RWTH Aachen

Department of Computer Science
Technical Report

Prefix-Recognisable Graphs and Monadic Second-Order Logic

Achim Blumensath

ISSN 0935–3232 · Aachener Informatik Berichte · AIB-2001-06

RWTH Aachen · Department of Computer Science · May 2001

Prefix-Recognisable Graphs and Monadic Second-Order Logic

Achim Blumensath

LuFG Mathematische Grundlagen der Informatik
RWTH Aachen
blume@informatik.rwth-aachen.de

Abstract We present several characterisations of the class of prefix-recognisable graphs including representations via graph-grammars and MSO-interpretations. The former implies that prefix-recognisable graphs have bounded clique-width; the latter is used to extend this class to arbitrary relational structures. We prove that the prefix-recognisable groups are exactly the context-free groups. Finally, we develop methods to prove that certain structures are not prefix-recognisable and apply them to well-ordered structures.

1 Introduction

In recent years the investigation of algorithmic properties of *infinite* structures has become an established part of computer science. Its applications range from algorithmic group theory to databases and automatic verification. Infinite databases, for example, were introduced to model geometric and, in particular, geographical data (see [17] for an overview). In the field of automatic verification several classes of infinite transition systems and corresponding model-checking algorithms have been defined. For instance, model-checking for the modal μ-calculus over prefix-recognisable graphs is studied in [6,18]. A further point of interest in this context is the bisimulation equivalence of such transition systems as considered in [23].

Obviously, only restricted classes of infinite structures are suited for such an approach. In order to process a class $\mathcal{K}$ of infinite structures by algorithmic means two conditions must be met:

(i) Each structure $\mathfrak{A} \in \mathcal{K}$ must possess a *finite representation.*
(ii) The operations one would like to perform must be *effective* with regard to these representations.

One fundamental operation demanded by many applications is the evaluation of a query, that is, given a formula $\varphi(\bar{x})$ in some logic and the representation of a structure $\mathfrak{A} \in \mathcal{K}$ one wants to compute a representation of the set $\varphi^{\mathfrak{A}} := \{\, \bar{a} \mid \mathfrak{A} \models \varphi(\bar{a}) \,\}$. Slightly simpler is the model-checking problem which asks whether $\mathfrak{A} \models \varphi(\bar{a})$ for some given $\bar{a}$. The class of prefix-recognisable graphs investigated in the present article has explicitly been defined in such a way that model-checking for MSO, monadic second order logic, is decidable. To the authors knowledge it is the largest natural class with this property.

Class	Logic
context-free	$\mathrm{MSO}(\exists^\omega)$
HR-equational	$\mathrm{MSO}(\exists^\omega)$
prefix-recognisable	$\mathrm{MSO}(\exists^\omega)$
automatic	$\mathrm{FO}(\exists^\omega)$
rational	Σ_0
recursive	Σ_0

Table 1. Decidability

Several different notions of infinite graphs and structures have been considered in the literature:

- *Context-free graphs* [21,22] are the configuration graphs of pushdown automata.
- *HR-equational graphs* [9] are defined by equations of hyperedge-replacement grammars.
- *Prefix-recognisable graphs* have been introduced in [7]. Several characterisations are presented in this article.
- *Automatic graphs* [16,3,4] are graphs whose edge relation is recognised by synchronous multihead automata.
- *Rational graphs* [16,20] are graphs whose edge relation is recognised by asynchronous multihead automata.
- *Recursive graphs* [15] are graphs whose edge relation is recursive.

These classes of graphs form a strict hierarchy. In Table 1 it is shown for each class for which logic model-checking is still decidable. $\mathrm{FO}(\exists^\omega)$ and $\mathrm{MSO}(\exists^\omega)$ denote first- and monadic second-order logic extended by the quantifier "there are infinitely many", and Σ_0 is the set of quantifier-free first-order formulae.

In this article the following characterisation of the class of prefix-recognisable graphs is presented.

Theorem 1. *Let $\mathfrak{G}$ be a graph. The following statements are equivalent:*

(i) $\mathfrak{G}$ *is prefix-recognisable.*
(ii) $\mathfrak{G} = h^{-1}(\mathfrak{T}_2)|_C$ *for a rational substitution h and a regular language C.*
(iii) $\mathfrak{G}$ *is the configuration graph of a pushdown automaton with ε-transitions.*
(iv) $\mathfrak{G}$ *is* MSO*-interpretable in the binary tree $\mathfrak{T}_2$.*
(v) $\mathfrak{G}$ *is VR-equational.*

The first two items are due to Caucal [7]; Stirling [23] proved (iii); and Barthelmann [1] delivered the last two characterisations where (iv) was stated in terms of k-copying transducers as introduced by Courcelle. It will turn out that, with regard to the binary tree, these are equivalent to MSO-interpretations as defined below.

Besides giving new and much simpler proofs of (iv) and (v), the main new results presented in this article are: (1) a generalisation of (iv) to non-injective

MSO-interpretations, (2) an extension of prefix-recognisable graphs to arbitrary relational structures, and (3) a characterisation of prefix-recognisable groups and well-orderings.

The outline of the article is as follows. In Section 2 we define the class of prefix-recognisable graphs and recall several known characterisations of it.

Section 3 introduces monadic-second order logic and proves the equivalence of (i) and (iv) in Theorem 1. This result enables us to extend the definition of prefix-recognisable graphs to arbitrary relational structures which are investigated in the remainder of this section.

The characterisation in terms of vertex-replacement graph-grammars is given in Section 4 together with some results on the tree-width and clique-width of prefix-recognisable graphs.

In Section 5 we investigate the class of groups whose Cayley-graph is prefix-recognisable, and prove that they coincide with the context-free groups.

Finally, in Section 6 examples are given on how to prove that certain structures are not prefix-recognisable.

2 Prefix-recognisable graphs

Most of the classes mentioned in the introduction were originally defined as classes of edge-labelled graphs. We will represent such graphs as structures $(V, (E_a)_{a\in A})$ where each $E_a \subseteq V \times V$.

Definition 1. *A graph is* prefix-recognisable *if it is isomorphic to a graph of the form* $(S, (E_a)_{a\in A})$ *where* S *is a regular language over some alphabet* Σ *and each* E_a *is a finite union of relations of the form*

$$W(U \times V) := \{\, (wu, wv) \mid u \in U,\ v \in V,\ w \in W \,\}$$

for regular languages U, V, $W \subseteq \Sigma^*$. *The class of prefix-recognisable graphs is denoted by* PRG.

Actually in the usual definition the reverse order $(U \times V)W$ is used. The above formulation was chosen as it fits better to the usual conventions regarding trees.

Example 1. The structure $(\omega, s, \leq)$ is prefix-recognisable. If we represent the universe by a^* the relations take the form

$$s = a^*(\varepsilon \times a) \qquad \text{and} \qquad \leq\, = a^*(\varepsilon \times a^*).$$

This representation can easily be generalised to one of the complete binary tree $\mathfrak{T}_2 = \big(\{0,1\}^*, \sigma_0, \sigma_1, \preceq\big)$ which will become important in Section 3:

$$\sigma_i = \{0,1\}^*(\varepsilon \times i) \qquad \text{and} \qquad \preceq\, = \{0,1\}^*(\varepsilon \times \{0,1\}^*).$$

Originally, Caucal defined PRG in a different way. In order to obtain a class of graphs with decidable MSO-theory he defined two operations on graphs which preserve MSO-decidability and applied them to the binary tree $\mathfrak{T}_2$.

Definition 2. *Let $\mathfrak{G} = (V,\ (E_a)_{a\in A})$ be a graph with universe $V \subseteq \{0,1\}^*$.*

(1) The restriction $\mathfrak{G}|_C$ denotes the subgraph $(V \cap C,\ (E_a \cap C \times C)_a)$ induced by $C \subseteq \{0,1\}^*$.

(2) Let $\bar{A}$ be a disjoint copy of A and expand $\mathfrak{G}$ by the relations $E_{\bar{a}} := (E_a)^{-1}$ for $\bar{a} \in \bar{A}$. Given a set of labels B and a mapping h associating with every $b \in B$ a language $h(b) \subseteq (A \cup \bar{A})^*$, the inverse substitution $h^{-1}(\mathfrak{G})$ defines the graph $(V, (E'_b)_{b\in B})$ where E'_b consists of those pairs (u,v) such that in the expansion of $\mathfrak{G}$ there is a path from u to v labelled by some word in $h(b)$.

Proposition 1 (Caucal [7]). *A graph $\mathfrak{G}$ is prefix-recognisable if and only if it is isomorphic to $h^{-1}(\mathfrak{T}_2)|_C$ for some regular language C and mapping h such that $h(a) \subseteq \{0,1,\bar{0},\bar{1}\}^*$ is regular for all a.*

Example 2. $(\omega, s, \leq)$ can be defined with $C := 1^*$ and $h(s) := 1$, $h(\leq) := 1^*$.

In a similar way to the characterisation of context-free graphs as configuration graphs of pushdown automata one can describe the class of prefix-recognisable graphs via some model of automaton. To do so one considers pushdown automata with ε-transitions where each configuration has either no outgoing ε-transitions or no outgoing non-ε-transitions. Then the ε-transitions are "factored out" in the following way: one takes only those vertices without outgoing ε-transitions and adds an a-transition between two vertices iff in $\mathfrak{G}$ there is a path between them consisting of one a-transition followed by arbitrarily many ε-transitions.

Proposition 2 (Stirling [23]). *A graph $\mathfrak{G}$ is prefix-recognisable if and only if it is the configuration graph of a pushdown automaton with ε-transitions where the ε-transitions are factored out in the way describe above.*

Example 3. The pushdown automaton for $(\omega, s, \leq)$ has the following configuration graph:

$$\begin{tikzcd}
(q_0,\varepsilon) \arrow[r, "s"] \arrow[d, bend left, "\leq"] & (q_0,X) \arrow[r, "s"] \arrow[d, bend left, "\leq"] & (q_0,XX) \arrow[r, "s"] \arrow[d, bend left, "\leq"] & (q_0,XXX) \arrow[r, "s"] \arrow[d, bend left, "\leq"] & \cdots \\
(q_1,\varepsilon) \arrow[u, bend left, "\varepsilon"] & (q_1,X) \arrow[l, "\varepsilon"] \arrow[u, bend left, "\varepsilon"] & (q_1,XX) \arrow[l, "\varepsilon"] \arrow[u, bend left, "\varepsilon"] & (q_1,XXX) \arrow[l, "\varepsilon"] \arrow[u, bend left, "\varepsilon"] & \cdots \arrow[l, "\varepsilon"]
\end{tikzcd}$$

3 Interpretations in the binary tree

Interpretations are a general tool to obtain classes of finitely presented structures with a set of desired properties. One fixes some structure $\mathfrak{A}$ having these properties and chooses a kind of interpretation that preserves them. Then one considers the class of all structures which can be interpreted in $\mathfrak{A}$.

If one is interested in decidability of monadic second-order logic the canonical structure to consider is the infinite binary tree $\mathfrak{T}_2$. Thus, in the present section we investigate the class of all structures that are MSO-interpretable in $\mathfrak{T}_2$. It turns out that, restricted to graphs, this class coincides with PRG yielding a model-theoretic characterisation of the class of prefix-recognisable graphs.

Let us recall some basic definitions. MSO, monadic second-order logic, extends first-order logic FO by quantification over sets. A formula $\varphi(\bar{x})$ where each free variable is first-order defines on a given structure $\mathfrak{A}$ the relation $\varphi^{\mathfrak{A}} := \{ \bar{a} \mid \mathfrak{A} \models \varphi(\bar{a}) \}$.

Definition 3. *Let $\mathfrak{A} = (A, R_1, \dots, R_n)$ and $\mathfrak{B}$ be relational structures. A (one-dimensional)* MSO-interpretation *of $\mathfrak{A}$ in $\mathfrak{B}$ is a sequence*

$$\mathcal{I} = \langle \delta(x),\ \varepsilon(x,y),\ \varphi_{R_1}(\bar{x}), \dots,\ \varphi_{R_n}(\bar{x}) \rangle$$

of MSO*-formulae such that*

$$\mathfrak{A} \cong \big(\delta^{\mathfrak{B}},\ \varphi_{R_1}^{\mathfrak{B}}, \dots,\ \varphi_{R_n}^{\mathfrak{B}}\big)/\varepsilon^{\mathfrak{B}}.$$

To make this expression well-defined we require that $\varepsilon^{\mathfrak{B}}$ is a congruence of the structure $\big(\delta^{\mathfrak{B}},\ \varphi_{R_1}^{\mathfrak{B}}, \dots,\ \varphi_{R_n}^{\mathfrak{B}}\big)$.

We write $\mathcal{I} : \mathfrak{A} \leq_{\mathrm{MSO}} \mathfrak{B}$ if $\mathcal{I}$ is an MSO-interpretation of $\mathfrak{A}$ in $\mathfrak{B}$. Since $\mathfrak{A}$ is uniquely determined by $\mathfrak{B}$ and $\mathcal{I}$, we can regard $\mathcal{I}$ as a functor and denote $\mathfrak{A}$ by $\mathcal{I}(\mathfrak{B})$. The *coordinate map* from $\delta^{\mathfrak{B}}$ to A is also denoted by $\mathcal{I}$. We call $\mathcal{I}$ *injective* if the coordinate map is injective.

If it is not necessary to name the interpretation we simple write $\mathfrak{A} \leq_{\mathrm{MSO}} \mathfrak{B}$. $\mathfrak{A} \leftrightharpoons_{\mathrm{MSO}} \mathfrak{B}$ holds iff $\mathfrak{A} \leq_{\mathrm{MSO}} \mathfrak{B}$ and $\mathfrak{B} \leq_{\mathrm{MSO}} \mathfrak{A}$.

The decidability of the monadic-theory of $\mathfrak{T}_2$ was established by Rabin via automata theoretic techniques (see [24] for an overview). To fix our notation, a *tree over S* is a structure $(T, (\sigma_a)_{a \in S})$ where $T \subseteq S^*$ is prefix-closed and the a-successor relation σ_a contains all pairs (x, xa) for $x \in T$. The *complete tree over S* is $\mathfrak{T}_S := (S^*, (\sigma_a)_{a \in S})$. Finally, $\mathfrak{T}_n := \mathfrak{T}_{\{0,\dots,n-1\}}$. If convenient we regard trees as partial orders $(T, \preceq, \sqcap)$ where $\preceq$ is the prefix-ordering and $x \sqcap y$ denotes the longest common prefix of x and y. Further, we identify a prefix-closed set $T \subseteq S^*$ with the tree $(T, (\sigma_a)_{a \in S})$. An A-labelled tree is either represented as structure $\big(T, (\sigma_a)_{a \in S}, (P_a)_{a \in A}\big)$ or simply as a mapping $T \to A$. Finally, a *regular tree* is a tree with only finitely many subtrees up to isomorphism.

We denote automata by tuples $(Q, \Sigma, \Delta, q_0, F)$ with set of states Q, alphabet Σ, transition relation Δ, initial state q_0, and acceptance condition F. Sometimes, if the automaton is deterministic, Δ is replaced by a function δ.

To state the equivalence result we need to code tuples of sets as labelled trees.

Definition 4. *For sets $X_0, \dots, X_{n-1} \subseteq \{0,1\}^*$ denote by $T_{\bar{X}}$ the $\{0,1\}^n$-labelled binary tree such that the ith component of $T_{\bar{X}}(y)$ is 1 iff $y \in X_i$. For singletons $X_i = \{x_i\}$ we also write $T_{\bar{x}}$.*

The relation between automata and MSO logic on trees is given by Rabin's famous tree theorem:

Theorem 2. *For each* $\varphi(\bar{X}, \bar{x}) \in \mathrm{MSO}$ *there is a tree-automaton* $\mathcal{A}$ *such that* $L(\mathcal{A}) = \{\, T_{\bar{X}\bar{x}} \mid \mathfrak{T}_2 \models \varphi(\bar{X}, \bar{x}) \,\}$.

Below we will prove that a graph is prefix-recognisable if and only if it is MSO-interpretable in $\mathfrak{T}_2$. Since we want to extend this class to arbitrary relational structures it is natural to use the following definition.

Definition 5. *The class of* prefix-recognisable structures *is defined as*

$$\mathrm{PRStr} := \{\, \mathfrak{A} \mid \mathfrak{A} \leq_{\mathrm{MSO}} \mathfrak{T}_2 \,\}.$$

By the above remarks it immediately follows that MSO model-checking is decidable for every prefix-recognisable structure. This can be slightly extended to $\mathrm{MSO}(\exists^\omega)$, MSO extended by the quantifier "there are infinitely many".

Lemma 1. *For each* $\mathfrak{A} \in \mathrm{PRStr}$ *model-checking for* $\mathrm{MSO}(\exists^\omega)$ *is decidable.*

Proof. It is sufficient to construct a formula $\varphi_{\mathrm{inf}}(X)$ such that $\mathfrak{T}_2 \models \varphi_{\mathrm{inf}}(A)$ if and only if A is infinite. Then

$$\exists^\omega x \psi(x, \bar{Y}, \bar{y}) \equiv \exists X \big[\forall x \big(Xx \leftrightarrow \psi(x, \bar{Y}, \bar{y})\big) \wedge \varphi_{\mathrm{inf}}(X)\big].$$

By König's lemma $A \subseteq \{0,1\}^*$ is infinite if and only if there is an infinite path P such that for all $x \in P$ there is some $y \in A$ with $x \preceq y$. This condition can be expressed in MSO.

The first direction of the desired characterisation is a straightforward modification of the well-known translation of automata to MSO-formulae.

Lemma 2. $\mathrm{PRG} \subseteq \mathrm{PRStr}$.

Proof. Let $\mathfrak{A} \in \mathrm{PRG}$. Clearly, the prefix-ordering $\preceq$ is MSO-definable in $\mathfrak{T}_2$. Further, for each regular language $L \subseteq \{0,1\}^*$ there is a MSO-formula $\varphi_L(u, v)$ stating that $u \preceq v$ and the labeling $u^{-1}v$ of the path from u to v is in L. Thus, relations $R = \bigcup_{i<n} W_i(U_i \times V_i)$ can be defined by

$$Rxy \;:\text{iff}\; \bigvee_{i<n} \exists z \big(\varphi_{W_i}(\varepsilon, z) \wedge \varphi_{U_i}(z, x) \wedge \varphi_{V_i}(z, y)\big).$$

In order to prove the converse we first simplify the involved MSO-interpretation to injective ones.

Lemma 3. *Let* $D \subseteq \{0,1\}^*$ *be regular and* $E \subseteq D \times D$ *a prefix-recognisable equivalence relation. There is a regular language* $D' \subseteq D$ *such that* D' *contains exactly one element of each* E*-class.*

Proof. Denote the E-class of x by $[x]$, define $p_{[x]} := \inf_{\preceq}[x]$ and $s_x := (p_{[x]})^{-1}x$. Let $\varphi_p(x,y)$ be a MSO-definition of the function $x \mapsto p_{[x]}$. Finally, let s be the number of states of the automaton associated with E. We claim that each class $[x]$ contains an element of length less than $|p_{[x]}| + s$. Thus, one can define

$$D' := \{\, x \in D \mid s_x \leq s_y \text{ for all } y \in [x] \,\}$$

where $\leq$ is the lexicographic ordering which is definable since the length of the words is bounded so that we only need to consider finitely many cases.

To prove the claim choose $x_0, x_1 \in [x]$ such that $x_0 \sqcap x_1 = p_{[x]}$. Since $(x_0, x_1) \in E$ there are regular languages U, V, and W such that $x_0 = wu$, $x_1 = wv$ for $u \in U$, $v \in V$, and $w \in W$ with $w \subseteq p_{[x]}$. If $|wu| \geq |p_{[x]}| + s$ then, by a pumping argument, there exists some $u' \in U$ such that $|p_{[x]}| \leq |wu'| < |p_{[x]}| + s$. Hence, $(wu', x_1) \in E$ is an element of the desired length.

Corollary 1.

(i) PRG *is closed under prefix-recognisable congruences.*
(ii) *Each structure in* PRStr *has an injective interpretation in* $\mathfrak{T}_2$*.*

The next result was stated in Barthelmann [1] using a slightly different notion of interpretation. We present a much shorter prove based on Theorem 2.

Theorem 3. *A graph* $\mathfrak{G}$ *is in* PRG *if and only if* $\mathfrak{G} \leq_{\text{MSO}} \mathfrak{T}_2$.

Proof. It remains to prove ($\Leftarrow$). Assume that $\mathcal{I} : \mathfrak{G} \leq_{\text{MSO}} \mathfrak{T}_2$ for injective $\mathcal{I}$. Let E be an edge relation of $\mathfrak{G}$, and let $\mathfrak{A} = (Q, \{0,1\}, \Delta, q_0, \Omega)$ be the tree-automaton associated with the MSO-definition of E in $\mathfrak{T}_2$. Let $\text{Occ}(t)$ denote the set of labels which occure at some vertex of t. We classify the states of $\mathfrak{A}$ according to the set of labelles which can appear in trees that are accepted from this state.

$$\begin{aligned}
Q_\emptyset &:= \big\{ q \in Q \bigm| \text{Occ}(t) = \big\{[0,0]\big\} \text{ for all trees } t \text{ accepted from } q \big\} \\
Q_0 &:= \big\{ q \in Q \bigm| \text{Occ}(t) = \big\{[0,0],[1,0]\big\} \text{ for all trees } t \text{ accepted from } q \big\} \\
Q_1 &:= \big\{ q \in Q \bigm| \text{Occ}(t) = \big\{[0,0],[0,1]\big\} \text{ for all trees } t \text{ accepted from } q \big\} \\
Q_{0,1} &:= \big\{ q \in Q \bigm| \text{Occ}(t) = \big\{[0,0],[1,1]\big\} \text{ or } \big\{[0,0],[1,0],[0,1]\big\} \\
&\qquad\qquad \text{for all trees } t \text{ accepted from } q \big\}
\end{aligned}$$

Let $Q_\emptyset \subseteq Q$ be the set of states from which $\mathfrak{A}$ accepts the tree labelled by $[0,0]$ everywhere; Q_0 the set of states from which only trees are accepted which contain on vertex labelled $[1,0]$ and whose other vertices are labelled by $[0,0]$. Q_1 is defined analogously with $[1,0]$ repaced by $[0,1]$. Finally, let $Q_{0,1}$ contain the states from which trees are accepted that contain either a vertex labelled by $[1,1]$ or two vertex labelled by $[1,0]$ and $[0,1]$ respectively.

We construct languages U_q, V_q, and W_q such that $u \in U_q$, $v \in V_q$, and $w \in W_q$ if and only if there is an accepting run of $\mathfrak{A}$ on $T_{\{wu\},\{wv\}}$ where the node w is labelled by q. Then

$$E = \bigcup_{q \in Q} W_q(U_q \times V_q).$$

Let $W_q := L\big(Q, \{0,1\}, \Delta_{W_q}, q_0, \{q\}\big)$ where

$$\begin{aligned} \Delta_{W_q} := {} & \{\, (p, 0, p') \mid (p, [0,0], p', p_0) \in \Delta,\ p_0 \in Q_\emptyset \,\} \\ & \cup \{\, (p, 1, p') \mid (p, [0,0], p_0, p') \in \Delta,\ p_0 \in Q_\emptyset \,\} \end{aligned}$$

and $U_q := L\big(Q \cup \{q_f\}, \{0,1\}, \Delta_{U_q}, q, \{q_f\}\big)$ where

$$\begin{aligned} \Delta_{U_q} := {} & \{\, (p, 0, p') \mid (p, [0,c], p', p_0) \in \Delta,\ p_0 \in Q_\emptyset \cup Q_1,\ c \in \{0,1\} \,\} \\ & \cup \{\, (p, 1, p') \mid (p, [0,c], p_0, p') \in \Delta,\ p_0 \in Q_\emptyset \cup Q_1,\ c \in \{0,1\} \,\} \\ & \cup \{\, (p, c, q_f) \mid (p, [1,d], p_0, p_0') \in \Delta,\ p_0,\ p_0' \in Q_\emptyset \cup Q_1,\ c, d \in \{0,1\} \,\} \end{aligned}$$

V_q is defined analogously.

Although the characterisation of prefix-recognisable graphs by interpretations is quite elegant and allows a natural generalisation to arbitrary relational structures, in actual proofs it is most of the time easier to work with a more concrete characterisation in terms of languages. Hence, we will try to generalise our original definition of prefix-recognisable graphs which uses relations of the form $\bigcup_i W_i(U_i \times V_i)$ to PRStr.

Definition 6. *The* branching structure *of words* $x_0, \dots, x_{n-1} \in \{0,1\}^*$ *is the partial order* $(X, \preceq)$ *where*

$$X := \{\varepsilon\} \cup \{x_i \sqcap x_j \mid i,\ j < n\}.$$

The elements of X *are called* branching points*.*

Example 4. The branching structure of 1111, 1011, 101011 is depicted in Figure 1.

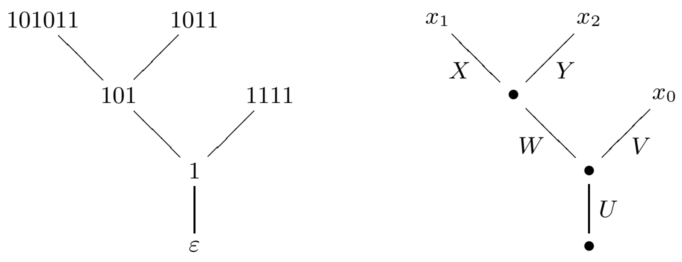

Figure 1. The branching structure of 1111, 1011, 101011 and its isomorphism type

Note that for fixed number of words there are only finitely many non-isomorphic branching structures.

Proposition 3. *An n-ary relation $R \subseteq (\{0,1\}^*)^n$ is* MSO*-definable in* $\mathfrak{T}_2$ *if and only if R is a finite union of relations of the following form:*

(a) *All tuples $\bar{x} \in R$ have the same branching structure (up to isomorphism).*
(b) *For each pair of adjacent branching points u and v there is a regular language $W_{u,v}$ such that $\bar{x} \in R$ if and only if for each such pair u, v the path between them is labelled by a word in $W_{u,v}$.*

Proof. The case of $n = 2$ was proved in Theorem 3 above. The proof of the general case is similar, so it is only sketched. ($\Leftarrow$) Clearly, each such relation is MSO-definable. ($\Rightarrow$) If R is MSO-definable the labels of paths between branching points are regular.

Example 5. For the branching structure in the previous example, a relation would be defined by five regular languages U, V, W, X, and Y (see Figure 1):

$$R = U(V \times W(X \times Y))$$

Automatic structures were first defined by Khoussainov and Nerode [16]. In the following we will use the characterisation of automatic and unary-automatic structures in terms of FO-interpretations given in Blumensath [3].

Proposition 4. PRStr $\subset$ AutStr.

Proof. Strictness follows from the fact that model-checking for MSO is decidable for PRStr but not for AutStr.

We have to show that $\mathfrak{A} \leq_{\mathrm{MSO}} \mathfrak{T}_2$ implies $\mathfrak{A} \leq_{\mathrm{FO}} (\mathfrak{T}_2, \preceq, \mathrm{el})$. Using the characterisation that is given in the preceding proposition it is sufficient to construct an FO-definition of a relation R which is defined by a certain branching-structure and regular languages W_i as described above. By a simple modification of the usual translation of automata to FO-formulae on $(\mathfrak{T}, \preceq, \mathrm{el})$ (see e.g. [5,3]) one obtains for each W_i a formula $\varphi_{W_i}(x, y)$ which states that $x \preceq y$ and the path from x to y is labelled by a word in W_i. Obviously, there also is a formula $\beta(\bar{x}, \bar{y})$ which holds iff $\bar{x}$ has a given branching-structure with universe $\bar{x}, \bar{y}$. Thus, one can define R by

$$\psi(\bar{x}) := \exists \bar{y}\Big(\beta(\bar{x}, \bar{y}) \wedge \bigwedge_i \varphi_{W_i}(z_i, z_i')\Big)$$

where $z_i,\ z_i' \in \{x_0, \dots, y_0, \dots\}$ are the branching-points corresponding to W_i.

Proposition 5. 1AutStr $\subset$ PRStr

Proof. Since $(\omega, s, \leq) \leq_{\mathrm{MSO}} \mathfrak{T}_2$, by Corollary 7.5 of [3], it is enough to construct an interpretation $(\mathbb{N}, \leq, (n \mid x)_n) \leq_{\mathrm{MSO}} (\omega, s, \leq)$. To do so we only need to define the the divisibility predicates.

$$\varphi_{n|x}(x) := \forall X(X\varepsilon \wedge \forall y(Xy \rightarrow Xs^n y) \rightarrow Xx).$$

For strictness, note that $\mathfrak{T}_2 \in$ PRStr $\setminus$ 1AutStr.

4 VR-equational graphs

The class of prefix-recognisable graphs can also be characterised via graph-grammars. This characterisation was first given by Barthelmann [1]. Following the notation of Courcelle [9,10,11] we present simplified versions of the proofs, give some corollaries, and study the class of finite induced subgraphs of a prefix-recognisable graph.

Infinite graphs can be obtained as canonical solutions of a system of equations of graph-terms over some signature Ξ.

Definition 7. *A* regular system of equations over Ξ *is a system*

$$x_1 = t_1, \quad \dots, \quad x_n = t_n$$

where $x_1, \dots, x_n$ are unknowns and $t_1, \dots, t_n$ are terms of signature Ξ with free variables among $x_1, \dots, x_n$. Furthermore, we require that no t_i consists solely of an unknown.

To define the solution of such a system we introduce infinite terms. An *countable term* over Ξ is a countable tree every vertex of which is labelled by some function of Ξ such that for every node the number of successors equals the arity of its label. The free algebra of all such terms is denoted by $\mathfrak{F}(\Xi)$.

Lemma 4 (Courcelle [8]).

(i) *Each regular system of equations has a unique solution in $\mathfrak{F}(\Xi)$ which is a tuple of regular trees.*
(ii) *Each regular tree can be obtained as solution of a regular system of equations.*

This lemma justifies the following definition.

Definition 8. *Let $\mathfrak{A}$ be a Ξ-algebra and $h : \mathfrak{F}(\Xi) \to \mathfrak{A}$ the canonical homomorphism. The* canonical solution *of a regular system of equations in $\mathfrak{A}$ is the image of the unique solution in $\mathfrak{F}(\Xi)$ under h.*

Since we are interested in defining infinite graphs by systems of equations we have to define an algebra of graphs and corresponding graph operations.

Definition 9. *The algebra of VR-equational graphs consists of all vertex-coloured graphs with countable many vertices and n colours for some fixed n. The operations are*

- *disjoint union $\oplus$,*
- *the graph a whose single vertex is coloured by a for all $a < n$,*
- *the relabelling ϱ_β for $\beta : \mathbb{N} \to \mathbb{N}$ which assigns colour $\beta(a)$ to each vertex formerly coloured by a, and*
- *the function $\eta^c_{a,b}$ which adds E_c-edges between all pairs of vertices which are coloured, respectively, a and b.*

The signature corresponding to n is denoted by $\Upsilon_n := \{\oplus, a, \varrho_\beta, \eta^c_{a,b} \mid a, b < n\}$.

Definition 10. *A graph is* VR-equational *iff it is the canonical solution of a regular system of equations over Υ_n.*

Countable terms are not only a technical tool to define VR-equational graphs. They also provide a connection to trees. The next result is Proposition 3.11 of [1]. It can also be proved analogous to Theorem 4.4 in Courcelle [10].

Proposition 6. *Let $\mathfrak{G}$ be a graph denoted by a countable term T of signature Υ_n for some $n < \omega$. Then $\mathfrak{G} \leq_{\mathrm{MSO}} T$.*

This result immediately implies one direction of the characterisation we are aiming at.

Corollary 2. *The class of VR-equational graphs is contained in* PRG.

Proof. Let $\mathfrak{G}$ be VR-equational, and let T be the regular countable term denoting $\mathfrak{G}$. Since T is regular, it can be interpreted in $\mathfrak{T}_2$. Thus, $\mathfrak{G} \leq_{\mathrm{MSO}} T \leq_{\mathrm{MSO}} \mathfrak{T}_2$.

The converse is slightly more involved.

Proposition 7 (Barthelmann [1]). *Let $\mathfrak{G} \leq_{\mathrm{MSO}} \mathfrak{T}_2$ be a graph. There is a regular countable term of signature Υ_n denoting $\mathfrak{G}$.*

Proof. We give a proof analogous to the ones of Theorem 4.6 and Corollary 4.9 in Courcelle [10].

For simplicity assume that $\mathfrak{G}$ has only one edge relation which is of the form $W(U \times V)$. Let $\mathcal{A} = (Q, \{0,1\}, \delta, q_0, F)$ and $\mathcal{A}' = (Q', \{0,1\}, \delta', q'_0, F')$ be automata recognising U and V respectively. The state transitions induced by $u \in \{0,1\}^*$ are $\delta_u(q) := \delta(q, u)$ and $\delta'_u(q) := \delta'(q, u)$. Let $D \subseteq \{0,1\}^*$ be the encoding of the universe.

For each $w \in \{0,1\}^*$ we construct a finite term $t_w(x_0, x_1)$ of signature Υ_n as described below. Let T_w be the countable term obtained from t_w by recursively substituting t_{wu0} for the occurrence of x_0 in t_{wu} and t_{wu1} for the one of x_1, i.e., $T_w = t_w(T_{w0}, T_{w1})$. The construction will ensure that T_w denotes the subgraph of $\mathfrak{G}$ induced by the set of vertices represented by $w\{0,1\}^*$. The colouring consists of pairs (f, f') of functions $f : Q \to Q$ and $f' : Q' \to Q'$. The unique vertex created by the subterm t_{wu} of T_w will be coloured (δ_u, δ'_u).

t_w is defined as follows: We update the colouring from δ_u to δ_{cu}, $c \in \{0,1\}$, and add edges between T_{w0} and T_{w1} if $w \in W$.

$$t_w(x_0, x_1) := \begin{cases} \mathrm{add}\big(\varrho_{\beta_0}(x_0) \oplus \varrho_{\beta_1}(x_1) \oplus v\big) & \text{if } w \in W, \\ \varrho_{\beta_0}(x_0) \oplus \varrho_{\beta_1}(x_1) \oplus v & \text{if } w \notin W, \end{cases}$$

where v introduces the vertex represented by w

$$v := \begin{cases} \big(\mathrm{id}_Q, \mathrm{id}_{Q'}\big) & \text{if } w \in D, \\ \emptyset & \text{otherwise,} \end{cases}$$

β_c updates the colouring

$$\beta_c(f, f') := (f \circ \delta_c, f' \circ \delta'_c),$$

and add is the composition of all functions $\alpha_{(f,f'),(g,g')}$ where $f(q_0) \in F$ and $g'(q'_0) \in F'$.

It remains to show that T_ε is regular. But this is obvious since the definition of the various t_w only depends on the membership of w in W and D which are both regular languages.

Summarising we have obtained the following characterisation of the class of prefix-recognisable graphs.

Theorem 4 (Barthelmann [1]). *Let $\mathfrak{G}$ be a graph. The following statements are equivalent:*

(i) $\mathfrak{G} \leq_{\mathrm{MSO}} \mathfrak{T}_2$.
(ii) $\mathfrak{G}$ *is VR-equational.*
(iii) $\mathfrak{G}$ *is denoted by a regular countable term of signature* Υ_n.

An infinite graph has clique-width at most n if it is denoted by a term of signature Υ_n (see Courcelle [11]). Hence, by the the third item of the above characterisation we immediately obtain the following result.

Corollary 3. *Each graph* $\mathfrak{G} \leq_{\mathrm{MSO}} \mathfrak{T}_2$ *has bounded clique-width.*

PRG contains graphs of unbounded tree-width on the other hand. For instance, $K_{\aleph_0}$, the infinite clique, is prefix-recognisable.

Proposition 8 (Barthelmann [2]). *Let* $\mathfrak{G} \in \mathrm{PRG}$. *The following statements are equivalent:*

(i) $\mathfrak{G}$ *is HR-equational.*
(ii) $\mathfrak{G}$ *has bounded tree-width.*
(iii) *There is some* $n \in \mathbb{N}$ *such that* $\mathfrak{G}$*, considered as undirected graph, does not contain the subgraph* $K_{n,n}$.

We conclude this section with an investigation of the class of finite subgraphs of a given VR-equational graph.

Theorem 5 (Engelfriet [13]). *A class* $\mathcal{K}$ *of finite graphs if generated by a VR-grammar if and only if there is a regular class* $\mathcal{T}$ *of finite binary trees such that* $\mathcal{K} \leq_{\mathrm{MSO}} \mathcal{T}$.

Lemma 5. *For all* MSO*-formulae* $\varphi(\bar{x})$ *there exists some formula* $\hat{\varphi}(\bar{x}) \in \mathrm{MSO}$ *which is equivalent modulo* $\mathrm{MTh}(\mathfrak{T}_2)$ *to* φ *such that for all subtrees* $\mathfrak{S} \subseteq \mathfrak{T}_2$ *and all elements* $\bar{a} \subseteq S$ *it holds that*

$$\mathfrak{S} \models \hat{\varphi}(\bar{a}) \text{ iff } \mathfrak{T}_2 \models \hat{\varphi}(\bar{a}).$$

Proof. Given φ consider the corresponding ω-tree automaton $\mathcal{A}$. We construct an automaton $\mathcal{A}'$ which takes labelled subtrees of $\mathfrak{T}_2$ as input and simulates the work of $\mathcal{A}$ on those. Whenever a node with some missing successors is encountered $\mathcal{A}'$ makes sure that from the state which would be assigned to these missing vertices the tree $T_\emptyset$ is accepted. Finally, let $\hat{\varphi}(\bar{x})$ be the formula associated with $\mathcal{A}'$. It follows that

$$\mathfrak{S} \models \hat{\varphi}(\bar{a}) \quad \text{iff} \quad T_{\bar{a}}|_S \in L(\mathcal{A}') \quad \text{iff} \quad T_{\bar{a}} \in L(\mathcal{A}) \quad \text{iff} \quad \mathfrak{T}_2 \models \varphi(\bar{a}).$$

Definition 11. *The* age *of a structure is the class of its finitely generated substructures.*

Lemma 6. *Let $\mathfrak{G} \leq_{\mathrm{MSO}} \mathfrak{T}_2$ be a graph. The age of $\mathfrak{G}$ can be obtained from the class of all finite labelled binary trees via an* MSO*-interpretation.*

Proof. Let $\mathcal{I} = \langle \delta(x),\ \varepsilon(x,y),\ \varphi_E(x,y) \rangle$ be the interpretation of $\mathfrak{G}$ in $\mathfrak{T}_2$. By the preceding lemma we can assume that

$$\psi^{\mathfrak{S}} = \psi^{\mathfrak{T}_2}|_S$$

for all finite subtrees $\mathfrak{S} \subseteq \mathfrak{T}_2$ where ψ is one of δ, ε, or φ_E. Let $\mathfrak{A} \subseteq \mathfrak{G}$ be a finite subgraph of $\mathfrak{G}$. Define the set $P := \mathcal{I}(A) \subseteq \{0,1\}^*$ and let $\mathfrak{S} \subseteq \mathfrak{T}_2$ be a subtree of $\mathfrak{T}_2$ whose universe contains P. Then

$$\mathcal{I}' = \langle \delta^P(x),\ \varepsilon^P(x,y),\ \varphi_E^P(x,y) \rangle$$

is an interpretation of $\mathfrak{A}$ in $(\mathfrak{S}, P)$. Conversely, each subtree of the form $(\mathfrak{S}, P)$ interprets a subgraph of $\mathfrak{G}$.

Lemma 7. *The class of finite labelled binary trees can be obtained from some regular class of finite unlabelled binary trees via an* MSO*-interpretation.*

Proof. Let $h : \{0,1\}^* \to \{0,1\}^*$ be the homomorphism defined by $h(0) = 00$ and $h(1) = 11$. Let $(\mathfrak{S}, P)$ be a labelled tree. We encode each node x of $\mathfrak{S}$ by $h(x)$. The unary predicate in encoded by appending 01 to those vertices in P. Thus, the universe of the corresponding unlabelled tree is the prefix-closure of $h(S)$ together with $h(P)01$. The interpretation is given by

$$\begin{aligned}
\delta(x) &:= \text{“}x \in (00+11)^*\text{”}, \\
\varepsilon(x,y) &:= x = y, \\
\varphi_{\sigma_c}(x,y) &:= y = xcc, \qquad \text{for } c \in \{0,1\} \\
\varphi_P(x) &:= \exists y (y = x01).
\end{aligned}$$

Clearly, the class of all encoding trees is regular.

Proposition 9. *Let $\mathfrak{G} \leq_{\mathrm{MSO}} \mathfrak{T}_2$ be a graph. The age of $\mathfrak{G}$ is generated by a VR-grammar.*

Proof. Combining the two previous lemmas the class can be obtained from a regular class of finite trees by an MSO-interpretation. By Theorem 5, this is equivalent to being generated by a VR-grammar.

5 Prefix-recognisable groups

The investigation of infinite structures with finite presentations has its origins in group theory. As this field remains an important area for the application of finitely presented structures it is natural to ask which groups are contained in PRStr.

There are basically two different ways to represent finitely generated groups as structures. Either multiplication is treated as binary function or one just includes several unary functions denoting the multiplication by a generator. If the first version is chosen it turns out that only finite groups are prefix-recognisable.

Proposition 10. *Each group* $(G, \cdot) \in \mathrm{PRStr}$ *is finite.*

Proof. Consider an MSO-interpretation $\mathcal{I} = \langle \delta, \varepsilon, \varphi_\cdot \rangle$ of $(G, \cdot)$ in $\mathfrak{T}_2$. Let k be the number of states of the automaton associated to $\varphi_\cdot$. Fix words $a_0 = xy_0$ and $a_1 = xy_1$ where $|x| \geq k$ and $|y_0| = |y_1| \geq k$. Let b represent their product. There is at least one a_i such that

(a) the distance between a_i and b is greater than k and
(b) the longest prefix of a_i which is also a prefix of b is at most of length $|x|$.

W.l.o.g. assume that $i = 0$. Consider the run on input $\{a_0\}$, $\{a_1\}$, $\{b\}$. At least one state occurs twice at the path from x to a_0. Repeating the part in between we obtain another accepting run on input $\{c\}$, $\{a_1\}$, $\{b\}$ with $c \neq a_0$. Thus (identifying an element with its representation) $c \cdot a_1 = b = a_0 \cdot a_1$ which implies $c = a_1$. Contradiction.

Example 6. $(\mathbb{Z}, +) \notin \mathrm{PRStr}$

The second type of presentation is called the *Cayley-graph* of a group. Given a set $S \subseteq G$ of semigroup generators, the Cayley-graph of $\mathfrak{G}$ is the structure

$$\Gamma(\mathfrak{G}, S) := \big(G, (f_e)_{e \in S}\big)$$

where $f_e(x) := xe$. Since $\Gamma(\mathfrak{G}, S) \leq_{\mathrm{MSO}} \mathfrak{G}$, the requirement that the Cayley-graph is prefix-recognisable is weaker than the one that $\mathfrak{G} \leq \mathfrak{T}_2$. It turns out that we indeed obtain a larger class of groups using this representations. Thus we will say that a finitely generated group is prefix-recognisable iff its Cayley-graph is.

Example 7. Let $\mathfrak{G}$ be the free group of two generators a and b. Its Cayley-graph is prefix-recognisable. Let $S := \{a, b, a^{-1}, b^{-1}\}$. The universe consists of all words over S which are reduced, that is, they do not contain one of the following factors:

$$aa^{-1}, \quad a^{-1}a, \quad bb^{-1}, \quad b^{-1}b.$$

The multiplication by a takes words w not ending in a^{-1} to wa and words of the form $w = ua^{-1}$ to u. Similarly for the other generators. Clearly, all of those relations are of the required form.

The obvious next step is to investigate which groups are prefix-recognisable. In the following we will prove that a group is prefix-recognisable if and only if it is context-free. A group is said to be *context-free* (Muller and Schupp [21,22]) if its Cayley-graph is a pushdown graph with the additional requirement that the alphabet used is the set of generators and the encoding function $S^* \to G$ is the natural homomorphism. The class of context-free groups is well investigated and has several characterisations.

Theorem 6. *Let $\mathfrak{G}$ be a finitely generated group. The following statements are equivalent:*

(i) *$\mathfrak{G}$ is context-free.*
(ii) *$\mathfrak{G}$ is virtually free.*
(iii) *$\Gamma(\mathfrak{G}, S)$ has only finitely many non-isomorphic ends.*
(iv) *The labels of circles in $\Gamma(\mathfrak{G}, S)$ form a context-free language.*
(v) *$\Gamma(\mathfrak{G}, S)$ is κ-triangulable for some finite κ.*

In this section we will add the following items to the above characterisation:

(vi) $\Gamma(\mathfrak{G}, S)$ is prefix-recognisable.
(vii) $\Gamma(\mathfrak{G}, S) \leq_{\mathrm{MSO}} \mathfrak{T}_2$.
(viii) $\Gamma(\mathfrak{G}, S)$ is VR-equational.
(ix) $\Gamma(\mathfrak{G}, S)$ has bounded strong tree-width.

The next lemma implies that a group is prefix-recognisable if and only if its Cayley-graph is a context-free graph.

Lemma 8. *Let f be an unary injective function of a structure in* PRStr*. Then f is a context-free function.*

Proof. Let $f := \bigcup_i W_i(U_i \times V_i)$. Each V_i is a singleton as f is a function, and each U_i is a singleton since f is injective.

Corollary 4. $\Gamma(\mathfrak{G}, S) \in \mathrm{PRG}$ *if and only if* $\Gamma(\mathfrak{G}, S)$ *is a context-free graph.*

It remains to show that one can always find a representation using the canonical encoding.

Remark 1 (Senizergues). Note that in the case of automatic groups this is not the case. According to Epstein et.al. [14], the Heisenberg group $\mathfrak{H}$ does not have an automatic presentation with the canonical encoding. On the other hand, using the matrix representation of $\mathfrak{H}$ one can easily construct a (3-dimensional) interpretation of $\mathfrak{H}$ in $(\mathbb{N}, +)$ showing that $\mathfrak{H} \in \mathrm{AutStr}$.

First, we need some technical definitions and results. In the remainder of this section let $\Gamma(\mathfrak{G}, S) = (G, (f_e)_{e \in S}) \in \mathrm{PRStr}$ be the Cayley graph of a finitely generated group. For $e \in S$ let

$$f_e = \bigcup_{i < r^e} W_i^e\big(\{u_i^e\} \times \{v_i^e\}\big),$$
$$k := \max\big\{\, |u_i^e|, |v_i^e| \;\big|\; e \in S,\ i < r^e \,\big\},$$

for some u_i^e, $v_i^e \in \{0,1\}^*$, and regular languages W_i^e. Fix automata for each W_i^e and let s be their maximal number of states. We denote by $\operatorname{prf}_k x$ and $\operatorname{suf}_k x$ the, respectively, prefix and suffix of x of length k. u/k is the prefix of u of length $|u| - k$.

With each element $a \in G$ we associate the finite amount of information $\tau(a) := (\bar{q},\ \operatorname{suf}_k a)$ where $\bar{q}$ consists of the states the automata recognising W_i^e have reached after having read a/k. The important property of τ is that under certain conditions one can compute $\tau(f_e a)$ from $\tau(a)$.

Lemma 9. *Let $\tau(a) = \tau(a')$. If $|f_e a| \geq |a|$ then $\tau(f_e a) = \tau(f_e a')$.*

A *path* from a to b is a sequence $a_0, \ldots, a_n$ with $a_0 = a$ and $a_n = b$ such that $a_{i+1} = f_{e_i} a_i$ for some $e_i \in S$ and all $i < n$. The word $e_0 \cdots e_{n-1}$ is called the *label* of the path. We say that a path connects a and b *above* if $a \sqcap b$ is a prefix of each a_i. The next lemma is proved similarly to the previous one.

Lemma 10. *Let $a_0, \ldots, a_n$ be a path connecting a_0 and a_n above with label $e_1 \cdots e_n \in S^*$. For all a'_0, a'_n with*

$$\tau(a'_0) = \tau(a_0), \quad \tau(a'_n) = \tau(a_n), \quad \textit{and} \quad |a'_n| - |a'_0| = |a_n| - |a_0|$$

it holds that $f_{e_n} \cdots f_{e_1} a'_0 = a'_n$.

This result allows us to obtain bounds on the distance of vertices in the Cayley-graph. The following consequence is the key tool in the construction of a triangulation below.

Lemma 11. *There is a constant L such that, if any vertices a and b with $|a|, |b| \leq |a \sqcap b| + k$ are connected by a path above, then they are connected above by a path of length at most L.*

Proof. For a, $b \in G$ with $|a|, |b| \leq |a \sqcap b| + k$ let $n(a,b)$ be the length of the shortest path connecting a and b above. If there is no such path let $n(a,b) := \infty$. By Lemma 10, $n(a,b)$ only depends on $\tau(a)$, $\tau(b)$, and $|a| - |b|$. Since the latter is bounded by k these can take only finitely many different values. Hence, so can $n(a,b)$ and the maximum exists. Thus,

$$L := \max\{\, n(a,b) \mid n(a,b) \neq \infty \,\}$$

is well-defined.

Theorem 7. *Let $\Gamma(\mathfrak{G}, S)$ be the Cayley graph of a finitely generated group. The group is context-free if and only if $\Gamma(\mathfrak{G}, S) \leq_{\mathrm{MSO}} \mathfrak{T}_2$.*

Proof. ($\Rightarrow$) is immediate. To prove ($\Leftarrow$) we show that $\Gamma(\mathfrak{G}, S)$ is L-triangulable. Let $a_0, \ldots, a_{n-1}$ be a cycle where each edge is either an actual edge of $\Gamma(\mathfrak{G}, S)$ or represents a path of length at most L connecting its endpoints above. Suppose that each edge (a,b) of the cycle satisfies $|a|,\ |b| \leq |a \sqcap b| + k$. (Note that this is always the case for actual edges.) By induction on the length n we construct an L-triangulation.

The case $n \leq 3$ being trivial suppose $n > 3$. Let a_i be an element of maximal length. W.l.o.g. assume that $|a_{i-1} \sqcap a_i| \leq |a_{i+1} \sqcap a_i|$. Then $a_{i-1} \sqcap a_{i+1} = a_{i-1} \sqcap a_i$ and the induction hypothesis implies that

$$a_i/k \preceq a_{i-1} \sqcap a_i \preceq a_i \sqcap a_{i+1} \preceq a_i .$$

Thus $|a_{i-1}|,\ |a_{i+1}| \leq |a_{i-1} \sqcap a_{i+1}| + k$. As there is a path connecting a_{i-1} and a_{i+1} above this implies that there is one of length at most L. This path together with the L-triangulation of $a_0, \dots, a_{i-1}, a_{i+1}, \dots, a_{n-1}$ which exists by induction yields the desired triangulation.

In a similar way one can prove the following characterisation of the class of context-free groups. We call a tree-decomposition *strong* if each of its components is connected.

Theorem 8 (Ly [19]). *A finitely generated group $\mathfrak{G}$ is context-free if and only if $\Gamma(\mathfrak{G}, S)$ has a strong tree-decomposition of finite width.*

Proof. ($\Rightarrow$) Let $A_w := w\{0,1\}^{\leq k}$. Note that $\{\, A_w \mid w \in \{0,1\}^* \,\}$ is a tree-decomposition of $\Gamma(\mathfrak{G}, S)$ since

$$|u|,\ |f_e(u)| \leq |u \sqcap f_e(u)| + k,$$

which implies that $u, f_e(u) \in A_{u \sqcap f_e(u)}$ for all $u \in \{0,1\}^*$. Furthermore, each component A_w is connected since the universe of $\Gamma(\mathfrak{G}, S)$ is prefix-closed.

($\Leftarrow$) Let k be the strong tree-width of $\Gamma(\mathfrak{G}, S)$. We prove the claim by constructing a k-triangulation. Let $a_0, \dots, a_{n-1}$ be a cycle where each edge is either an actual edge of $\Gamma(\mathfrak{G}, S)$ or represents a path of length at most k. Let T be the subtree of the decomposition which contains edges of the cycle. The triangulation is constructed by induction on the size of T. Consider a leaf X of T containing $a_i, \dots, a_{i+m}$ but neither a_{i-1} nor a_{i+m+1}. Note that any two vertices in X are connected by a path of length at most k. If $m = 1$ the edge (a_i, a_{i+1}) is also contained in the predecessor of X which, thus, can be deleted from T. Otherwise, add a path of length at most k from a_i to a_{i+k} to the cycle. By induction hypothesis there is a k-triangulation of the cycle $a_0, \dots, a_i, a_{i+k}, \dots, a_n$. By adding paths, say, from a_i to each a_{i+j}, for $j < k$, it can be completed to a triangulation of the whole cycle.

6 Well-Orderings

The biggest obstacle in proving that some structure is not prefix-recognisable is that the used encoding of elements is unknown. In this section we develop methods to gain information about the possible encodings based on cardinality considerations.

Lemma 12. *Let $\mathfrak{A} \in \mathrm{PRStr}$ and $\varphi(x, y) \in \mathrm{MSO}$ such that the set*

$$\varphi(a, A) := \{ b \in A \mid \mathfrak{A} \models \varphi(a, b) \}$$

is finite for every $a \in A$. There are constants s and k such that $\varphi(a, A)$ contains at most k words of each length less than $|a| + s$ and no words of greater length. In particular, $|\varphi(a, A)| \in \mathcal{O}(|a|)$.

Proof. Let s be the number of states of the automaton associated with φ. If a and b are words satisfying φ then $|b| < |a \sqcap b| + s$ since otherwise there would be a repetition of states on the path from $a \sqcap b$ to b and one could apply pumping to obtain infinitely many accepted pairs a, b'. Since, for fixed $m := |a \sqcap b|$, $\varphi(a, A)$ contains at most 2^{l-m} different words of length l, and since $l < m + s$, we obtain the following bound for the number of words of length l in $\varphi(a, A)$:

$$k := \sum_{m=l-s}^{l} 2^{l-m} = \sum_{i=0}^{s} 2^i = 2^{s+1} - 1.$$

We apply this result to well-ordered structures.

Lemma 13. *Let $\mathfrak{A} \in \mathrm{PRStr}$ be well-ordered. Then $|a_n| \in \Theta(n)$ for $n < \omega$ where a_n is the nth element of $\mathfrak{A}$.*

Proof. Clearly, $|a_n| \in \mathcal{O}(n)$ since the successor function is definable and, thus, $|a + 1| < |a| + s$ for some constant s.

To show the other bound consider some element a_n. The preceding lemma implies that $|a| \leq |a_n| + s$ for every $a < a_n$ and that there are at most $k(|a_n| + s)$ such elements. Thus, $n \leq k(|a_n| + s)$, that is, $|a_n| \geq n/k - s \in \Omega(n)$.

This implies that, as far as well-orderings are concerned, PRStr collapses to 1AutStr.

Proposition 11. *Let $\mathfrak{A}$ be well-ordered of order type $\alpha < \omega^2$. $\mathfrak{A} \in \mathrm{PRStr}$ if and only if $\mathfrak{A} \in \mathrm{1AutStr}$.*

Proof. Since $\mathrm{1AutStr} \subseteq \mathrm{PRStr}$ it remains to show the other direction. Let $\mathfrak{A} \in \mathrm{PRStr}$ be well-ordered or order-type α where $\omega(n-1) \leq \alpha < \omega n$. For $i < n$, we denote the elements a of A with $\omega i \leq a < \omega(i+1)$ which are of length l by $a^i_{l,0}, \ldots, a^i_{l,m}$ where, according to Lemma 12, m is bounded by some constant k. Further, we require that $a^i_{l,0}, \ldots, a^i_{l,m}$ are sorted lexicographically. By applying the homomorphism $0 \mapsto 0^{nk}$ and $1 \mapsto 1^{nk}$ we can assume that A contains only elements whose length is a multiple of nk. To construct a unary presentation of $\mathfrak{A}$ we encode the element $a^i_{nkl,j}$ by the word $1^{nkl+ki+j}$. It remains to define, for each relation R, formulae φ_R such that

$$(\mathfrak{T}_2, \preceq, \mathrm{el}) \models \varphi_R(1^{kl_0+i_0}, \ldots, 1^{kl_r+i_r}) \quad \text{iff} \quad (a^{i_0}_{nkl_0,j_0}, \ldots, a^{i_r}_{nkl_r,j_r}) \in R.$$

Since PRStr $\subseteq$ AutStr there is a formula ψ_R defining R in $(\mathfrak{T}_2, \preceq, \mathrm{el})$ which can be used to define φ_R if we are able to decode $1^{nkl+ki+j}$ into $a^i_{nkl,j}$. Given a word $1^{nkl+ki+j}$ we search for all elements $a^i_{nkl,0}, \dots, a^i_{nkl,m}$ and pick the $(ki+j)$-th one. Let $\delta(x)$ be the formula defining the universe, define

$$\vartheta_i(x) := \bigvee_{j<k} |x| \equiv (ki+j) \pmod{nk}$$

and let

$$\chi(x,y) := \delta(y) \ \wedge\ |x| - nk < |y| \le |x| \ \wedge\ \bigwedge_{i<n} (\vartheta_i(x) \leftrightarrow \vartheta_i(y))$$

which states that y is one of the $a^i_{nkl,j'}$, $j' < k$, if x is the encoding $1^{nkl+ki+j}$. For simplicity, consider the case of unary relations only. We can define R by

$$\varphi_R(x) := \bigvee_{\substack{i<n\\ m\le k}} \exists x_0 \cdots x_{m-1} \Big(\exists^{=m} y \chi(x,y) \wedge \bigwedge_{i<m} \chi(x,x_i) \wedge \bigwedge_{i<j} x_i <_{\mathrm{lex}} x_j$$
$$\wedge \bigvee_{j<m} \big(|x| = |x_0| + ki + j \ \wedge\ \psi_R(x_i)\big)\Big).$$

We considered the interpretation-closure of the binary tree in order to obtain a class of countable structures for which model-checking for MSO is decidable. To the authors knowledge there is no larger natural class with this property. Thus, one might ask whether PRStr already includes *all* countable structures with decidable MSO model-checking. Our final application of Lemma 12 shows that this is not the case.

Proposition 12. *There is a structure* $\mathfrak{A}$ *for which* MSO *model-checking is decidable but* $\mathfrak{A} \not\le_{\mathrm{MSO}} \mathfrak{T}_2$

Proof. Let $\mathfrak{A} := (\omega, <, P_{\mathrm{fac}})$ where $P_{\mathrm{fac}} := \{\, n! \mid n \in \mathbb{N} \,\}$. According to Elgot and Rabin [12] the monadic theory of $\mathfrak{A}$ is decidable. Since each element of this structure is definable this implies that MSO model-checking is decidable as well.

Assume that $\mathfrak{A} \le_{\mathrm{MSO}} \mathfrak{T}_2$. The regularity of P_{fac} implies that there is a sequence $a_0, a_1, \ldots \in P_{\mathrm{fac}}$ such that $|a_0|, |a_1|, \ldots$ forms an arithmetical progression. In particular, $|a_n| \in \Theta(n)$. By the above Lemma this implies that the set $\{\, a \in \omega \mid a < a_n \,\}$ has size $\Theta(n)$. Contradiction.

References

1. K. Barthelmann, *On equational simple graphs*, Tech. Rep. 9, Universität Mainz, Institut für Informatik, 1997.
2. ———, *When can an equational simple graph be generated by hyperedge replacement?*, LNCS, 1450 (1998), pp. 543–552.
3. A. Blumensath, *Automatic Structures*, Diploma Thesis, RWTH Aachen, 1999.
4. A. Blumensath and E. Grädel, *Automatic structures*, in Proc. 15th IEEE Symp. on Logic in Computer Science, 2000, pp. 51–62.

5. V. Bruyère, G. Hansel, C. Michaux, and R. Villemaire, *Logic and p-recognizable sets of integers*, Bull. Belg. Math. Soc., 1 (1994), pp. 191–238.
6. O. Burkart, *Model checking rationally restricted right closures of recognizable graphs.*, ENTCS, 9 (1997).
7. D. Caucal, *On infinite transition graphs having a decidable monadic theory*, LNCS, 1099 (1996), pp. 194–205.
8. B. Courcelle, *Fundamental properties of infinite trees*, Theoretical Computer Science, 25 (1983), pp. 95–169.
9. ———, *The monadic second-order logic of graphs II: Infinite graphs of bounded width*, Math. System Theory, 21 (1989), pp. 187–221.
10. ———, *The monadic second-order logic of graphs VII: Graphs as relational structures*, Theoretical Computer Science, 101 (1992), pp. 3–33.
11. ———, *Clique-width of countable graphs: A compactness property.* unpublished, 2000.
12. C. C. Elgot and M. O. Rabin, *Decidability and undecidability of extensions of second (first) order theory of (generalized) sucessor*, J. of Symbolic Logic, 31 (1966), pp. 169–181.
13. J. Engelfriet, *A characterization of context-free NCE graph languages by monadic second-order logic on trees*, LNCS, 532 (1991), pp. 311–327.
14. D. B. A. Epstein, J. W. Cannon, D. F. Holt, S. V. F. Levy, M. S. Paterson, and W. P. Thurston, *Word Processing in Groups*, Jones and Bartlett Publishers, Boston, 1992.
15. Y. L. Ershov, S. S. Goncharov, A. Nerode, and J. B. Remmel, *Handbook of Recursive Mathematics*, North-Holland, 1998.
16. B. Khoussainov and A. Nerode, *Automatic presentations of structures*, LNCS, 960 (1995), pp. 367–392.
17. G. Kuper, L. Libkin, and J. Paredaens, *Constraint Databases*, Springer-Verlag, 2000.
18. O. Kupferman and M. Y. Vardi, *An automata-theoretic approach to reasoning about infinite-state systems*, LNCS, 1855 (2000), pp. 36–52.
19. O. Ly, 2000. personal communication.
20. C. Morvan, *On rational graphs*, LNCS, 1784 (1996), pp. 252–266.
21. D. E. Muller and P. E. Schupp, *Groups, the theory of ends, and context-free languages*, J. of Computer and System Science, 26 (1983), pp. 295–310.
22. ———, *The theory of ends, pushdown automata, and second-order logic*, Theoretical Computer Science, 37 (1985), pp. 51–75.
23. C. Stirling, *Decidability of bisimulation equivalence for pushdown processes.* unpublished, 2000.
24. W. Thomas, *Languages, automata, and logic*, in Handbook of Formal Languages, G. Rozenberg and A. Salomaa, eds., vol. 3, Springer, New York, 1997, pp. 389–455.

Aachener Informatik-Berichte

This is a list of recent technical reports. To obtain copies of technical reports please consult http://aib.informatik.rwth-aachen.de/ or send your request to: Informatik-Bibliothek, RWTH Aachen, Ahornstr. 55, 52056 Aachen, Email: `biblio@informatik.rwth-aachen.de`

95-11 * M. Staudt / K. von Thadden: Subsumption Checking in Knowledge Bases

95-12 * G.V. Zemanek / H.W. Nissen / H. Hubert / M. Jarke: Requirements Analysis from Multiple Perspectives: Experiences with Conceptual Modeling Technology

95-13 * M. Staudt / M. Jarke: Incremental Maintenance of Externally Materialized Views

95-14 * P. Peters / P. Szczurko / M. Jeusfeld: Business Process Oriented Information Management: Conceptual Models at Work

95-15 * S. Rams / M. Jarke: Proceedings of the Fifth Annual Workshop on Information Technologies & Systems

95-16 * W. Hans / St. Winkler / F. Sáenz: Distributed Execution in Functional Logic Programming

96-1 * Jahresbericht 1995

96-2 M. Hanus / Chr. Prehofer: Higher-Order Narrowing with Definitional Trees

96-3 * W. Scheufele / G. Moerkotte: Optimal Ordering of Selections and Joins in Acyclic Queries with Expensive Predicates

96-4 K. Pohl: PRO-ART: Enabling Requirements Pre-Traceability

96-5 K. Pohl: Requirements Engineering: An Overview

96-6 * M. Jarke / W. Marquardt: Design and Evaluation of Computer–Aided Process Modelling Tools

96-7 O. Chitil: The ς-Semantics: A Comprehensive Semantics for Functional Programs

96-8 * S. Sripada: On Entropy and the Limitations of the Second Law of Thermodynamics

96-9 M. Hanus (Ed.): Proceedings of the Poster Session of ALP'96 — Fifth International Conference on Algebraic and Logic Programming

96-10 R. Conradi / B. Westfechtel: Version Models for Software Configuration Management

96-11 * C. Weise / D. Lenzkes: A Fast Decision Algorithm for Timed Refinement

96-12 * R. Dömges / K. Pohl / M. Jarke / B. Lohmann / W. Marquardt: PRO-ART/CE* — An Environment for Managing the Evolution of Chemical Process Simulation Models

96-13 * K. Pohl / R. Klamma / K. Weidenhaupt / R. Dömges / P. Haumer / M. Jarke: A Framework for Process-Integrated Tools

96-14 * R. Gallersdörfer / K. Klabunde / A. Stolz / M. Eßmajor: INDIA — Intelligent Networks as a Data Intensive Application, Final Project Report, June 1996

96-15 * H. Schimpe / M. Staudt: VAREX: An Environment for Validating and Refining Rule Bases

96-16 * M. Jarke / M. Gebhardt, S. Jacobs, H. Nissen: Conflict Analysis Across Heterogeneous Viewpoints: Formalization and Visualization

96-17 M. Jeusfeld / T. X. Bui: Decision Support Components on the Internet

96-18 M. Jeusfeld / M. Papazoglou: Information Brokering: Design, Search and Transformation

96-19 * P. Peters / M. Jarke: Simulating the impact of information flows in networked organizations

96-20 M. Jarke / P. Peters / M. Jeusfeld: Model-driven planning and design of cooperative information systems

96-21 * G. de Michelis / E. Dubois / M. Jarke / F. Matthes / J. Mylopoulos / K. Pohl / J. Schmidt / C. Woo / E. Yu: Cooperative information systems: a manifesto

96-22 * S. Jacobs / M. Gebhardt, S. Kethers, W. Rzasa: Filling HTML forms simultaneously: CoWeb architecture and functionality

96-23 * M. Gebhardt / S. Jacobs: Conflict Management in Design

97-01 Jahresbericht 1996

97-02 J. Faassen: Using full parallel Boltzmann Machines for Optimization

97-03 A. Winter / A. Schürr: Modules and Updatable Graph Views for PROgrammed Graph REwriting Systems

97-04 M. Mohnen / S. Tobies: Implementing Context Patterns in the Glasgow Haskell Compiler

97-05 * S. Gruner: Schemakorrespondenzaxiome unterstützen die paargrammatische Spezifikation inkrementeller Integrationswerkzeuge

97-06 M. Nicola / M. Jarke: Design and Evaluation of Wireless Health Care Information Systems in Developing Countries

97-07 P. Hofstedt: Taskparallele Skelette für irregulär strukturierte Probleme in deklarativen Sprachen

97-08 D. Blostein / A. Schürr: Computing with Graphs and Graph Rewriting

97-09 C.-A. Krapp / B. Westfechtel: Feedback Handling in Dynamic Task Nets

97-10 M. Nicola / M. Jarke: Integrating Replication and Communication in Performance Models of Distributed Databases

97-13 M. Mohnen: Optimising the Memory Management of Higher-Order Functional Programs

97-14 R. Baumann: Client/Server Distribution in a Structure-Oriented Database Management System

97-15 G. H. Botorog: High-Level Parallel Programming and the Efficient Implementation of Numerical Algorithms

98-01 * Jahresbericht 1997

98-02 S. Gruner/ M. Nagel / A. Schürr: Fine-grained and Structure-oriented Integration Tools are Needed for Product Development Processes

98-03 S. Gruner: Einige Anmerkungen zur graphgrammatischen Spezifikation von Integrationswerkzeugen nach Westfechtel, Janning, Lefering und Schürr

98-04 * O. Kubitz: Mobile Robots in Dynamic Environments

98-05 M. Leucker / St. Tobies: Truth — A Verification Platform for Distributed Systems

98-07 M. Arnold / M. Erdmann / M. Glinz / P. Haumer / R. Knoll / B. Paech / K. Pohl / J. Ryser / R. Studer / K. Weidenhaupt: Survey on the Scenario Use in Twelve Selected Industrial Projects

98-08 * H. Aust: Sprachverstehen und Dialogmodellierung in natürlichsprachlichen Informationssystemen

98-09 * Th. Lehmann: Geometrische Ausrichtung medizinischer Bilder am Beispiel intraoraler Radiographien

98-10 * M. Nicola / M. Jarke: Performance Modeling of Distributed and Replicated Databases

98-11 * A. Schleicher / B. Westfechtel / D. Jäger: Modeling Dynamic Software Processes in UML

98-12 * W. Appelt / M. Jarke: Interoperable Tools for Cooperation Support using the World Wide Web

98-13 K. Indermark: Semantik rekursiver Funktionsdefinitionen mit Striktheitsinformation

99-01 * Jahresbericht 1998

99-02 * F. Huch: Verifcation of Erlang Programs using Abstract Interpretation and Model Checking — Extended Version

99-03 * R. Gallersdörfer / M. Jarke / M. Nicola: The ADR Replication Manager

99-04 M. Alpuente / M. Hanus / S. Lucas / G. Vidal: Specialization of Functional Logic Programs Based on Needed Narrowing

99-07 Th. Wilke: CTL+ is exponentially more succinct than CTL

99-08 O. Matz: Dot-Depth and Monadic Quantifier Alternation over Pictures

2000-01 * Jahresbericht 1999

2000-02 Jens Vöge / Marcin Jurdziński: A Discrete Strategy Improvement Algorithm for Solving Parity Games

2000-04 Andreas Becks, Stefan Sklorz, Matthias Jarke: Exploring the Semantic Structure of Technical Document Collections: A Cooperative Systems Approach

2000-05 * Mareike Schoop: Cooperative Document Management

2000-06 * Mareike Schoop (Ed.): Proceedings of the Fifth International Workshop on the Language-Action Perspective on Communication Modelling

2000-07 * Markus Mohnen / Pieter Koopman (Eds.): Proceedings of the 12th International Workshop of Functional Languages

2000-08 Thomas Arts / Thomas Noll: Verifying Generic Erlang Client-Server Implementations

2001-01 * Jahresbericht 2000

2001-02 Benedikt Bollig / Martin Leucker: Deciding LTL over Mazurkiewicz Traces

2001-03 Thierry Cachat: The power of one-letter rational languages

2001-04 Benedikt Bollig / Martin Leucker / Michael Weber: Local Parallel Model Checking for the Alternation free μ-calculus

2001-05 Benedikt Bollig / Martin Leucker / Thomas Noll: Regular MSC languages

* These reports are only available as a printed version.

Please contact `biblio@informatik.rwth-aachen.de` to obtain copies.