t. non-degenerate guarded pairs specified in $I_{\sim_\mathrm{g}}(\mathfrak{A})$ would just tell us that every pair (a, a') of this type must have an overlap in a with some pair (a'', a) of the same type and in a' with some pair (a', a''') of the same type. In $I_{\sim_\mathrm{g}}(\mathfrak{A})$, we would just find correspondingly labelled edges and loops at the nodes representing the two mirror-symmetric types of non-degenerate guarded pairs in $\mathfrak{A}$. But $\mathfrak{A}$ is the minimal realisation, because a structure with fewer than three realisations of these types cannot consistently realise the required types and overlaps.

This trivial example may wrongly suggest that some locally acyclic graph cover of $I_{\sim_\mathrm{g}}(\mathfrak{A})$ could avoid the difficulty. However, the situation is more complicated for relational structures in vocabularies of width greater than 2, however, because sub-configurations of the tuples represented in individual elements of $I_{\sim_\mathrm{g}}(\mathfrak{A})$ may be carried through sequences of overlaps of any length. Again, dangerous cycles cannot a priori be bounded in length. The N-acyclic covers of Theorem 2.12 deal with this difficulty but are not polynomial. Instead, the construction of weakly N-acyclic covers of [5], cf. Theorem 2.14, can be used to transform $I_{\sim_g}(\mathfrak{A})$ into a finite $\hat{I}(\mathfrak{A}) \sim I_{\sim_g}(\mathfrak{A})$ that is a valid building plan for a finite τ-structure that is guarded bisimilar to $\mathfrak{A}$, and can be computed in polynomial time from $I_{\sim_g}(\mathfrak{A})$ and hence from $\mathfrak{A}$. The resulting representative of the $\sim_g$-class of $\mathfrak{A}$ is canonical since it is constructed from just $I_{\sim_g}(\mathfrak{A})$. So we obtain the following in [5].

Theorem 5.2. *There is a polynomial time computable canonisation for finite τ-structures (with guarded tuples of parameters) w.r.t. guarded bisimulation.*

Following the reasoning that was reviewed for bisimulation of graph-like structures above, this directly leads to an abstract capturing result.

Corollary 5.3. *The class of Ptime computable properties (of guarded tuples in τ-structures) that are closed under guarded bisimulation,* Ptime/$\sim_g[\tau]$*, admits an effective syntactic representation.*

In other words, the *guarded fragment of Ptime*, Ptime/$\sim_g$, can be captured in the sense of descriptive complexity, just like the modal fragment of Ptime Ptime/$\sim$. Concrete logical syntax that is expressively complete for this class Ptime/$\sim_g[\tau]$ could also be given. These results are of some systematic interest, because the question whether Ptime itself (which really is Ptime/$\simeq$) can be captured remains one of the great open problems of descriptive complexity.

Acknowledgement. I am grateful for discussions with Vince Bárány and Balder ten Cate concerning the guarded negation fragment.

References

[1] N. Alon. Tools from higher algebra. In R. Graham, M. Grötschel, and Lovász, editors, *Handbook of Combinatorics*, volume II, pages 1749–1783. North-Holland, 1995.

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

[3] A. Atserias, A. Dawar, and M. Grohe. Preservation under extensions on well-behaved finite structures. In *Proceedings of 32nd International Colloquium on Automata, Languages and Programming ICALP'05*, volume 3580 of *LNCS*, pages 1437–1449, 2005.

[4] A. Atserias, A. Dawar, and P. Kolaitis. On preservation under homomorphisms and unions of conjunctive queries. *Journal of the ACM*, 53:208–237, 2006.

[5] V. Bárány, G. Gottlob, and M. Otto. Querying the guarded fragment. In *Proceedings of the 25th IEEE Symposium on Logic in Computer Science LICS'2010*, pages 2–11, 2010.

[6] V. Bárány, B. ten Cate, and L. Segoufin. Guarded negation. In *Proceedings of ICALP*, pages 356–367, 2011.

[7] C. Beeri, R. Fagin, D. Maier, and M. Yannakakis. On the desirability of acyclic database schemes. *Journal of the ACM*, 30:497–513, 1983.

[8] C. Berge. *Graphs and Hypergraphs.* North-Holland, 1973.

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

[10] A. Dawar and M. Otto. Modal characterisation theorems over special classes of frames. *Annals of Pure and Applied Logic*, 161:1–42, 2009.

[11] E. Grädel. Finite model theory and descriptive complexity. In *Finite Model Theory and Its Applications*, pages 125–230. Springer, 2007.

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

[13] I. Hodkinson and M. Otto. Finite conformal hypergraph covers and Gaifman cliques in finite structures. *Bulletin of Symbolic Logic*, 9:387–405, 2003.

[14] D. Janin and I. Walukiewicz. On the expressive completeness of the propositional μ-calculus with respect to monadic second-order logic. In *Proceedings of CONCUR*, pages 263–277, 1996.

[15] M. Otto. Bisimulation-invariant Ptime and higher-dimensional μ-calculus. *Theoretical Computer Science*, 224:237–265, 1999.

[16] M. Otto. Modal and guarded characterisation theorems over finite transition systems. *Annals of Pure and Applied Logic*, 130:173–205, 2004.

[17] M. Otto. Bisimulation invariance and finite models. In *Colloquium Logicum 2002*, Lecture Notes in Logic, pages 276–298. ASL, 2006.

[18] M. Otto. Highly acyclic groups, hypergraph covers and the guarded fragment. *Journal of the ACM*, 59, 2012.

[19] E. Rosen. Modal logic over finite structures. *Journal of Logic, Language and Information*, 6:427–439, 1997.

[20] B. Rossman. Homomorphism preservation theorems. *Journal of the ACM*, 55, 2008.

[21] B. ten Cate and L. Segoufin. Unary negation. In *Proceedings of STACS*, pages 344–355, 2011.