HAL open science

The monadic theory of finite representations of infinite words

Anuj Dawar, David Janin

To cite this version:

Anuj Dawar, David Janin. The monadic theory of finite representations of infinite words. Information Processing Letters, 2007, 103, pp.94-101. hal-00306381

HAL Id: hal-00306381

https://hal.science/hal-00306381v1

Submitted on 3 Nov 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The Monadic Theory of Finite Representations of Infinite Words

Anuj Dawar* David Janin†

January 16, 2007

Abstract

In this paper, we consider existential monadic second-order logic (monadic Σ_1) on finite unary graphs, that is finite graphs with functional edge relation. These can be seen as finite encodings of ultimately periodic words. We show that, on these graphs, the bisimulation invariant fragment of monadic Σ_1 equals the bisimulation invariant fragment of monadic second order logic itself, and that MSO-definable bisimulation closed classes of graphs coincide with classes of graphs definable by means of (an extension of) finite state ω-word automata. This result can be seen as a translation, onto finite representations of infinite words, of Büchi's automata-theoretic characterization of S1S. In terms of descriptive complexity, this result contrasts with the situation on arbitrary unary structures where bisimulation invariant monadic Σ_1 properties only define languages that are closed with respect to the prefix topology.
Keywords: Automata on infinite words, descriptive complexity, logic, specification languages.

1 Introduction

Models and properties of infinite words

Given a finite alphabet A, an infinite word (or ω-word) over A is a sequence $(a_i : i \in \omega)$ where each $a_i \in A$. When treating such words as structures that interpret formulas of some logic, it is common to take the underlying domain to be the set of natural numbers, and to allow a unary predicate symbol P_a for each element $a \in A$ so that P_a holds of integer i if, and only if, the ith element in the word is a. In this way a sentence in a predicate logic, such as first-order or second-order logic, can be seen as defining a language of infinite words, namely the set of words whose encoding satisfies the sentence. It is with such an encoding that Büchi, seeking a decision procedure for S1S, the monadic theory of the natural numbers, obtained his famous automata-theoretic characterisation of languages (i.e. sets of words) definable in monadic second-order logic [3]. The consequences of this result continue to have a deep impact on research in computer science.

We are especially interested in those infinite words that are eventually periodic. That is, they are of the form $u.v^\omega$ for finite strings u and v. We can represent such words as *finite structures* consisting of a (directed) graph that contains a simple path labeled u leading to a cycle labelled with the word v. We examine properties of this class of structures, which we call *lassos*, that are definable in monadic second-order logic. It is clear that any property of eventually periodic words can be reformulated as a property of lassos. We aim to study the definability and descriptive complexity of languages of eventually periodic words when they are represented in the form of finite graphs, especially as lassos.

Properties of words vs properties of their finite encodings

Given an infinite word w that is eventually periodic, the u and v such that $w = u.v^\omega$ are not uniquely determined. Thus, if we have a sentence φ of some logic that is interpreted over lassos and we wish to

*University of Cambridge Computer Laboratory, Cambridge CB3 0FD, UK, anuj.dawar@cl.cam.ac.uk.
†LaBRI, Université Bordeaux I, 33405 Talence, France, janin@labri.fr.

interpret the models of φ as defining a language of infinite words, it is important to consider those sentences that do not distinguish between different lassos that encode the same word.

This invariance condition can be seen as a special case of *bisimulation invariance* which has been much studied in the context of modal logics [2]. It has also been studied specifically in the context of monadic second-order logic in [7], where it was shown that any property that is definable in monadic second-order logic and is bisimulation-invariant is definable in the modal μ-calculus.

In the particular context of finite encodings of eventually periodic words, an interesting question poses itself. What is the cost or benefit, from the point of view of definability in a given logical formalism, of representing words in terms of their finite encodings instead of explicitly as infinite structures? Of course, finite representability is crucial when we want to study computational aspects of the definable properties. However, it is also the case that the particular representation, such as lassos, makes certain information explicit that is only implicitly available in the infinite word. For instance, any lasso contains a special vertex (which we will call the *knot*) that marks the beginning of the cycle and is easily distinguished from other vertices. As we shall see, this may allow us to give simpler definitions of some languages of infinite, eventually periodic words. Indeed, it is possible that certain languages may become definable in their finite representations that are not definable at all on infinite words.

Our contribution

In this paper we study the power of monadic second-order logic (MSO) on finite representations of eventually periodic words and give a fairly complete account of it.

First, we establish that, as might be expected, for any MSO definable language of ω-words L there is an MSO formula φ_L that defines the set of all finite encodings of the eventually periodic words of L.

Moreover, we prove that such a formula φ_L can always be found in the Σ_1 fragment of the quantifier alternation hierarchy of monadic second order logic. This contrasts with the case where MSO formulas are interpreted directly over infinite words as we know [6] that in this case the monadic Σ_1 formulas only define closed languages in the prefix topology. As these correspond to Büchi automata where any infinite run is accepting, they represent only a fragment of the languages that are definable in MSO.

An intuitive explanation for this contrast is that on finite lassos it is possible to state infinitary Büchi automata acceptance conditions using a monadic Σ_1 formula by requiring that the knot vertex is labelled by an accepting state. On infinite words, in the absence of such a knot vertex, the condition can be stated but at the cost of an increase in the descriptive complexity.

Nonetheless, we are able to show that, apart from this increase in complexity, the expressive power of MSO is the same in the two cases. To be precise, we show that any class of lassos that is MSO-definable and that describes word properties (i.e. that is invariant under the equivalence on lassos relating structures encoding the same infinite word) is already MSO definable as a class of infinite words. This means that MSO definability on lassos can still be captured by finite automata.

These two results can be seen as a translation, onto finite representations of infinite words, of the Büchi automata characterization theorem [3] of S1S.

These results also shed some light on the question of characterising the bisimulation invariant fragment of MSO in the finite. As was shown by Janin and Walukiewicz [7], any MSO formula that is invariant under bisimulations on all structures is equivalent to a formula of the modal μ-calculus (L_μ). It has long been an open question whether this is true in the finite. That is, if an MSO formula φ is invariant under bisimulation on finite structures, is it the case that φ is equivalent *on finite structures* to a formula of L_μ. Similarly, the bisimulation invariant fragment of MSO with counting is known to correspond to the counting extension of L_μ. More fine-grained results have established that the Σ_1 fragment of MSO, when restricted to bisimulation-invariant properties corresponds exactly to the fragment of L_μ which only uses greatest fixed points and similarly for the counting extensions [6].

It is these last results where we show that the situation on finite structures is drastically different. The results on eventually periodic words show that the correspondence between bisimulation-invariant Σ_1 with counting and the greatest fixed-point fragment of counting L_μ simply fails in the finite. We can construct a formula of Σ_1 that is invariant under counting bisimulations on finite structures but is not equivalent to

a formula of L_μ with counting that uses only greatest fixed points. Similarly, the correspondence between Σ_1 and L_μ with only greatest fixed points, even without counting, fails, at least in relation to finite unary graphs.

Related work

The interest in bisimulation preservation theorems goes back to van Benthem [13], who showed that a formula of first-order logic is preserved under bisimulations on all structures if, and only if, it is equivalent to a formula of modal logic. While many preservation theorems of classical model theory fail when we restrict ourselves to finite structures, this is one exception. It was shown by Rosen [12] that a formula of first-order logic is preserved under bisimulations *on finite structures* if, and only if, it is equivalent *on finite structures* to a formula of modal logic. This result has prompted further investigation of preservation under bisimulations on finite structures. For instance, Otto in[10] gives a unified treatment of the van Benthem and Rosen results and considers their extensions to other variants of bisimulation and in [9] studies similar preservation theorems for guarded logics. The status of the first-order modal preservation theorem on specific classes of finite structures is explored by Dawar and Otto in [5]. The modal preservation theorems for monadic second-order logic of Janin and Walukiewicz and Janin and Lenzi have also received attention and it has been an active open question whether the corresponding results hold in the finite. In our paper [4], we examined the case of the latter property on specific classes of finite structures. Some of the results in this paper were presented in preliminary form there.

2 Background, definitions and main results

2.1 Graphs

The logic we consider is interpreted over rooted labeled directed graphs (often called transition systems or Kripke structures). Fix a set *Prop* of atomic propositions. A graph for *Prop* is a structure

$$M = \langle V, r, E, \{p^M\}_{p \in Prop} \rangle$$

with universe V (vertices), distinguished element $r \in V$ (root), binary relation $E \subseteq V \times V$ (directed edges) and monadic relations $p^M \subseteq V$ for each atomic proposition $p \in Prop$ (labeling).

A graph M is a *unary graph* when its edge relation E is functional, i.e. for any vertex u in M there is one and only one vertex v in M such that $(u, v) \in E$.

A (directed) *path* in graph M is a (finite or infinite) sequence of vertices such that for any two consecutive vertices v_1 and v_2 in the sequence one has $(v_1, v_2) \in E$.

When M is unary, there is a unique infinite path $\rho \in V^\omega$ starting from the root. In this case, given $\Sigma = \mathcal{P}(Prop)$, we define the *infinite word encoded by the unary graph* M to be the word $w_M \in \Sigma^\omega$ defined, for each $i \in \mathbb{N}$, by $w_M[i] = \{p \in Prop : \rho[i] \in p^M\}$.

2.2 Bisimulation

A *bisimulation* between two graphs

$$M = \langle V, r, E, \{p^M\}_{p \in Prop} \rangle$$

and

$$M' = \langle V', r', E', \{p^{M'}\}_{p \in Prop} \rangle$$

is a relation $B \subseteq V \times V'$ such that, if $(v, v') \in B$ then:

- for each $p \in Prop$, $v \in p^M \Leftrightarrow v' \in p^{M'}$;
- for each w with $(v, w) \in E$ there is a w' with $(v', w') \in E'$ and $(w, w') \in B$; and

- for each w' with $(v', w') \in E'$ there is a w with $(v, w) \in E$ and $(w, w') \in B$.

A *counting bisimulation* between M and M' is a relation $B \subseteq V \times V'$ such that, if $(v, v') \in B$ then:

- for each $p \in Prop$, $v \in p^M \Leftrightarrow v' \in p^{M'}$;
- B contains a bijection between the sets $\{w \mid (v, w) \in E\}$ and $\{w' \mid (v', w') \in E'\}$.

Observe that any counting bisimulation is a bisimulation.

We say that M and M' are (counting) bisimilar if there is a (counting) bisimulation B between them with $(r, r') \in B$. More generally, we say that two states $v \in M$ and $v' \in M'$ (where M and M' are not necessarily distinct) are (counting) bisimilar if there is a (counting) bisimulation B between the two structures with $(v, v') \in B$.

Restricted to unary graphs, the two notions coincide. That is, unary graphs M and M' are bisimilar if, and only if, they are counting bisimilar.

2.3 Logic

We consider standard predicate logics, in particular first-order logic (FO) and monadic second-order logic (MSO). These are interpreted in graphs. In the sequel, we shall write $\varphi(x_1, \cdots, x_n)$ or simply $\varphi(\overline{x})$ for an FO or MSO-formula with free first-order variables among $\overline{x} = (x_1, \cdots, x_n)$ *regardless* of the free monadic predicate (or set) variables occurring in φ. More precisely, given the set $\{X_1, \cdots, X_n\}$ of all set variables occurring free in φ, we shall implicitly and whenever required interpret the formula φ on transition systems with the set of atomic proposition $Prop' = Prop \cup \{X_1, \cdots, X_n\}$.

A class of graphs C is (counting) bisimulation closed when, for any graph M, $M \in C$ whenever there is $N \in C$ such that M and N are (counting) bisimilar. We say that an FO or MSO-sentence φ is *(counting) bisimulation invariant* when the class C_φ of models of φ is (counting) bisimulation closed, i.e. for any two (counting) bisimilar models M and N one has $M \models \varphi$ if, and only if, $N \models \varphi$.

We say that a class of *graphs* is (counting) bisimulation closed *in the finite* when the above property holds restricted to finite graphs only. By extension, we say that an MSO-sentence φ is *(counting) bisimulation invariant in the finite* when the class C^f_φ of *finite* graphs it defines is (counting) bisimulation closed in the finite.

Observe that any unary graph M is characterized up to bisimulation by the infinite word $w_M \in \Sigma^\omega$ that is described by the infinite path emanating from the root of M. Also, as we observed earlier, on unary graphs, bisimulation coincides with counting bisimulation. In particular, this means that if a sentence φ is invariant under counting bisimulations in the finite then, in particular, on finite unary graphs it is invariant under bisimulations.

It follows that a counting bisimulation invariant class C of unary graphs is characterized by the language of words $L_C = \{w_M \in \Sigma^\omega : M \in C\}$, i.e. for any unary graph M, $M \in C$ if and only if $w_M \in L_C$. This, along with Büchi's characterization of S1S by means of finite state automata [3] gives us the following theorem (following arguments similar to those developed in [7] but restricted to unary graphs).

Theorem 1 (Büchi [3]) *A bisimulation closed class of (finite or infinite) unary graphs C is MSO-definable if, and only if, the language L_C is ω-regular.*

2.4 Main results

In this paper, we focus our attention on bisimulation invariance on *finite* unary graphs. As above, a bisimulation invariant class of a finite unary graph C is still characterized by the language L_C it induces, and counting bisimulation coincides with bisimulation. However, in this case, since all graphs in C are finite, all words in L_C are ultimately periodic, i.e. of the form $u.v^\omega$ for some finite non empty words u and $v \in \Sigma^+$.

These leads to the following definition:

Definition 2 *Given a language of infinite words $L \subseteq \Sigma^\omega$, the* kernel *$kern(L) \subseteq \Sigma^\omega$ of L is defined to be the set of all ultimately periodic words of L, i.e. $kern(L) = \{u.v^\omega : u, v \in \Sigma^+, u.v^\omega \in L\}$.*

Our main result is:

Theorem 3 *A bisimulation closed class of finite unary graphs C is definable by a MSO-sentence φ if, and only if, there is an ω-regular language $L \subseteq \Sigma^\omega$ such that $L_C = kern(L)$.*

Moreover, φ is equivalent, over C to a formula in the Σ_1 level of the monadic second order logic hierarchy.

Thus, we get the following in terms of descriptive complexity.

Corollary 4 *Restricted to bisimulation invariant properties on finite unary graphs, the monadic quantifier alternation depth hierarchy collapses to the Σ_1 level.*

3 From languages of words to classes of unary graphs

We first show that for any regular language of infinite words, there is an existential monadic formula that defines, in the finite, the set of all graph encodings of the ultimately periodic words of this language.

More precisely:

Theorem 5 *For any regular ω-language $L \subseteq \Sigma^\omega$ there is a* monadic *Σ_1 formula φ_L such that for any finite unary graph M, $M \models \varphi_L$ if, and only if, $w_M \in L$.*

Proof. Let L be an ω-regular language.

Observe first, that there is a nondeterministic finite Büchi automaton

$$\mathcal{A}_L = \langle Q, q_0, \delta, F \rangle$$

that recognizes L and such that, for any infinite word of L of the form $u.v^\omega$, there is an accepting state $q \in F$ such that, there is a path in $\mathcal{A}_L$ from q_0 to q reading u, and a cycle in $\mathcal{A}_L$ from q to q reading v. In fact, following [11], such an automaton can be taken as the Büchi automaton one can canonically build out of an ω-semigroup recognizing L.

Now, the formula φ_L can be defined as follows: there is a collection of disjoint sets X_q, one for each state $q \in Q$, such that: (i) $r \in X_{q_0}$; (ii) for each q, for each vertex $x \in X_q$, x has a single successor y and there is a state $q' \in \delta(\lambda(x))$ such that $y \in X_{q'}$ with $\lambda(x) = \{p \in Prop : p(x) \text{ holds}\}$; and (iii) any element with two predecessors in the union of the X_qs belongs to some X_q with $q \in F$.

The formula φ_L defined in this way (1) checks that there is a unique path from the root, (2) necessarily defines on this path a labeling that is a run of automaton $\mathcal{A}_L$, and (3) confirms that there is a vertex on this path with two predecessors labeled by an accepting state. Since M is finite, this suffices to ensure that this run is accepting. □

One may ask whether a converse of this theorem holds. More precisely, given an MSO-formula φ, let L_φ be the language of all infinite words encoded by the unary finite models of φ, i.e.

$$L_\varphi = \{w_M \in \Sigma^\omega : M \models \varphi, M \text{ unary and finite}\}$$

One may ask, for instance, if there is some ω-regular language $L \subseteq \Sigma^\omega$ such that $L_\varphi = kern(L)$. Recall that, by construction, all words in L_φ are ultimately periodic.

The answer is no: there are MSO sentences φ such that L_φ is not the kernel of an ω-regular language. In fact, on the alphabet $\Sigma = \{a, b\}$, take the formula φ that defines unary graphs such that, on the unique path from the root, there is a single b on the cycle. One has $L_\varphi = \{a^m.(b.a^n)^\omega \in \Sigma^\omega : m \in \mathbb{N}, n \in \mathbb{N}\}$ but, by the pumping lemma, for any ω-regular language L such that $L_\varphi \subseteq L$, there are integers m, n and $p > n$ such that $a^m.(b.a^n.b.a^p)^\omega \in L$ hence $L_\varphi \neq kern(L)$.

However, we prove in the next two sections that a converse to Theorem 5 can be obtained under the assumption that the formula φ is invariant under counting bisimulations in the finite.

4 The monadic second order theory of lassos

In this section, we consider monadic second order logic on lassos, i.e. unary graphs that are canonical encodings of ultimately periodic words.

More formally, a unary graph M is called a *lasso* if all nodes in M except two have exactly one predecessor. The two exceptions are the root, which has no predecessors and a node we call the *knot* which has two.

Any finite unary graph is bisimilar to a lasso. Moreover, any lasso M is completely characterized by the two non empty finite words u and v, in the alphabet Σ, that are described respectively by the simple path leading from the root to the knot of M (excluding the knot) and by the simple cycle leading from the knot back to itself (excluding the knot at the end of the path). We write $M_{u,v}$ for such a lasso. The following characterization of MSO on lassos follows from the Decomposition Theorem proved for MSO in [8].

Lemma 6 *For any MSO sentence φ, there exists a finite set of pairs of regular languages $\{(U_i, V_i) \subseteq \Sigma^+ \times \Sigma^+\}_{i\in I}$ such that:*

$$M_{u,v} \models \varphi \text{ if, and only if, } (u,v) \in \bigcup_{i\in I} U_i \times V_i \quad ((u,v) \in \Sigma^+ \times \Sigma^+)$$

Proof. The mapping that maps any pair of non empty finite words $(u,v) \in \Sigma^+ \times \Sigma^+$ to the lasso $M_{u,v}$ is an FO-definable transduction. It follows, by the decomposition theorem [8] that there is a finite set of pairs of MSO-formulas $\{(\varphi_i, \psi_i)\}_{i\in I}$ over finite Σ-words such that for any two words u and $v \in \Sigma^+$,

$$M_{u,v} \models \varphi$$

if, and only if, there is some $i \in I$ such that $u \models \varphi_i$ and $v \models \psi_i$. By Büchi's theorem, for all $i \in I$, the MSO-formulas φ_i and ψ_i define regular languages U_i and V_i. □

Remark: It follows, in particular, that $L_\varphi \subseteq kern(\bigcup_{i\in I} U_i.(V_i)^\omega)$. However, the inclusion may be proper. To see this, conisder the formula φ in the two letter alphabet $\Sigma = \{a, b\}$ that asserts that on the cycles of lasso either only as or only bs occur. We have $L_\varphi = (a+b)^*(a^\omega + b^\omega)$. But φ is also characterized by the languages $U = (a+b)^+$ and $V = a^+ + b^+$ with $U.V^\omega = (a+b)^\omega$.

At a more conceptual level, the inclusion is not proper because, for a given pair of languages U and V, there may be eventually periodic words in $U.V^\omega$ which are not of the form $u.(v)^\omega$ for any $u \in U$ and $v \in V$. The following lemma gives a precise characterisation of the eventually periodic words in $U.V^\omega$

Lemma 7 *Let U and V be two languages of finite non empty words and let $w \in U.V^\omega$. Word w is ultimately periodic if and only if there is $u_0 \in U$ and $v_1, \ldots, v_m, v_{m+1}, \ldots, v_{m+n} \in V$ such that*

$$w = u_0.v_1.\cdots.v_m.(v_{m+1}.\cdots.v_{m+n})^\omega$$

i.e. $w = u.v^\omega$ with $u \in U.V^+$ and $v \in V^+$.

Thus, in regard to Lemma 6, given some $i \in I$, $u \in U_i.V_i^+$ and $v \in V_i^+$, nothing ensures that $M_{u,v} \models \varphi$ (equivalently $u.v^\omega \in L_\varphi$).

5 Bisimulation invariance on finite unary graphs

We stated, at the end of Section 3 that a converse to Theorem 5 is possible for formulas φ closed under counting bisimulation. So far, we have not used the assumption of invariance under counting bisimulation. We now do this, in the proof of the following theorem.

Theorem 8 *For any MSO formula φ that is counting-bisimulation invariant on finite graphs, there exists a finite set of pairs of languages $(D_t, E_t) \subseteq \Sigma^+ \times \Sigma^+$ such that, for any lasso $M_{u,v}$:*

$$M_{u,v} \models \varphi \text{ iff } \exists r \in D_t.E_t^+, \exists s \in E_t^+ \text{ such that } u.v^\omega = r.s^\omega$$

Proof. Let φ be a formula as above and let $(U_i, V_i)_{i \in I}$ be the regular languages obtained by applying Lemma 6.

.

Lemma 9 *For any $i \in I$, any $(u, v) \in U_i \times V_i$, there is a triple $t = (j, r, s) \in I \times \Sigma^+ \times \Sigma^+$ such that:*

1. *$r.s^\omega = u.v^\omega$ (hence $M_{u,v}$ and $M_{r,s}$ are bisimilar),*
2. *for all $n > 0$, $r.s^n \in U_j$ and $s^n \in V_j$.*

Proof. Let i, u and v be as above, so $M_{u,v} \models \varphi$. By invariance of φ, for each $k > 0$, we also have $M_{u.v^k, v^k} \models \varphi$. Hence, by Lemma 6 for each $k > 0$ there is some $i_k \in I$ such that $(u.v^k, v^k) \in U_{i_k} \times V_{i_k}$. Since I is finite, there is some $j \in I$ such that $j = i_k$ for infinitely many k. Now, since both U_j and V_j are regular languages and there are infinitely many k such that $u.v^k \in U_j$ and $v^k \in V_j$ there must be some $p > 0$ such that $u.v^{pn} \in U_j$ and $v^{pn} \in V_j$ for all $n > 0$. Taking $r = u.v^p$ and $s = v^p$ gives us the desired triple t. □

A triple $t = (j, r, s)$ as in Proposition 9 is called special. Write $\mathcal{S}$ for the set of all special triples.

To continue the proof of Theorem 8, we need some standard definitions from formal language theory. Recall that the *congruence class* $[w]_L$ of a finite word $w \in \Sigma^+$ with respect to a language $L \subseteq \Sigma^+$ is defined as the sets of words

$$[w]_L = \{w' \in \Sigma^+ | \forall u, v \in \Sigma^*, u.w.v \in L \Leftrightarrow u.w'.v \in L\}$$

We know that if L is regular there are only finitely many distinct sets $[w]_L$ for $w \in \Sigma^*$ and each one is a regular language.

For any special triple $t = (j, r, s)$ we define the languages

$$D_t = [r]_{U_j}.([s]_{U_j} \cap [s]_{V_j})$$

and

$$E_t = ([s]_{U_j} \cap [s]_{V_j})$$

By construction, since U_j and V_j are regular languages, D_t and E_t are also regular languages. Also, even though there are infinitely many special triples, there are still only finitely many distinct sets D_t and E_t. Moreover, we have the following lemma.

Lemma 10 *For any special triple $t = (j, r, s)$, $r.s \in D_t$, $s \in E_t$, $D_t.E_t^+ \subseteq U_j$ and $E_t^+ \subseteq V_j$.*

Proof. Let $t = (j, r, s)$ be a special triple as above.

Recall first that, for any u and $v \in \Sigma^+$, any $L \subseteq \Sigma^+$, we have $[u]_L.[v]_L \subseteq [u.v]_L$. It follows that:

$$D_t.E_t^+ \subseteq \bigcup_{n>0} [r.s^n]_{U_j}$$

and

$$E_t^+ \subseteq \bigcup_{n>0} [s^n]_{V_j}$$

Moreover, we know that for any $u \in \Sigma^+$ and $L \subseteq \Sigma^+$, if $u \in L$ then $[u]_L \subseteq L$. So we conclude the proof of the lemma by observing that, following Lemma 9, we *do* have, for any $n > 0$, $r.s^n \in U_j$ and $s^n \in V_j$. □

Lemmas 6, 9 and 10 conclude the proof of Theorem 8 □

Corollary 11 *For any MSO formula φ that is counting-bisimulation invariant on finite graphs, there exists a regular language L such that $L_\varphi = kern(L)$.*

Proof. Take $L = \bigcup_{t \in T} D_t.(E_t)^\omega$ where the languages D_t and E_t are as given by Theorem 8.

We first show that $L_\varphi \subseteq kern(L)$. Let w be in L_φ. By definition, there are u and v such that $u.v^\omega = w$ and $M_{u,v} \models \varphi$. By Lemma 9, this means that there is a special triple $t = (j, r, s)$ such that $u.v^\omega = r.s^\omega$ with $r.s \in D_t$ and $s \in E_t$ hence $w = r.s^\omega \in D_t.E_t^\omega$.

For the converse, let w be an ultimately periodic word in L, i.e. $w \in kern(L)$. By definition of L, this means that there is a special triple $t = (j, r, s)$ such that $w \in D_t.(E_t)^\omega$. By Lemma 7, this means that $w = u.v^\omega$ with $u \in D_t.E_t^+$ and $v \in E_t^+$. By Lemma 10, this means $w = u.v^\omega$ with $u \in U_j$ and $v \in V_j$, hence $M_{u,v} \models \varphi$ and thus $w \in L_\varphi$. □

6 Application to bisimulation invariance in the finite

As a corollary, putting together the two theorem above:

Corollary 12 *Any MSO formula that is bisimulation invariant on finite unary graphs is equivalent to a* monadic Σ_1 *formula.*

Moreover, since any class of unary graphs that is equivalent to a regular language is definable, among finite unary graphs, in the (counting or modal) mu-calculus, we obtain the following corollary.

Corollary 13 *The bisimulation invariant fragment of* monadic Σ_1 *on finite unary graphs is equivalent, on finite unary graphs, to the mu-calculus* L_μ*.*

Writing N_1 for the first alternation level of L_μ (i.e. formulas of L_μ with only greatest fixed-point operators), we see that monadic Σ_1 is not only not equivalent to N_1, but it is equivalent to all of L_μ. This result is rather unexpected since, on arbitrary (finite or infinite) unary graphs, the counting bisimulation invariant fragment of monadic Σ_1 only induces topologically closed regular languages [6], i.e. languages definable by finite Büchi automata with only accepting states, which is the same as N_1.

More generally, it is also known that any formula of monadic Σ_1 that is invariant under *counting bisimulations* on all graphs is equivalent to a formula in the counting extension of N_1. Our results show that this simply fails when we restrict ourselves to finite graphs. This is another striking illustration that finite model theory can be dramatically different to the infinite variety.

Observe however that there is no hope to extend Corollary 13 to arbitrary finite graphs.

In fact, the mu-calculus formula

$$\mu X.(p \vee (\neg p \rightarrow \Diamond X))$$

that defines the set of vertices from which there is a (directed) path to a vertex where p holds, is bisimulation invariant but not definable by a monadic Σ_1 formula. If it was, this would imply that directed reachability would also be definable in monadic Σ_1 and that is not the case [1].

7 Conclusions

The question of whether bisimulation-invariant MSO is equivalent on finite structures to L_μ remains a challenging open problem. By investigating this question at the first level of the alternating hierarchy on unary graphs we have been able to show that the nature of the problem is radically different to its counterpart on arbitrary (finite or infinite) structures. In particular, we have constructed counterexamples to what might be considered the natural ways of adapting the automata construction that works on infinite structures.

References

[1] M. Ajtai and R. Fagin. Reachability is harder for directed rather than undirected finite graphs. *Journal of Symbolic Logic*, 55:113–150, 1990.

[2] J.F.A.K. van Benthem. *Modal Correspondance Theory*. PhD thesis, University of Amsterdam, 1976.

[3] J. R. Büchi. On a decision method in restricted second order arithmetic. In *Proceedings of the international congress on Logic, Methodology and Philosophy of Science 1960*, pages 1–11. Stanford University Press, 1962.

[4] A. Dawar and D. Janin. On the bisimulation invariant fragment of monadic σ_1 in the finite. In *Proc. 24th Conf. on Foundations of Software Technology and Theoretical Computer Science*, volume 3328 of *LNCS*, pages 224–236. Springer, 2004.

[5] A. Dawar and M. Otto. Modal characterisation theorems over special classes of frames. In *Proc. of 20th IEEE Symp. on Logic in Computer Science*, pages 21–30, 2005.

[6] D. Janin and G. Lenzi. On the logical definability of topologically closed recognizable languages of infinite trees. *Computing and Informatics*, 21:185–203, 2002.

[7] D. Janin and I. Walukiewicz. On the expressive completeness of the modal mu-calculus with respect to monadic second order logic. In *Conf. on Concurrency Theory (CONCUR'96)*, pages 263–277. LNCS 1119, 1996.

[8] J.A. Makowski and E. Ravve. Incremental model checking for decomposable strutures. In J. Wiedermann and P. Hajek, editors, *Mathematical Foundation of Comp.Sci (MFCS'95)*, LNCS 969, pages 540–551, 1995.

[9] M. Otto. Modal and guarded characterisation theorems over finite transition systems. In *Proc. of the 17th IEEE Symp. on Logic in Computer Science (LICS)*, pages 371–380, 2002.

[10] M. Otto. Bisimulation invariance and finite models. In W. Pohlers Z. Chatzidakis, P. Koepke, editor, *Logic Colloquium '02*, volume 27 of *Lecture Notes in Logic*. ASL, 2006.

[11] D. Perrin and J.E. Pin. Semigroups and automata on infinite words. In J. Fountain, editor, *Semigroups, Formal Languages and Groups*, NATO Advanced Study Institute, pages 49–72. Kluwer academic, 1995.

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

[13] J. van Benthem. *Modal Logic and Classical Logic*. Bibliopolis, 1983.