

THE COMPLEXITY OF BISIMULATION AND SIMULATION ON FINITE SYSTEMS

MOSES GANARDI, STEFAN GÖLLER, AND MARKUS LOHREY

University of Siegen, Germany
e-mail address: ganardi@eti.uni-siegen.de

Laboratoire Specification et Verification (LSV), ENS de Cachan & CNRS, France
e-mail address: goeller@lsv.ens-cachan.fr

University of Siegen, Germany
e-mail address: lohrey@eti.uni-siegen.de

ABSTRACT. In this paper the computational complexity of the (bi)simulation problem over restricted graph classes is studied. For trees given as pointer structures or terms the (bi)simulation problem is complete for logarithmic space or NC^1 , respectively. This solves an open problem from Balcázar, Gabarró, and Sántha. Furthermore, if only one of the input graphs is required to be a tree, the bisimulation (simulation) problem is contained in AC^1 (LogCFL). In contrast, it is also shown that the simulation problem is P -complete already for graphs of bounded path-width.

1. INTRODUCTION

Courcelle’s theorem states that every problem definable in monadic second-order logic (MSO) is solvable in linear time on graphs of bounded tree-width. In recent works by Elberfeld, Jakoby, and Tantau, techniques have been developed to transfer this famous result to low space and circuit complexity classes [EJT10, EJT12]. In particular, the following logspace (resp., NC^1) version of Courcelle’s theorem was shown (see Section 2 for the necessary definitions):

Theorem 1.1 [EJT10, EJT12]. *For a fixed MSO-sentence ψ and a fixed constant k one can check in logspace whether a given structure \mathcal{A} of tree-width at most k satisfies ψ . If a tree decomposition of \mathcal{A} of width k is given in term representation, then one can check in DLOGTIME-uniform NC^1 whether \mathcal{A} satisfies ψ .*

This result is a very powerful metatheorem, which can be applied to many computational problems. On the other hand, there are important problems solvable in logspace on graphs of bounded tree-width that are not covered by Theorem 1.1. One example is the graph isomorphism problem. Graph isomorphism is not MSO-definable even over finite paths since two finite paths are isomorphic if they have the same length, but one cannot express in MSO

Key words and phrases: bisimulation, computational complexity, tree-width.

that two finite sets have the same size. Lindell [Lin92] has shown that isomorphism of trees is in logspace, and only very recently Elberfeld and Schweitzer [ES16] extended this result to graphs of bounded tree-width.

In this paper, we are concerned with the complexity of simulation and bisimulation, which are of fundamental importance in the theory of reactive systems, see e.g. [AI07] for more background. It is known that on finite state systems simulation and bisimulation are both \mathbf{P} -complete [BGS92], and hence have no efficient parallel algorithm unless $\mathbf{P} = \mathbf{NC}$. Surprisingly, no results on the complexity of (bi)simulation on natural subclasses of finite state systems are known (whereas there exists an extensive literature on (bi)simulation problems for various classes of infinite state systems, like pushdown systems or Petri nets; see e.g. [Srb04]). The authors of [BGS92] pose this open question and suggest to consider the bisimulation problem on trees. The above remark that tree isomorphism cannot be expressed in MSO applies to bisimulation on trees as well (two finite paths are bisimilar if and only if they are isomorphic). Moreover, it is not clear whether there is a natural reduction of the bisimulation problem on trees to the logspace-solvable isomorphism problem for trees (or even bounded tree-width graphs).

In this paper, we determine the complexity of the bisimulation problem and simulation problem on several subclasses of finite state systems. More precisely, we show the following results; see also Table 1.

- (1) On trees the (bi)simulation problem is complete for logarithmic space (resp., \mathbf{NC}^1) if the trees are given as pointer structures (resp., in term representation).
- (2) The bisimulation problem between a tree and a dag (or arbitrary graph) belongs to \mathbf{AC}^1 and is \mathbf{NL} -hard.
- (3) The simulation problem is \mathbf{P} -hard (and hence \mathbf{P} -complete) already for graphs of bounded path-width.
- (4) Simulation of a tree by a dag as well as simulation of a dag by a tree is \mathbf{LogCFL} -complete.

Whether the bisimulation problem on graphs of bounded tree-width is in \mathbf{NC} remains open.

We prove our results for the bisimulation problem for trees (statement (1) above) by a reduction to the evaluation problem for a new class of Boolean circuits that we call *tree-shaped circuits*. These are circuits that are composed in a tree-like fashion of smaller subcircuits. The important parameter of such a circuit is its *width* which is defined to be the maximal number of different paths from the root to an input gate in one of the above mentioned subcircuits. We emphasize that the size of the subcircuits (number of gates) may be unbounded. The main technical contributions of this paper are logspace- and \mathbf{NC}^1 -evaluation algorithms (depending on the input representation) for tree-shaped circuits of bounded width. These circuits should not be confused with circuits of bounded tree-width, which are known to have logspace- and \mathbf{NC}^1 -evaluation algorithms (depending on the representation) by the above Theorem 1.1. We show how to partially unfold tree-shaped circuits of bounded width into circuits of bounded tree-width. This unfolding is possible in \mathbf{TC}^0 (assuming the right representation of the circuit). Finally, the resulting bounded tree-width circuit can be evaluated using Theorem 1.1. For the above logspace result we actually prove a stronger statement: A given tree-shaped circuit of size N and width m can be evaluated in space $O(\log N \cdot \log m)$.

One should also mention the paper [GGHL12], where a logic LREG (which extends classical first-order logic) is introduced. It is shown that LREG captures logspace on directed trees. Hence, bisimulation for trees is expressible in LREG. Due to the very technical

	trees (term rep.)	trees (pointer rep.)	tree, graph	bounded path-width	general graphs
bisimulation	NC ¹ -compl.	L-compl.	in AC ¹ , NL-hard	in P, L-hard	
simulation			LogCFL-compl.	P-compl.	

TABLE 1. Parallel complexity of the (bi)simulation problem on restricted classes of finite graphs.

definition of LREG we think that it is not easier to express the bisimulation problem in LREG than giving a direct logspace algorithm. Moreover expressibility in LREG does not imply that bisimulation for trees in term representation is in NC¹.

A conference version of this paper appeared in [GGL16].

2. PRELIMINARIES

2.1. Graphs and trees. A (*directed*) graph $G = (V, E)$ consists of a set of nodes V and a set of edges $E(G) = E \subseteq V \times V$. If $(u, v) \in E$, we call v a *successor* of u . A *path* (of length $n \geq 0$ from v_0 to v_n) in G is a node sequence v_0, v_1, \dots, v_n such that $(v_i, v_{i+1}) \in E$ for all $0 \leq i < n$. A graph is *acyclic* if there is no path of length ≥ 1 from a node to itself. We say that two nodes $u, v \in V$ are *connected* if there exists a path from u to v in the underlying undirected graph $(V, E \cup \{(v, u) \mid (u, v) \in E\})$. A set of nodes $U \subseteq V$ is *connected* if any two nodes in U are connected.

A graph T is a (*rooted*) *tree* if there exists a node $r \in V(T)$, called the *root* of T , such that for all $v \in V(T)$ there exists exactly one path from r to v . The *size* $|u|$ of a node u in a tree is the size of the subtree rooted in u . The *depth* of a node u , denoted by $\text{depth}(u)$, is the length of the unique path from the root to u . A node u is an *ancestor* (resp., *proper ancestor*) of v , briefly $u \preceq v$ (resp., $u \prec v$), if there exists a path (resp., a path of length ≥ 1) in T from u to v . Given a tree T we say that a node v is *between* nodes u and w if either $v \in \{u, w\}$, or u and w are connected in $T \setminus v$ (which is the subgraph of T obtained by removing v).

A *node-labelled graph* (V, E, β) is a graph (V, E) together with a labelling function $\beta: V \rightarrow A$ into a finite set A . An *edge-labelled graph* (V, E) consists of a set of nodes V and a labelled edge relation $E \subseteq V \times A \times V$. We also write $u \xrightarrow{a} v$ instead of $(u, a, v) \in E$. Unlabelled graphs are also regarded as labelled graphs over a singleton label set. An edge-labelled tree is an edge-labelled graph (V, E) where the sets $E_a = \{(u, v) \mid (u, a, v) \in E\}$ are pairwise disjoint for $a \in A$ and $(V, \bigcup_{a \in A} E_a)$ is a rooted tree.

We denote by Graphs, Unlabelled-Graphs, Trees and Unlabelled-Trees the classes of (unlabelled) graphs and (unlabelled) trees, respectively.

2.2. Circuits. A (*Boolean*) *circuit* $C = (G, \beta)$ is a node-labelled graph, where G is acyclic and $\beta: V(G) \rightarrow \{x_1, \dots, x_n, 0, 1, \neg, \wedge, \vee\}$ for some n . Nodes of C are usually called *gates*. Gates labelled by 0, 1 (constant gates) or by a variable x_i (input gates) have no successors. Gates labelled by \wedge or \vee have at least one successor, and gates labelled by \neg have exactly

one successor. A *variable-free* circuit is a circuit without input gates. In a variable-free circuit every gate can be evaluated to either 0 or 1. The question of deciding whether a given gate of a given variable-free circuit evaluates to 1 is known as the *circuit value problem*. It is one of the classical P-complete problems, see [GHR95] for more details.

2.3. Computational complexity. In this paper we will work with the following complexity classes:

$$\text{AC}^0 \subseteq \text{TC}^0 \subseteq \text{NC}^1 \subseteq \text{L} \subseteq \text{NL} \subseteq \text{LogCFL} \subseteq \text{AC}^1 \subseteq \text{NC} \subseteq \text{P}.$$

In the following we briefly define these classes. Of course, P denotes deterministic polynomial time. A function $f: \Sigma^* \rightarrow \Gamma^*$ is *logspace-computable* if it can be computed on a deterministic Turing-machine with a read-only input tape, a write-only output tape and a working tape whose length is bounded logarithmically in the input length; such a machine is also called a logspace transducer. We denote by L the class of languages which can be decided in logspace, i.e., for which the characteristic function is logspace-computable. Throughout the paper we will use implicitly that compositions of logspace-computable functions are logspace-computable again. The class NL is the set of languages that can be decided by nondeterministic Turing machine in logarithmic space. It is closed under complement by the famous theorem of Immerman-Szelepcsényi [Imm88, Sze88].

The complexity class NC^i ($i \geq 0$) contains all languages $L \subseteq \{0, 1\}^*$ such that there exists a circuit family $(C_n)_{n \in \mathbb{N}}$ where

- (1) C_n has the input gates $\{x_1, \dots, x_n, \neg x_1, \dots, \neg x_n\}$, \wedge -gates and \vee -gates of fan-in two, and a distinguished output gate,
- (2) C_n has size $n^{O(1)}$ and depth $O((\log n)^i)$,
- (3) and C_n accepts $x \in \{0, 1\}^n$ if and only if $x \in L$.

If we allow in (1) \wedge -gates and \vee -gates of unbounded fan-in, we obtain the class AC^i . If we allow in (1) \wedge -gates, \vee -gates and majority-gates – all of unbounded fan-in – then we obtain the class TC^i . Finally, the class SAC^i is obtained by taking \wedge -gates of fan-in two and unbounded fan-in \vee -gates in (1). One can easily show that $\text{NC}^i \subseteq \text{SAC}^i \subseteq \text{AC}^i \subseteq \text{TC}^i \subseteq \text{NC}^{i+1}$ for all $i \geq 0$. For the classes above (and inclusively) SAC^1 , we assume logspace uniformity, which means that there is a logspace transducer that computes from the unary encoding of the number n the n -th circuit C_n . For the classes AC^0 , TC^0 and NC^1 logspace uniformity is too weak (the class NC^0 will not be used). For these classes one usually imposes the stronger **DLOGTIME**-uniformity condition on the circuit families. For the constant depth classes AC^0 and TC^0 , **DLOGTIME**-uniformity means that for given binary coded gate numbers u, v of the n -th circuit C_n , one can (i) compute the type of gate u in time $O(\log n)$ and (ii) check in time $O(\log n)$ whether u is an input gate for v . Since C_n has polynomially many gates, the gates of C_n can be encoded by bit strings of length $O(\log n)$. Thus the time bound $O(\log n)$ is linear in the input length $|u| + |v|$. To define **DLOGTIME**-uniformity for NC^1 -circuits one needs the so-called extended connection language. We do not have to go into details (which can be found in [BIS90, Ruz81, Vol99]), since we will not work with uniformity explicitly. The union of all classes NC^i is denoted by NC.

The definitions of the above circuit complexity classes can be easily extended to functions $I: \{0, 1\}^* \rightarrow \{0, 1\}^*$. This can be done by encoding I by the language $L_I = \{1^i 0 w \mid w \in \{0, 1\}^*, \text{ the } i\text{-th bit of } f(w) \text{ is } 1\}$. The main complexity class used in this paper is **DLOGTIME**-uniform TC^0 . It can be seen as the extension of AC^0 by the ability of counting. Typical problems in TC^0 are the computation of the sum, product and integer quotient of

two binary encoded integers, and the sum and product of an arbitrary number of binary encoded integers [HAB02].

The class LogCFL contains all languages $L \subseteq \{0, 1\}^*$ which are logspace-reducible to a fixed context-free language. It coincides with logspace-uniform SAC^1 [Ven91]. Another characterization uses alternating logspace Turing machines with polynomial proof trees. Recall that an alternating Turing machine is a Turing machine where the set of states is partitioned into existential states and universal states. A configuration where the current state is existential (resp., universal) is called a existential (resp., universal) configuration. A existential (resp., universal) configuration is accepting if and only if there exists an accepting successor configuration (resp., if all successor configurations are accepting). In our definition of alternating machines there is no explicit accepting state and to accept the machine has to branch to a universal configuration without successor configurations. It is well-known that P is the class of languages that are accepted by alternating Turing machines in logarithmic space. To get the subclass LogCFL one has to require that there exists a polynomial $p(n)$ such that for every accepted input word w of length n there exists a proof tree of size at most $p(n)$. A proof tree of an alternating Turing machine for input w is a finite tree where the nodes are labelled with configurations such that the following conditions hold:

- The root is labelled with the initial configuration for input w .
- If a node v is labelled with an existential configuration c , then v has exactly one child that is labelled with a successor configuration of c .
- If a node v is labelled with a universal configuration c , then v has exactly one child v' for each successor configuration c' of c , and v' is labelled with c' .

Ruzzo [Ruz80] proved that LogCFL coincides with the class of languages that can be accepted by an alternating Turing machine in logarithmic space and such that for every accepted input there exists a polynomial size proof tree. It was shown by Borodin et al. that LogCFL is closed under complement [BCD⁺89].

For more details on space and circuit complexity we refer to [Vol99].

2.4. Tree representations. The complexity of tree problems often depends on how the trees are represented. Firstly, trees can be given as *pointer structures* where the edge relation is given explicitly as a list of pairs consisting of two node names, which is the standard encoding of graphs in general. Secondly, trees can be given in *term representation* (or *bracket representation*): The string $()$ represents a tree of size 1. If a tree T has a root and direct subtrees T_1, \dots, T_n which have term representations r_1, \dots, r_n , then the string $(r_1 \cdots r_n)$ is a term representation of T . Notice that a tree can have multiple term representations because trees are unordered. Thirdly, trees can be represented in *ancestor representation* where we specify a list of all pairs (u, v) where u is an ancestor of v . Elberfeld et al. showed that term and ancestor representations can be converted into each other in TC^0 [EJT12]. We will utilize this result since operating on ancestor representations of trees is technically easier than on term representations.

To represent node-labelled trees (e.g., tree decompositions), in the pointer and ancestor representation we append a list of pairs consisting of a node name and a node label. In the term representation node labelled trees can be encoded by introducing for each label a an opening bracket symbol $(_a$. For instance, $(_a(b)(_a))$ encodes a tree with an a -labelled root and two children which are labelled by b and a . If the set of node labels is not fixed (this is the case for tree decompositions) we choose an arbitrary binary block code for the opening

brackets $(a$. To represent an edge-labelled tree we transform it into a node-labelled tree, e.g. by assigning the label of an edge (u, v) to its end point v and labelling the root by a special symbol. Let us finally mention that these coding details for labelled trees are only relevant for our NC^1 lower bound in Section 3.1.5, which refers to AC^0 -reductions. The reason is that different codings of labelled trees in term representation can be transformed in TC^0 , but not necessarily in AC^0 , into each other.

2.5. Tree-width and path-width. A *tree decomposition* (T, β) of a directed graph G consists of a tree T and a function $\beta: V(T) \rightarrow 2^{V(G)}$ which assigns to each node of T a so called *bag* such that

- for all $v \in V(G)$ the set $\{t \in V(T) \mid v \in \beta(t)\}$ is non-empty and connected, and
- for all $(u, v) \in E(G)$ there exists $t \in V(T)$ such that $u, v \in \beta(t)$.

The *width* of (T, β) is $\max_{t \in V(T)} |\beta(t)| - 1$ and the *tree-width* of a graph G is the minimum width over all tree decompositions of G . A tree decomposition (T, β) is a *path decomposition* if T is a path. The *path-width* of a graph G is the minimum width of all path decompositions of G . Tree-width and path-width are also defined for node- and edge-labelled graphs via their underlying unlabelled graph. More background on the notions of tree- and path-width can be found for instance in [DF13].

2.6. Bisimulation and simulation.

Definition 2.1. A *bisimulation* on an edge-labelled graph (V, E) is a binary relation $R \subseteq V \times V$ such that for all $(u, v) \in R$ the following conditions hold:

- (1) For all $u \xrightarrow{a} u'$ there exists $v \xrightarrow{a} v'$ such that $(u', v') \in R$.
- (2) For all $v \xrightarrow{a} v'$ there exists $u \xrightarrow{a} u'$ such that $(u', v') \in R$.

A relation R that only satisfies condition 1 for all $(u, v) \in R$ is called a *simulation*. A (bi)simulation on two edge-labelled graphs is a (bi)simulation on their disjoint union. Two nodes u, v are called *bisimilar* if there exists a bisimulation R such that $(u, v) \in R$. We say that u is *simulated by* v if there exists a simulation R such that $(u, v) \in R$.

It is easy to see that the union of all bisimulations on a graph G is a bisimulation again and an equivalence relation, called the *bisimulation equivalence* \sim_G , or simply \sim if G is clear from the context. Similarly, the union of all simulations on a graph is a simulation again and a preorder, called the *simulation preorder* \sqsubseteq_G , or simply \sqsubseteq if G is clear from the context. If u belongs to a graph G and v belongs to a graph $H \neq G$, we write $u \sqsubseteq v$ (resp., $u \sim v$) if $u \sqsubseteq_{G \uplus H} v$ (resp., $u \sim_{G \uplus H} v$), where $G \uplus H$ is the disjoint union of G and H . Note that in general $u \sim v$ is not equivalent to $u \sqsubseteq v \sqsubseteq u$. For two graph classes $\mathcal{C}_1, \mathcal{C}_2$ we define the bisimulation problem and the simulation problem:

$\text{BISIM}(\mathcal{C}_1, \mathcal{C}_2)$

Given: Two graphs $G_1 \in \mathcal{C}_1$, $G_2 \in \mathcal{C}_2$ and two nodes $v_1 \in V(G_1)$, $v_2 \in V(G_2)$

Question: Are v_1 and v_2 bisimilar?

$\text{SIM}(\mathcal{C}_1, \mathcal{C}_2)$

Given: Two graphs $G_1 \in \mathcal{C}_1$, $G_2 \in \mathcal{C}_2$ and two nodes $v_1 \in V(G_1)$, $v_2 \in V(G_2)$

Question: Is v_1 simulated by v_2 ?

Note that $\text{BISIM}(\mathcal{C}_1, \mathcal{C}_2)$ and $\text{BISIM}(\mathcal{C}_2, \mathcal{C}_1)$ are equivalent, since bisimulation is symmetric, but this is not the case for simulation. If $\mathcal{C}_1 = \mathcal{C}_2$ then we just write $\text{BISIM}(\mathcal{C}_1)$ or $\text{SIM}(\mathcal{C}_1)$. It is known that on a finite edge-labelled graph $G = (V, E)$ bisimulation equivalence and simulation preorder can be computed as a fixed point as follows. First we set $u \sim_0 v$ for all $u, v \in V$. Inductively, we define $u \sim_{k+1} v$ if

- for all $u \xrightarrow{a} u'$ there exists $v \xrightarrow{a} v'$ such that $u' \sim_k v'$, and
- for all $v \xrightarrow{a} v'$ there exists $u \xrightarrow{a} u'$ such that $u' \sim_k v'$.

One easily sees that $\sim_i \supseteq \sim_{i+1}$ for all $i \in \mathbb{N}$. Hence there exists $n \in \mathbb{N}$ such that $\sim_n = \sim_{n+i}$ for all $i \in \mathbb{N}$ and \sim_n coincides with bisimulation equivalence \sim . Using the approximants above the bisimulation equivalence can be decided on finite graphs in polynomial time by a partition refinement algorithm [KS90]. The fastest known algorithm is due to Paige and Tarjan, which runs in time $O(m \log n)$ where m is the number of edges and n is the number of nodes [PT87]. The simulation problem on finite graphs is also decidable in polynomial time by a refinement algorithm [CS01]. Similar as above, one defines preorders \sqsubseteq_i such that $\sqsubseteq_n = \sqsubseteq$ for some $n \in \mathbb{N}$. It is known that both problems are P-hard [BGS92, SJ05, Srb01], and therefore:

Theorem 2.2 [BGS92, SJ05]. *$\text{BISIM}(\text{Graphs})$ and $\text{SIM}(\text{Graphs})$ are P-complete. For both problems P-hardness already holds for dags.*

3. COMPLEXITY OF BISIMULATION

3.1. Bisimulation on trees. In this section we consider the bisimulation problem on edge-labelled trees. Here we have two versions, depending on whether the trees are given as pointer structures or in term representation. If the trees are given in term representation then the nodes can be specified by numbers which refer to an opening bracket in the term. Clearly, we can restrict ourselves to the problem of deciding whether the roots of the two input trees are bisimilar, since one can compute subtrees rooted in a given node in the pointer (term) representation in logspace (TC^0).

Theorem 3.1. *$\text{BISIM}(\text{Trees})$ is in L if the trees are given as pointer structures.*

Theorem 3.2. *$\text{BISIM}(\text{Trees})$ is in NC^1 if the trees are given in term representation.*

We remark that the same complexity bounds hold for the tree isomorphism problem [Bus97, JKMT03, Lin92]. In fact, converting a tree given as a pointer structure into term representation is FL-complete [BM95], where FL is the class of functions computable in deterministic logspace. In that sense, with the proper abstraction of the input representation the actual complexity of $\text{BISIM}(\text{Trees})$ is NC^1 .

3.1.1. Tree-shaped Boolean circuits. Bisimilarity between two edge-labelled trees T_1 and T_2 can be expressed as a Boolean circuit $C_{\sim}(T_1, T_2)$: For all $u \in V(T_1), v \in V(T_2)$ such that $\text{depth}(u) = \text{depth}(v)$ the circuit contains a gate $x_{u,v}$, which evaluates to true if and only if u is bisimilar to v . Since we want to test whether the roots of T_1 and T_2 are bisimilar, it suffices to consider the reachable gates $x_{u,v}$ where $\text{depth}(u) = \text{depth}(v)$. We define

$$x_{u,v} = \bigwedge_{u \xrightarrow{a} u' \ v \xrightarrow{a} v'} \bigvee x_{u',v'} \wedge \bigwedge_{v \xrightarrow{a} v' \ u \xrightarrow{a} u'} \bigvee x_{u',v'}. \quad (3.1)$$

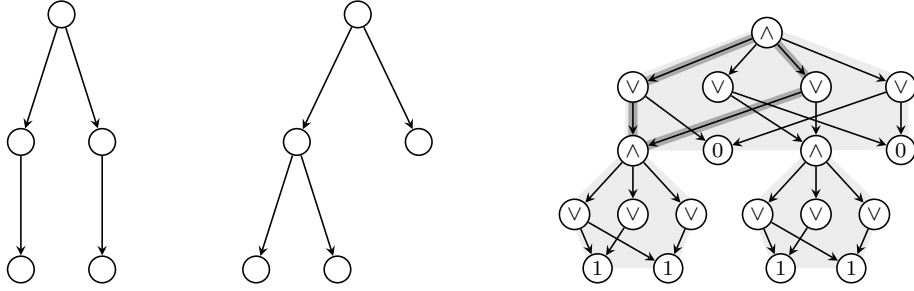


FIGURE 1. Two trees T_1, T_2 and the tree-shaped circuit $C_{\sim}(T_1, T_2)$ for bisimulation equivalence.

As usual we regard an empty conjunction (resp., disjunction) as 1 (resp., 0). In particular, if both u and v are leaves, then $x_{u,v} = 1$, and if exactly one of u and v is a leaf, then $x_{u,v} = 0$. An example circuit is shown in Figure 1, where T_1 and T_2 are unlabelled. Note that the circuit is composed in a tree-shaped form from smaller circuits. These smaller circuits correspond to the definition in (3.1) and have the crucial property that there exist exactly two paths from the root (of the smaller circuit) to an arbitrary leaf, which are highlighted in one subcircuit in Figure 1. This is the case because each gate variable $x_{u',v'}$ occurs once in the first conjunction and once in the second conjunction in (3.1). In fact, we will show that circuits with such a path property can be evaluated in logspace, and even in NC^1 under a suitable representation.

We use a more syntactic definition of such circuits: A *tree-shaped circuit* is a sequence of Boolean equations $\mathcal{S} = (x_i = \varphi_i)_{1 \leq i \leq n}$ where $\varphi_1, \dots, \varphi_n$ are Boolean formulas over the variables x_1, \dots, x_n such that the graph

$$T_{\mathcal{S}} = (\{x_1, \dots, x_n\}, \{(x_i, x_j) \mid x_j \text{ occurs in } \varphi_i\}) \quad (3.2)$$

is a tree with root x_1 . This implies that there are no cyclic definitions in \mathcal{S} and that no variable x_k appears in two distinct formulas φ_i and φ_j ($i \neq j$). The *size* of \mathcal{S} is defined as the sum of the sizes of all formulas φ_i . The *width* of \mathcal{S} is defined as the maximal number of occurrences of a variable x_j in a formula φ_i . An example of a tree-shaped circuit of width two is given by the formulas (3.1) for bisimulation. We can view \mathcal{S} as an ordinary Boolean circuit by taking the disjoint union of the formula trees of the φ_i and then merging all x_i -labelled leaves with the root of the formula tree of φ_i . For example, the tree-shaped circuit $C_{\sim}(T_1, T_2)$ for bisimulation equivalence can be regarded as a tree-shaped circuit of width 2, which can be computed in logspace from T_1 and T_2 . The main goal of this section is to show that the circuit value problem restricted to tree-shaped circuits of bounded width belongs to logspace.

Note that a variable-free Boolean circuit can be represented by a relational structure with a binary edge relation and unary relations for the labels 0, 1, \neg , \wedge , \vee . Moreover, there is a formula of monadic-second order logic (MSO) expressing that a variable-free Boolean circuit evaluates to 1. As a consequence, by Theorem 1.1, we can evaluate variable-free Boolean circuits of tree-width at most k in logspace for every fixed k . Moreover, the complexity can be improved to NC^1 if we also provide for the input circuit a bounded width tree decomposition in term representation. For this, one stores a tree decomposition (T, β) by an expression for the node-labelled tree T , where every node $t \in V(T)$ is labelled by the

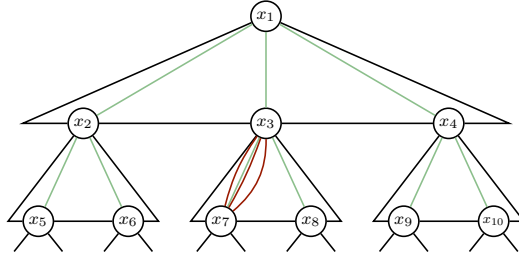


FIGURE 2. A tree-shaped circuit \mathcal{S} where the underlying tree $T_{\mathcal{S}}$ is drawn in green.

bag $\beta(t)$. Note that $\beta(t)$ is a set of gates of the circuit, and these gates are stored by their addresses. To sum up, we have:

Theorem 3.3 [EJT10, EJT12]. *For every fixed $k \in \mathbb{N}$, the circuit value problem restricted to circuits of tree-width at most k can be solved in logspace. If in addition to the input circuit C a width- k tree decomposition of C in term representation is given, then the circuit value problem can be solved in NC^1 .*

3.1.2. *(Un)related restricted circuit classes.* We cannot directly apply Theorem 3.3 to tree-shaped circuits since neither their tree-width nor clique-width is bounded by the following lemma. Clique-width is another graph measure defined by so called k -expressions, see e.g. [KLM09] for a survey. It is easy to see that Theorem 3.3 can be generalized to circuits of clique-width k , provided a k -expression for the circuit is part of the input.

Lemma 3.4. *For $n \geq 1$, let T_n be the tree of size $n+1$ whose root has exactly n children. The set of circuits $\{C_{\sim}(T_n, T_n) \mid n \geq 1\}$ has unbounded tree-width and unbounded clique-width.*

Proof. The underlying undirected graph of $C_{\sim}(T_n, T_n)$ contains the complete bipartite graph $K_{n,n}$ as a (topological) minor. This can be seen by removing the output gate and dissolving all constant gates (*dissolving* a node of degree 2 means deleting it and connecting its two neighbors). Figure 3 shows an example for $n = 3$. Since the tree-width of $K_{n,n}$ is n and the tree-width of every minor of a graph G is bounded by the tree-width of G , it follows that the tree-width of $C_{\sim}(T_n, T_n)$ is at least n .

It is known that a set of graphs which has bounded clique-width and for which there exist only finitely many n such that the bipartite graph $K_{n,n}$ is contained as a subgraph (not only minor) also has bounded tree-width [GW00]. We claim that for each $n \geq 1$ the undirected graph of $C_{\sim}(T_n, T_n)$ does not contain $K_{3,3}$ as a subgraph, which implies that $\{C_{\sim}(T_n, T_n) \mid n \geq 1\}$ also has unbounded clique-width (note that $C_{\sim}(T_n, T_n)$ contains a $K_{2,2}$, i.e., a cycle on four nodes). Note that in $C_{\sim}(T_n, T_n)$ for every simple path of four nodes, one of the four nodes has degree 2 (these are the gates in the middle layer of Figure 3). But this is not possible in a $K_{3,3}$. \square

Let us also compare tree-shaped circuits of constant width with a class of circuits that is presented in [GGHL12, page 5]. For the latter, the authors require that for every path in the circuit the product of the fan-outs (i.e., indegrees, since we direct circuits towards the input gates) of the gates on the path is bounded polynomially by the circuit size. It is shown that circuits with this property can be evaluated in logspace. Note that tree-shaped

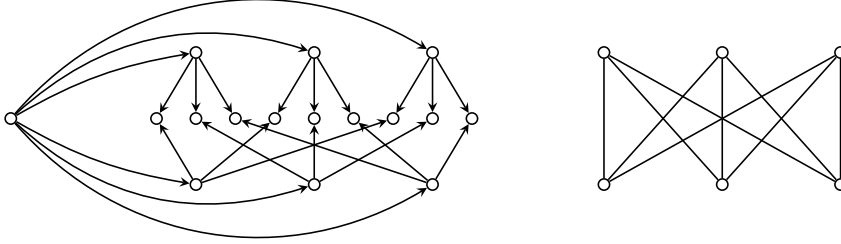


FIGURE 3. The circuit $C_{\sim}(T_3, T_3)$ for bisimulation equivalence. Removing the output gate on the left and dissolving the constant gates in the middle layer shows that $K_{3,3}$ is a minor of $C_{\sim}(T_3, T_3)$.

circuits of constant width do not have this path property from [GGHL12]. On the other hand, the circuits from [GGHL12] can have nodes with large fan-out, which is not possible for tree-shaped circuits of constant width. Hence, the two circuit classes are incomparable.

3.1.3. Evaluating tree-shaped circuits given in pointer representation. Before we show how to evaluate tree-shaped circuits of bounded width in logspace, let us briefly recall the standard logspace algorithm which performs a depth-first traversal of a tree T . It stores (i) a pointer to the current node of T using $O(\log |T|)$ bits and (ii) a single bit, indicating the current direction of the traversal (up or down). We will call this procedure *logspace depth-first traversal*. If T is a formula then one can evaluate the formula using the logspace algorithm above by additionally storing the value of the current subformula.

Next we introduce some notions. Recall that the size of the subtree rooted in a node u is denoted by $|u|$. We say that a node u is *heavy* if for all siblings v of u we either have (i) $|u| > |v|$ or (ii) $|u| = |v|$ and $u < v$ (where $<$ denotes some fixed order on the nodes of the tree). Otherwise a node is called *light*. Notice that the root is heavy and that every inner node has exactly one heavy child. Using the logspace depth-first traversal algorithm we can compute in logspace for each node its size and determine whether it is heavy. Note that every path in a tree contains at most $O(\log n)$ light nodes.

By the following result, tree-shaped circuits of bounded width can be evaluated in logspace (take $m = O(1)$). Also note that a tree-shaped circuit of width $m = 1$ is a tree and therefore can be evaluated in logspace.

Theorem 3.5. *A given tree-shaped circuit $\mathcal{S} = (x_i = \varphi_i)_{1 \leq i \leq n}$ of width $m \geq 2$ can be evaluated in space $O(\log s + \log n \cdot \log m)$, where $s = \max\{|\varphi_i| \mid 1 \leq i \leq n\}$ is the maximal size of one of the formulas φ_i . In particular, for every fixed $m \in \mathbb{N}$, there is a logspace algorithm which evaluates a given tree-shaped circuit of width at most m .*

Proof. Let $\mathcal{S} = (x_i = \varphi_i)_{1 \leq i \leq n}$ be a tree-shaped circuit of width m and let $s = \max\{|\varphi_i| \mid 1 \leq i \leq n\}$. Recall the definition of the tree $T_{\mathcal{S}}$ with node set $\{x_1, \dots, x_n\}$ from (3.2).

First of all, we label every node x_i of $T_{\mathcal{S}}$ by (i) the size $|x_i|$ of the subtree rooted in x_i and (ii) a single bit indicating whether x_i is a heavy node of $T_{\mathcal{S}}$. This information can be computed in space $O(\log n)$ using a logspace depth-first traversal.

The circuit \mathcal{S} is evaluated in a recursive way using a pointer to one of the nodes x_1, \dots, x_n (which needs space $O(\log n)$) and a stack of height $O(\log n \cdot \log m)$ as follows. Initially the

pointer is set to the root x_1 . Assume that x_k is the heavy child of x_1 in T_S . Then, the pointer is moved to x_k without writing anything on the stack. Next, the subcircuit rooted at x_k is evaluated recursively. By induction, space $O(\log s + \log n \cdot \log m)$ is used for this. Once the algorithm returns from the recursion the pointer is back on x_k and one can release the space. The value of x_k is stored on the stack, which now contains a single bit. Note that the algorithm knows (using the labeling computed before) that x_k is the heavy child of its parent node x_1 . This information now triggers the evaluation of the Boolean formula φ_1 . This is done by a logspace depth-first traversal of the Boolean formula tree of φ_1 . Each time, this depth-first traversal of φ_1 arrives at a leaf node, the following is done:

- If the leaf is labelled with the variable x_k (the heavy child of x_1 in T_S), then the value of x_k is retrieved from the stack and the algorithm continues the evaluation of φ_1 .
- If the leaf is a light child labelled with the variable $x_i \neq x_k$, then let $1 \leq c \leq m$ such that the leaf corresponds to the c -th occurrence of x_i in φ_1 . The algorithm stores c on the stack (which needs space $O(\log m)$) and continues recursively with the evaluation of the subcircuit rooted at x_i . Once it comes back from the recursion, the pointer is back on x_i . The algorithm sees that x_i is a light child of x_1 . Using this information and the number c stored on the stack, it continues the evaluation of φ_1 at the right position in φ_1 .

Note that in the second case, we have $|x_i| \leq |x_1|/2$. This implies that the number of bits stored on the stack is bounded by $O(\log n \cdot \log m)$. The total space consumption is therefore $O(\log n + \log s + \log n \cdot \log m) = O(\log s + \log n \cdot \log m)$ (since we assume $m \geq 2$). \square

Note that we do not assume m to be a constant in Theorem 3.5. In particular, since k , n , and m are all bounded by the size of the tree-shaped circuit (which is the sum of the sizes of the formulas φ_i), it follows that a tree-shaped circuit of size N can be evaluated in space $O(\log^2 N)$. For the special case $m = O(1)$ (which is used for the bisimulation problem) we give an alternative proof below. This proof prepares our handling of trees in term representation in the next section (proof of Theorem 3.6).

Alternative proof of Theorem 3.5 for constant width. Let N be the size of \mathcal{S} and $m \in O(1)$ its width. We will construct from \mathcal{S} in logspace an equivalent polynomially sized circuit with constant tree-width, which can be seen as a partial tree unfolding of the circuit corresponding to \mathcal{S} . By Theorem 3.3 the resulting circuit can be evaluated in logspace.

As before, we view the Boolean formulas φ_i as node-labelled trees. For simplicity we assume that all φ_i have at least size two, which can be ensured by replacing φ_i by $\varphi_i \wedge 1$, so that \mathcal{S} contains no ‘‘chain rules’’. Moreover, we assume that the trees φ_i have disjoint node sets. Let x_i be an inner node of T_S whose heavy child is x_k . Inductively we define a circuit C_i as follows: We take the formula φ_i , viewed as a tree, and merge all x_k -labelled leaves into a single node. Note that x_j -labelled leaves for $j \neq k$ are not merged. Then we insert into each leaf labelled by some variable x_j a copy of the circuit C_j . Finally let C be C_1 , which clearly evaluates to the same truth value as \mathcal{S} . Figure 4 shows the circuit resulting from the circuit $C \sim (T_1, T_2)$ on the right in Figure 1.

Note that the number of copies of C_i in C is bounded by m^{ℓ_i} where ℓ_i is the number of light nodes on the path from x_1 to x_i in T_S . Since $\ell_i \leq \log n$, C has size at most $m^{\log n} \cdot N$, which is bounded by $n^{O(1)} \cdot N$ since m is a constant.

Furthermore C can be computed in logspace from \mathcal{S} . To make this explicit, we introduce a naming scheme for the gates in C . The set $\text{Addr}(x_i)$ (addresses for the copies of x_i) contains finite words over the alphabet $\{1, \dots, m\}$ defined inductively: We set $\text{Addr}(x_1) = \{\varepsilon\}$ for the root x_1 . If x_j is the heavy child of x_i in T_S , we set $\text{Addr}(x_j) = \text{Addr}(x_i)$. If x_j is a light

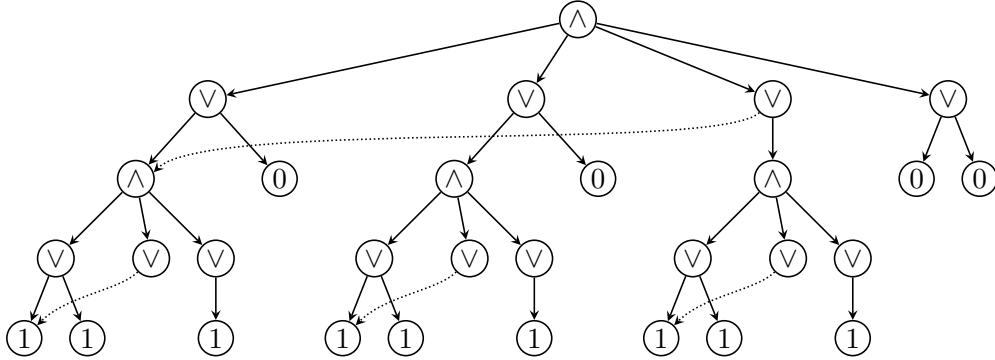


FIGURE 4. The partial unfolding of the tree-shaped circuit $C_{\sim}(T_1, T_2)$ from Figure 1.

child of x_i in $T_{\mathcal{S}}$, we set $\text{Addr}(x_j) = \text{Addr}(x_i) \cdot \{1, \dots, k\}$ where $k \leq m$ is the number of occurrences of x_j in φ_i . The sets $\text{Addr}(x_i)$ contain words of length $O(\log n)$ over the constant sized alphabet $\{1, \dots, m\}$. Moreover, given a word of length $O(\log n)$ over $\{1, \dots, m\}$ we can easily check in logspace whether it belongs to $\text{Addr}(x_i)$ by traversing the path from x_i to the root of $T_{\mathcal{S}}$. Now we can define the circuit C over the gate set

$$V(C) = \bigcup_{i=1}^n \{(u, a) \mid a \in \text{Addr}(x_i), u \text{ is a non-input node of the tree } \varphi_i\}.$$

For every edge (u, u') of φ_i ($1 \leq i \leq n$) and $a \in \text{Addr}(x_i)$, we add the following edges to $E(C)$:

- (1) If u' is a non-input gate, add the edge $((u, a), (u', a))$.
- (2) If u' is labelled by x_j , let v be the root of φ_j and add the edge
 - (a) $((u, a), (v, a \cdot d))$ if x_j is a light child of x_i and u' is its d -th occurrence of x_j in φ_i ,
 - (b) $((u, a), (v, a))$ if x_j is the heavy child of x_i .

The labels of the gates in C are inherited from the formula trees φ_i . Note that a pair (u, a) from $V(C)$ can be stored in logspace. Moreover, whether a pair belongs to $V(C)$ and whether a pair of nodes from $V(C)$ belongs to $E(C)$ can be checked in logspace. Hence, the circuit C can be constructed in logspace.

Finally we show that the tree-width of C is at most 2. Let us call an edge in C *crossing* if it is of type 2b in the above definition of C and u' is the d -th occurrence of x_j in φ_i for some $d > 1$. In Figure 4 the crossing edges are drawn as dotted lines. Notice that for each address a there is at most one crossing edge $((u, a), (v, a)) \in E(C)$. Consider the subgraph T of C obtained by removing all crossing edges. The resulting subgraph T is indeed a tree, on which we define a tree decomposition (T, β) of C . First let (T, β') be the “canonical” tree decomposition of T of width 1, i.e. $\beta'(u, a)$ contains (u, a) and its parent node in T (if existent). To obtain β from β' we add for each address a the crossing edge $((u, a), (v, a))$ (if existent) to all bags $\beta(x, a)$. One can verify that (T, β) is indeed a valid tree decomposition, which has width at most 2. \square

Theorem 3.5 applied to the tree-shaped circuit defined by (3.1) immediately yields Theorem 3.1 for the problem $\text{BISIM}(\text{Trees})$ when trees are given in pointer representation.

3.1.4. *Evaluating tree-shaped circuits given in term representation.* In this section we will prove Theorem 3.2 for the problem $\text{BISIM}(\text{Trees})$ when trees are given in term representation. For that we prove an NC^1 -version of Theorem 3.5 for the case that m is constant, which uses the NC^1 -part of Theorem 3.3, where a tree decomposition in term representation is part of the input. Here we require that the tree-shaped circuit $\mathcal{S} = (x_i = \varphi_i)_{1 \leq i \leq n}$ must be given together with the underlying tree $T_{\mathcal{S}}$ in term representation. We also assume that the Boolean formulas φ_i are given in term representation. Recall the ancestor representation of a tree from Section 2 and that it can be transformed into the term representation of a tree in TC^0 and vice versa.

Theorem 3.6. *For every fixed $m \in \mathbb{N}$, one can evaluate in NC^1 a given tree-shaped circuit $\mathcal{S} = (x_i = \varphi_i)_{1 \leq i \leq n}$ (with all φ_i given in term representation) of width at most m that is given together with the term representation of the tree $T_{\mathcal{S}}$.*

Proof. Let $\mathcal{S} = (x_i = \varphi_i)_{1 \leq i \leq n}$ be a tree-shaped circuit of width at most m , where every φ_i is given in term representation. Moreover, we assume to have the term (or ancestor) representation of the tree $T_{\mathcal{S}}$. We will show how to compute in TC^0 the partial unfolding C of \mathcal{S} and the tree decomposition of C in ancestor representation from the alternative proof of Theorem 3.5 on page 11.

Let N be the size of \mathcal{S} . For each node x_i of $T_{\mathcal{S}}$ we can compute in TC^0 its depth (by counting ancestors) and the size of the subtree below x_i (by counting descendants). Hence, we can also compute the heavy child of every inner node x_i in $T_{\mathcal{S}}$. Additionally, if x_i has a parent node x_j , we can compute the number $d(x_i) \in \{1, \dots, m\}$ of occurrences of x_i in φ_j . We transform all formulas φ_i of \mathcal{S} into ancestor representation in TC^0 where we can assume that the encodings of all nodes have length $O(\log N)$. Also we can ensure that all formulas have at least size two.

Recall that every node u of a formula φ_i has multiple copies (u, a) in C where $a \in \text{Addr}(x_i)$ is an address of length $O(\log n)$ defined on page 11. Given a string $a \in \{1, \dots, m\}^*$ of length $O(\log n)$ and a node x_i we can verify in TC^0 whether $a \in \text{Addr}(x_i)$: From the ancestor representation of $T_{\mathcal{S}}$ we can compute the sequence of all light nodes $x_{i_1} \prec x_{i_2} \prec \dots \prec x_{i_k}$ in $T_{\mathcal{S}}$ on the path from the root to x_i . This can be done by sorting all light ancestors of x_i by their depth in ascending order. It is known that sorting n numbers with n bits each is in TC^0 [Vol99]. Then we have $a_1 \dots a_k \in \text{Addr}(x_i)$ if and only if $a_j \leq d(x_{i_j})$ for all $j \in \{1, \dots, k\}$. Hence, we can also encode all gates of the partial unfolding C by strings of length $O(\log N)$ and can compute $V(C)$ in TC^0 . With the previous preparation the edge relation $E(C)$ can be computed in AC^0 using the definition on page 12.

It remains to show that we can compute in TC^0 the ancestor representation of the width-2 tree decomposition (T, β) of C from page 12. We set $V(T) = V(C)$. Let $(u, a), (u', a') \in V(T)$ be nodes where u (resp., u') belongs to φ_i (resp., φ_j). Then (u, a) is an ancestor of (u', a') in T if and only if $x_i \preceq x_j$ in $T_{\mathcal{S}}$, a is a prefix of a' and the following holds:

- (1) If $x_i = x_j$, then $a = a'$ and $u \preceq u'$ in φ_i .
- (2) If $x_i \prec x_j$ in $T_{\mathcal{S}}$, let x_k be the unique child of x_i which is an ancestor of x_j .
 - (a) If x_k is a light node, let $d \in \{1, \dots, m\}$ be the number in a' at position $|a| + 1$. Then the d -th occurrence of x_k in φ_i is a descendant of u .
 - (b) If x_k is a heavy node, then the first occurrence of x_k in φ_i is a descendant of u .

The last condition forbids those edges that were deleted when constructing the tree T from the partial unfolding C (the dotted edges in Figure 4). The bag-function β can be computed

straightforwardly in TC^0 using its definition on page 12. By Theorem 3.3 we can evaluate C in NC^1 , which concludes the proof. \square

Finally we can apply Theorem 3.6 to prove Theorem 3.2:

Proof of Theorem 3.2. First we convert the term representations of T_1 and T_2 into ancestor representations in TC^0 . We can compute the tree-shaped circuit \mathcal{S} corresponding to the circuit $C_{\sim}(T_1, T_2)$ and an ancestor representation of $T_{\mathcal{S}}$ in TC^0 . The set of variables

$$\{x_{u,v} \mid u \in V(T_1), v \in V(T_2), \text{depth}(u) = \text{depth}(v)\}$$

can be clearly computed in TC^0 since for a given node of a tree in ancestor representation one can compute its depth by counting ancestors. The term representations of the formulas $\varphi_{u,v}$ in (3.1) can then be computed in AC^0 . The ancestor representation of $T_{\mathcal{S}}$ is also AC^0 -computable, since $x_{u,v}$ is an ancestor of $x_{u',v'}$ if and only if u is an ancestor of u' and v is an ancestor of v' . By Theorem 3.6 we can evaluate \mathcal{S} in NC^1 . \square

3.1.5. *Lower Bounds.* In this section we prove matching lower bounds for the upper bounds from Theorem 3.1 and 3.2.

Theorem 3.7. *BISIM(Unlabelled-Trees) is L-hard if the trees are given as pointer structures and NC^1 -hard if they are given in term representation (both with respect to many-one AC^0 -reductions).*

Before we prove Theorem 3.7 let us first show the following lemma:

Lemma 3.8. *BISIM(Trees) is many-one AC^0 -reducible to BISIM(Unlabelled-Trees) in both term and pointer representation.*

Proof. We only show the lemma for the term representation; the same construction also works for the pointer representation. In [Srb01] Srba presents a reduction from the bisimulation problem for edge-labelled graphs to the bisimulation problem for unlabelled graphs. In fact, this construction transforms trees into trees. We slightly modify the reduction to ensure AC^0 -computability and assume that there are only two labels, say a and b (which is the case for the trees constructed in the proof of Theorem 3.7).

Consider a tree T with edge labels a and b . First every labelled edge of T is subdivided into two edges. In Figure 5 (middle tree), the new node added for an x -labelled edge ($x \in \{a, b\}$) is labelled with x . To distinguish the original nodes from the new nodes, we attach to each original node two paths of length 3. To each new node we attach one of two small trees depending on the label of the original edge that is represented by the new node, see the right tree in Figure 5. Let us denote the resulting unlabelled tree with $\text{ul}(T)$. It is not hard to prove that two labelled trees T_1 and T_2 are bisimilar if and only if $\text{ul}(T_1)$ and $\text{ul}(T_2)$ are bisimilar. The proof is basically given in [Srb01].

It remains to prove that the term representation of $\text{ul}(T)$ can be computed in AC^0 from the term representation of T . Consider the term representation t of T . Recall that we identify the opening brackets in t with the nodes of T . We assume that the term representation t contains the following three opening bracket types: $(_a$ and $(_b$, which represent nodes with an incoming a -labelled (resp., b -labelled) edge, and $($ for the root. For example, $t = ((_a) (_b))$ is a term representation for the left tree in Figure 5. The transformation $T \mapsto \text{ul}(T)$ can be described by two isometric homomorphisms. A homomorphism $h: \Sigma^* \rightarrow \Gamma^*$ is *isometric* if

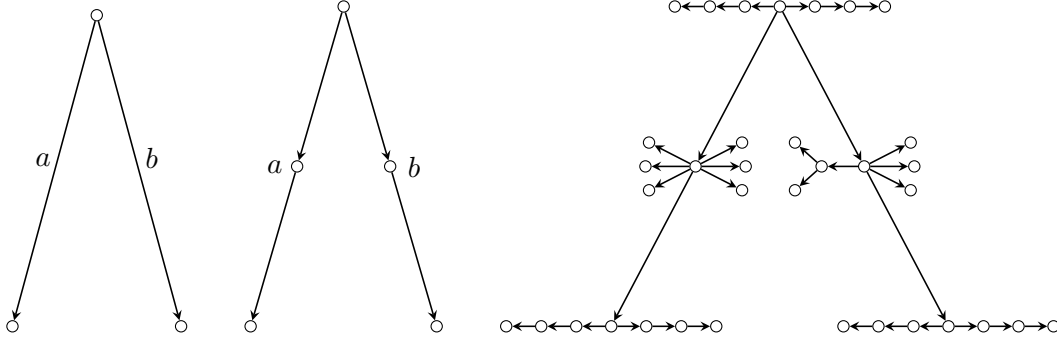


FIGURE 5. From labelled to unlabelled trees

there is an $\ell \geq 1$ such that $|h(c)| = \ell$ for all $c \in \Sigma$. In [LM98] it is shown that for a given isometric homomorphism $h: \Sigma^* \rightarrow \Gamma^*$ and a word $w \in \Sigma^*$ one can compute $h(w)$ in AC^0 .¹

We will proceed in two steps. Define the isometric homomorphism $h_1: \{(a, (b, \cdot))\}^* \rightarrow \{(a, (b, (,),])\}^*$ by:

$$(a \mapsto (a(\quad (b \mapsto (b(\quad) \mapsto))$$

Let $u \in \{(a, (b, \cdot))\}^*$ be the word such that $t = (u)$ and consider the string $(h_1(u))$. Formally, it is not a term representation (since we have two types of closing brackets). Nevertheless, it describes the tree obtained from T by subdividing every edge and labelling each new node with the former edge label. For example $t = ((a)(b))$ is transformed into $(h_1(u)) = ((a()] (b()])$, which describes the node-labelled tree in Figure 5. The second isometric homomorphism $h_2: \{(a, (b, (,),])\}^* \rightarrow \{(,)\}^*$ is defined by:

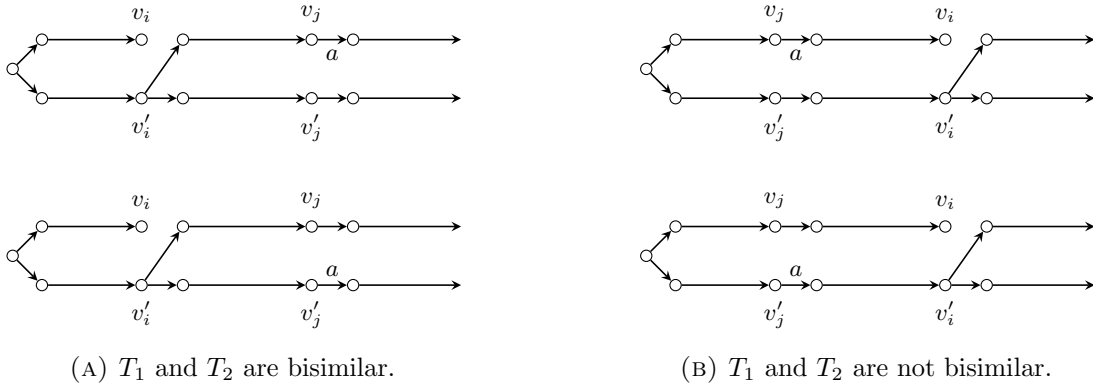
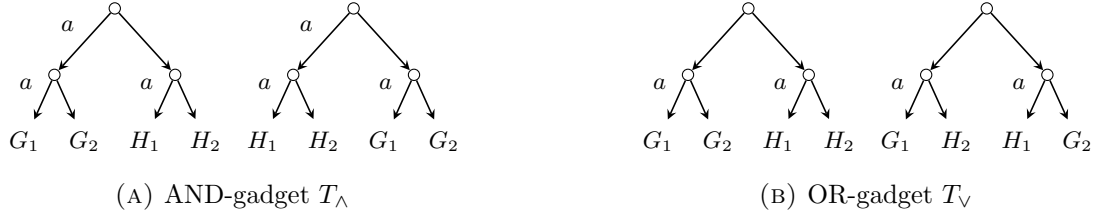
- $(\mapsto ((()))$ (an opening bracket followed by a path of length 3)
- $) \mapsto ((()))$ (a path of length 3 followed by a closing bracket)
- $] \mapsto () ()$ (3 leaves followed by a closing bracket)
- $(a \mapsto (() ()$ (an opening bracket followed by 3 leaves)
- $(b \mapsto ((() ()$ (an opening bracket followed by the tree $\delta^{\circledast} b$)

Then, the string $h_2((h_1(u)))$ is indeed a term representation for the desired unlabelled tree $\text{ul}(T)$. □

Proof of Theorem 3.7. By Lemma 3.8 it suffices to show the lower bounds for edge-labelled trees. We reuse the proofs from [JKMT03], where it is shown that the tree isomorphism problem is L -hard (NC^1 -hard, respectively) with respect to AC^0 -reductions if the trees are given as pointer structures (in term representation, respectively). Let us start with the bisimulation problem for trees given in pointer representation. Here, Jenner et al. reduce from the L -complete reachability problem on paths, i.e., the question whether for a given directed path graph G and two nodes $v_i, v_j \in V(G)$, there is a path from v_i to v_j [Ete97]. Without loss of generality, v_i and v_j are distinct and have successors v_{i+1} and v_{j+1} , respectively.

Consider the tree with a root node which has two copies of G as direct subtrees. We refer to nodes of the two copies by v_1, \dots, v_n and v'_1, \dots, v'_n , respectively. Additionally we replace the edge (v_i, v_{i+1}) by the new edge (v'_i, v_{i+1}) . Now let T_1 (resp., T_2) be the tree

¹If the homomorphism is fixed, this is even possible in NC^0 . Moreover, if the homomorphism is not isometric then the problem is TC^0 -complete [LM98].

(A) T_1 and T_2 are bisimilar.(B) T_1 and T_2 are not bisimilar.FIGURE 6. The two possible forms of T_1, T_2 depending on whether v_j is reachable from v_i or not(A) AND-gadget T_\wedge (B) OR-gadget T_\vee FIGURE 7. The trees for the NC^1 lower bound

where the edge (v_j, v_{j+1}) (resp., (v'_j, v'_{j+1})) is labelled by a symbol a (all unlabelled edges are assumed to be labelled with a symbol $b \neq a$). Clearly, T_1 and T_2 can be computed in AC^0 from G . There is a path from v_i to v_j in G if and only if T_1 and T_2 are bisimilar. See Figure 6 for an illustration of the reduction.

Secondly, Jenner et al. present in [JKMT03] an AC^0 -reduction from the NC^1 -complete evaluation problem of balanced Boolean expressions to the isomorphism problem for trees in term representation. They use the AND-gadget $T_\wedge(G_1, G_2, H_1, H_2)$ and the OR-gadget $T_\vee(G_1, G_2, H_1, H_2)$ which are depicted in Figure 7. Notice that for all trees G_1, G_2, H_1, H_2 the following holds, where \sim can both mean bisimilarity and isomorphism:

- $G_1 \sim H_1$ and $G_2 \sim H_2 \iff T_\wedge(G_1, G_2, H_1, H_2) \sim T_\wedge(H_1, H_2, G_1, G_2)$, and
- $G_1 \sim H_1$ or $G_2 \sim H_2 \iff T_\vee(G_1, G_2, H_1, H_2) \sim T_\vee(G_1, H_2, H_1, G_2)$.

Using the fact that the AND-gadget and the OR-gadget have the same tree structure (if we ignore labels), one can show that the term representations for the resulting trees can be computed in AC^0 from the balanced Boolean expression; see the arguments in [JKMT03]. This yields an AC^0 -reduction from the evaluation problem of balanced Boolean expressions to the bisimulation problem for trees in term representation. \square

3.2. Bisimulation between trees and general graphs. As remarked in Section 2.6, the bisimulation problem is \mathbf{P} -complete for the class of all finite graphs. For trees we have shown that the complexity reduces to logspace, respectively NC^1 , depending on the representation of the input trees. This leaves the question for the complexity of the bisimulation problem,

when one of the input graphs is a tree, and the other graph is not restricted. In this section we show that this problem belongs to AC^1 and that it is NL -hard, but we are not able to prove matching bounds. Let us start with the lower bound.

Theorem 3.9. $\text{BISIM}(\text{Unlabelled-Trees}, \text{Unlabelled-Graphs})$ is NL -hard.

Proof. We reduce from the following NL -complete problem: Given a graph $G = (V, E)$, two vertices $s, t \in V$ and a number $k \in \mathbb{N}$, does there exist a path of length exactly k from s to t ?

Let H be the dag defined on $V(H) = V \times \{1, \dots, k+1\} \cup \{*\}$ with the edge set

$$E(H) = \{((u, i), (v, i+1)) \mid (u, v) \in E, 1 \leq i \leq k\} \cup \{((t, k+1), *)\}.$$

Notice that t is reachable from s in exactly k steps if and only if the node $(s, 1)$ is not bisimilar to a path of length k . Clearly, H and the path graph can be constructed in logspace from G . The result follows, since NL is closed under complement. \square

In the rest of the section, we show that $\text{BISIM}(\text{Trees}, \text{Graphs})$ belongs to AC^1 . Using the reduction in [Srb01] (see also the proof of Lemma 3.8) we can assume that the input tree and graph are unlabelled. For a number $d \in \mathbb{N}$ we write $u \xrightarrow{d} u'$ if there exists a path of length d from u to u' . Let $G = (V, E)$ be a finite unlabelled directed graph, on which we define approximants \approx_k of the bisimulation equivalence \sim_G . First we set $u \approx_0 v$ for all $u, v \in V$. Inductively, we define $u \approx_{k+1} v$ if for all $d \in \mathbb{N}$ we have

- for all $u \xrightarrow{d} u'$ there exists $v \xrightarrow{d} v'$ such that $u' \approx_k v'$, and
- for all $v \xrightarrow{d} v'$ there exists $u \xrightarrow{d} u'$ such that $u' \approx_k v'$.

Notice that $\approx_i \supseteq \approx_{i+1}$ for all $i \in \mathbb{N}$. Hence there exists $n \in \mathbb{N}$ such that $\approx_n = \approx_{n+i}$ for all $i \in \mathbb{N}$ and \approx_n coincides with the bisimulation equivalence \sim . Before considering the bisimulation problem between trees and graphs, we show that the approximants \approx_k converge after $O(\log n)$ rounds on trees of size n . For nodes $u, v \in V$ we define

$$r(u, v) = \sup\{k \in \mathbb{N} \mid u \approx_k v\} \in \mathbb{N} \cup \{\infty\},$$

and define $r(G) = \sup\{r(u, v) \mid u \not\sim v \text{ for } u, v \in V\} \in \mathbb{N} \cup \{-\infty\}$. Notice that $r(G) = -\infty$ if and only if all nodes in G are bisimilar. The following proposition shows that $r(T) \leq 2 \log_2(n) + 2$ for every unlabelled tree T of size n .

Proposition 3.10. *Let $T = (V, E)$ be an unlabelled tree of size n . If $u_0, v_0 \in V$ are not bisimilar, then $r(u_0, v_0) \leq 2 \log_2(|u_0| + |v_0|)$.*

Proof. We prove the lemma by induction on $k = r(u_0, v_0)$. The cases $k = 0$ and $k = 1$ are clear since $2 \log_2(|u_0| + |v_0|) \geq 2$. Now assume that $k \geq 2$. Since $u_0 \approx_k v_0$, there exist nodes $u_1, u_2, v_1, v_2 \in V$ and numbers $d_1, d_2 \in \mathbb{N}$ such that

- $u_0 \xrightarrow{d_1} u_1 \xrightarrow{d_2} u_2$ and $v_0 \xrightarrow{d_1} v_1 \xrightarrow{d_2} v_2$,
- $r(u_1, v_1) = k - 1$ and $r(u_2, v_2) = k - 2$.

Let $d = d_1 + d_2$. Since $u_0 \xrightarrow{d} u_2$ there exists $v_0 \xrightarrow{d} v_3$ such that $u_2 \approx_{k-1} v_3$, and since $v_0 \xrightarrow{d} v_2$ there exists $u_0 \xrightarrow{d} u_3$ such that $u_3 \approx_{k-1} v_2$. Since $u_2 \not\approx_{k-1} v_2$, we get $u_2 \neq u_3$ and $v_2 \neq v_3$. From $r(u_2, v_2) = k - 2$ we get $u_3 \approx_{k-1} v_2 \approx_{k-2} u_2 \approx_{k-1} v_3$, which implies $u_3 \approx_{k-2} v_3$. On the other hand, $u_3 \approx_{k-1} v_3$ would imply $u_2 \approx_{k-1} v_2$, which contradicts $r(u_2, v_2) = k - 2$. We therefore have $r(u_3, v_3) = k - 2$. Since $u_0 \xrightarrow{d} u_2$, $u_0 \xrightarrow{d} u_3$, and

$u_2 \neq u_3$, the nodes u_2 and u_3 are roots of disjoint subtrees, which implies $|u_0| \geq |u_2| + |u_3|$. Analogously, we get $|v_0| \geq |v_2| + |v_3|$, and hence

$$|u_0| + |v_0| \geq |u_2| + |v_2| + |u_3| + |v_3|.$$

Observe that either

$$\frac{1}{2}(|u_0| + |v_0|) \geq |u_2| + |v_2|,$$

or

$$\frac{1}{2}(|u_0| + |v_0|) \geq |u_3| + |v_3|$$

holds. Since the two cases are symmetric we will assume the former. The induction hypothesis implies

$$k - 2 \leq 2 \log_2(|u_2| + |v_2|) \leq 2 \log_2(|u_0| + |v_0|) - 2,$$

and hence $k \leq 2 \log_2(|u_0| + |v_0|)$. \square

Lemma 3.11. *Let G be an unlabelled dag with root u_0 and T be an unlabelled tree with root v_0 of size n . If $u_0 \sim v_0$, then $r(G \uplus T) \leq 2 \log_2(n) + 2$.*

Proof. Assume that $u_0 \sim v_0$. Since u_0 and v_0 are the roots of the dag G and the tree T , respectively, every node in G is bisimilar to some node in T . Furthermore r is invariant under bisimulation, i.e., if $u \sim v$ and $u' \sim v'$, then $r(u, u') = r(v, v')$. Therefore $r(G \uplus T) = r(T) \leq 2 \log_2(n) + 2$. \square

Now we can show the AC^1 upper bound for $\text{BISIM}(\text{Trees}, \text{Graphs})$:

Theorem 3.12. $\text{BISIM}(\text{Trees}, \text{Graphs})$ is in AC^1 .

Proof. Let T be the input tree and G be the input graph. As explained before, we can assume that T and G are unlabelled. Assume we want to test whether node $u_0 \in V(T)$ is bisimilar to node $v_0 \in V(G)$. First we restrict G to the subgraph reachable from v_0 using a reachability oracle in $\text{NL} \subseteq \text{AC}^1$. If G is not acyclic, which can be also tested in NL , then u_0 and v_0 are not bisimilar. If G is a dag, we construct a logspace-uniform AC^1 -circuit family which computes the approximants \approx_k on the disjoint union $G \uplus T$ for all $0 \leq k \leq 2 \log_2 |T| + 2$. A similar construction can be found in [GV06]. To do so, we first compute for all nodes u, u' the set of path lengths from u to u' . Note that one can test in $\text{NL} \subseteq \text{AC}^1$ whether there is (no) path of a given length between two given nodes. Using this information we can compute \approx_{k+1} by an AC^0 -circuit from \approx_k .

Finally, we test whether $\approx_k = \approx_{k+1}$ for $k = \lceil 2 \log_2 |T| + 2 \rceil$. If this is not the case, then Lemma 3.11 implies $u_0 \not\sim v_0$ and the circuit outputs false. Otherwise, $\approx_k = \sim$. Hence the circuit outputs true if $u_0 \approx_k v_0$, and false otherwise. \square

4. FURTHER APPLICATIONS

Before we consider simulation problems, let us present two further application of the techniques from Section 3.

4.1. Equality of hereditarily finite sets. The bisimulation problem on trees in term representation arises in a very natural way. A *hereditarily finite set* is either the empty set $\{\}$ or a set $\{a_1, \dots, a_n\}$ containing finitely many hereditarily finite sets a_1, \dots, a_n . Hereditarily finite sets have a natural string representation over the bracket symbols $\{$ and $\}$. By counting brackets, one can check in TC^0 whether a string over $\{$ and $\}$ is well-bracketed [BC89]. As before, such a well-bracketed string corresponds to a tree. By induction over the height of trees, one can easily show that two well-bracketed strings over $\{$ and $\}$ represent the same set if and only if the corresponding trees are bisimilar. Hence, the tree bisimulation problem for (unlabelled) trees in term representation is equivalent to the *set equality problem*, which asks whether two such string representations represent the same set. For example $\{\{\}\{\}\}$ and $\{\{\}\}$ represent the same set. From Theorem 3.2 we obtain the NC^1 upper bound in the following result. The NC^1 lower bound follows from the NC^1 -hardness of the bisimulation problem for unlabelled trees in term representation (Theorem 3.7).

Corollary 4.1. *The set equality problem is NC^1 -complete with respect to AC^0 -reductions.*

4.2. Model checking on tree-shaped graphs. Recall that the starting point of our paper was the algorithmic meta theorem of Elberfeld et al. (Theorem 1.1) which states that a fixed MSO-property can be model-checked in logarithmic space on a graph class \mathcal{C} of bounded tree-width. On the other hand, if \mathcal{C} has unbounded tree-width, then Theorem 1.1 cannot be applied directly. In this section, we argue that we can extend Theorem 1.1 to some graph classes of unbounded tree-width if we restrict to bisimulation-invariant MSO (which coincides with the modal μ -calculus).

Consider an MSO-formula $\psi = \psi(x)$ with a single free element variable x . We would like to know, whether for a given graph $G \in \mathcal{C}$ and a node $v \in V(G)$, ψ is satisfied in v ; $(G, v) \models \psi$ for short. We say that ψ is *bisimulation-invariant* if for all graphs G and all nodes $u, v \in V(G)$ we have: if u and v are bisimilar and $(G, u) \models \psi$ holds, then $(G, v) \models \psi$. It is known that an MSO-formula ψ is bisimulation-invariant if and only if ψ is equivalent to a formula in the *modal μ -calculus* [JW96]. Assume now that there exist a constant k and a logspace transducer, that maps a graph $G \in \mathcal{C}$ to a pair (H, R) , where H is a graph of tree-width at most k and $R \subseteq V(G) \times V(H)$ is a bisimulation R such that for all $v \in V(G)$ there exists $(v, w) \in R$. Then we can check in logspace whether $(G, v) \models \psi$ holds, by computing a bisimilar node $w \in V(H)$ with $(v, w) \in R$ and checking $(H, w) \models \psi$ in logspace using Theorem 1.1.

Our second proof of Theorem 3.5 uses the above strategy. Whether a certain gate of a Boolean circuit (viewed as a node-labelled graph) evaluates to true is a bisimulation invariant property. The partial tree unfolding constructed in the second proof of Theorem 3.5 has the properties of the graph H from the previous paragraph.

The following definition presents a generalization of tree decompositions that allows to compute a pair (H, R) with H of bounded tree-width and $R \subseteq V(G) \times V(H)$ a bisimulation.

Definition 4.2. Let $G = (V, E)$ be an edge-labelled graph. A *tree-shaped decomposition* for G is a node-labelled tree (T, β) where $\beta: V(T) \rightarrow 2^V$ satisfies the following conditions:

- For all $v \in V$ there exists $t \in V(T)$ such that $v \in \beta(t)$.
- For all $u \xrightarrow{a} v$ there exists $t \in V(T)$ such that $u, v \in \beta(t)$.
- Let $u \xrightarrow{a} v$ and $t_1 \in V(T)$ such that $u \in \beta(t_1)$. Then there exists $t_3 \in V(T)$ such that $u, v \in \beta(t_3)$ and for all nodes $t_2 \in V(T)$ between t_1 and t_3 , we have $u \in \beta(t_2)$.

The *width* of (T, β) is $\max_{t \in V(T)} |\beta(t)|$.

Note that every tree decomposition is also a tree-shaped decomposition but not vice versa since we do not demand that $\{t \in V(T) \mid v \in \beta(t)\}$ is connected. The third property of tree-shaped decompositions ensures that directed edges in G can be “traced” by paths in T .

Proposition 4.3. *For every constant $k \in \mathbb{N}$, there exists a logspace transducer which, given a graph G and a tree-shaped decomposition (T, β) for G of width at most k , outputs a graph H of tree-width at most $k - 1$ and a bisimulation $R \subseteq V(G) \times V(H)$ such that for all $v \in V(G)$ there exists $(v, w) \in R$.*

Proof. Consider the set of pairs

$$P = \{(v, t) \in V(G) \times V(T) \mid v \in \beta(t)\}.$$

We say that two pairs $(v, t_1), (w, t_3) \in P$ are *related* if $v = w$ and for all nodes $t_2 \in V(T)$ on the path between t_1 and t_3 we have $v \in \beta(t_2)$, which can be tested in logspace. This relation is an equivalence relation on P and we denote with $[v, t]$ the equivalence class that contains (v, t) . We take the set of equivalence classes $\{[v, t] \mid (v, t) \in P\}$ as the node set of the graph H . For all $t \in V(T)$ and $v, w \in \beta(t)$ with $v \xrightarrow{a} w$ we add the edge $[v, t] \xrightarrow{a} [w, t]$ to H . Clearly, H is computable in logspace from G and (T, β) .

First, we claim that (T, γ) is a tree decomposition for H of width at most $k - 1$ where $\gamma(t) = \{[v, t] \mid v \in \beta(t)\}$ for all $t \in V(T)$. Clearly, every node and every edge of H is contained in some bag $\gamma(t)$. Moreover, every bag $\gamma(t)$ has size at most k . We need to show that for each node $[v, t] \in V(H)$ the set

$$\{t' \in V(T) \mid [v, t] \in \gamma(t')\}$$

is connected. Let $t_1, t_3 \in V(T)$ such that $[v, t] \in \gamma(t_1) \cap \gamma(t_3)$. By definition of γ the pairs $(v, t), (v, t_1)$ and (v, t_3) must be related. Hence for all $t_2 \in V(T)$ on the unique path between t_1 and t_3 we have $v \in \beta(t_2)$, which implies $[v, t_2] \in \gamma(t_2)$. Since (v, t_2) and (v, t) are also related, we conclude $[v, t] \in \gamma(t_2)$.

Now we show that $R = \{(v, [v, t]) \mid (v, t) \in P\}$ is a bisimulation between G and H . Let $(v, t) \in P$, i.e., $v \in \beta(t)$. If $v \xrightarrow{a} w$ in G , then there exists $t' \in V(T)$ such that $v, w \in \beta(t')$ and the pairs (v, t) and (v, t') are related. Hence we find the matching edge $[v, t] = [v, t'] \xrightarrow{a} [w, t']$ in H . Conversely, if $[v, t] \xrightarrow{a} [w, t]$ is an edge in H , then we know by definition that $v \xrightarrow{a} w$ is an edge in G . \square

Proposition 4.3 and Theorem 1.1 together yield:

Theorem 4.4. *Let $k \in \mathbb{N}$ be a constant and φ be a μ -calculus formula. There exists a logspace algorithm which, given a graph G , a node $v \in V(G)$ and a tree-shaped decomposition (T, β) for G of width at most k , tests whether $(G, v) \models \varphi$.*

Note that in Theorem 4.4 we need the tree-shaped decomposition (T, β) as part of the input. We do not know whether for every constant $k \in \mathbb{N}$ there exists a logspace transducer that computes from a given graph G a tree-shaped decomposition (T, β) for G of width at most k and polynomial size, or outputs 0 if such a decomposition does not exist. For tree decompositions, such a logspace transducer indeed exists [EJT10].

5. COMPLEXITY OF SIMULATION

In this section, we consider simulation problems on restricted classes of finite graphs. As for bisimulation, it is known that simulation on finite graphs is P-complete.

The upper bounds from Theorem 3.1 and 3.2 for BISIM(Trees) carry over to SIM(Trees). The proofs are in fact much easier, since the simulation problem for trees reduces to the evaluation of the Boolean circuit obtained from (3.1) by removing the second conjunction over all edges $v \xrightarrow{a} v'$; in fact, this circuit is a tree. Also the lower bound proofs for Theorem 3.7 hold for the simulation problem: Note that in Figure 6(A) the trees T_1 and T_2 are isomorphic and hence T_1 is simulated by T_2 . On the other hand, in Figure 6(B), T_1 is not simulated by T_2 . Similarly, for the AND-gadget $T_\wedge(G_1, G_2, H_1, H_2)$ and OR-gadget $T_\vee(G_1, G_2, H_1, H_2)$ (see Figure 7) from the proof of Theorem 3.7 we have

- $G_1 \sqsubseteq H_1$ and $G_2 \sqsubseteq H_2 \iff T_\wedge(G_1, G_2, H_1, H_2) \sqsubseteq T_\wedge(H_1, H_2, G_1, G_2)$, and
- $G_1 \sqsubseteq H_1$ or $G_2 \sqsubseteq H_2 \iff T_\vee(G_1, G_2, H_1, H_2) \sqsubseteq T_\vee(G_1, H_2, H_1, G_2)$.

Hence we get:

Theorem 5.1. *The problem SIM(Trees) is L-complete (resp., NC¹-complete) if trees are given in pointer representation (resp., term representation).*

On the other hand, the construction from the proof of Lemma 3.8 (elimination of edge labels) does not carry over to the simulation problem. Indeed, there is a difference in the complexities of the simulation problems for edge-labelled trees and unlabelled trees, if trees are given in term representation. The easy observation is that in unlabelled graphs, u is simulated by v if and only if the longest path starting in u is at most as long as the longest path starting in v (the path lengths can be infinite).

Theorem 5.2. *The following holds:*

- *The problem SIM(Unlabelled-Graphs) is NL-complete.*
- *The problem SIM(Unlabelled-Trees) is L-complete if the trees are given in pointer representation.*
- *The problem SIM(Unlabelled-Trees) is TC⁰-complete if the trees are given in term representation.*

Proof. The upper bounds are easy to show. In general graphs we can test in NL whether from a given node there exists a path of a given length $d \in \mathbb{N} \cup \{\infty\}$. Since NL is closed under complement, we can also check whether from a given node there does not exist a path of a given length $d \in \mathbb{N} \cup \{\infty\}$. To decide simulation for unlabelled trees we only need to compare the heights, which can be done in logspace if the trees are given in pointer representation, and in TC⁰ if the trees are given in term representation.

For the NL-hardness on general graphs we reduce from the NL-complete problem whether a cycle is reachable from a given node u in a given graph G . The latter holds if and only if u is simulated by the root of a path of length $|V(G)|$.

For the L-hardness on trees we again reduce from the reachability problem on directed paths [Ete97]. Assume that G is a path (which we view as a tree) with n nodes and we want to test whether there exists a path from $u \in V(G)$ to $v \in V(G)$. Note that such a path exists if and only if the longest path starting in v is at most as long as the longest path starting in u . By the observation before Theorem 5.2 this is equivalent to the fact that v is simulated by u .

For the TC^0 -hardness we reduce from the TC^0 -complete majority problem: Given a string $x = a_1 \cdots a_n \in \{(,)\}^*$, is the number of opening brackets in x at least the number of closing brackets in x ? Consider the path $T_1 = ({}^{2n+2})^{2n+2}$ of length $2n + 1$ and the tree

$$T_2 = ({}^{n+1}x ({}^{n+1})^{n+1} \bar{x})^{n+1}$$

where $\bar{x} = \overline{a_n} \cdots \overline{a_1}$ and $\bar{}$ swaps $($ and $)$. One can verify that T_2 is indeed a valid term representation and that both terms can be computed in AC^0 from the string x . Notice that the node in T_2 that is represented by the innermost bracket pair $($ has maximal depth. Now one can show that the root of T_1 is simulated by the root of T_2 if and only if the height of T_2 is at least $2n + 1$ if and only if the number of $($'s in x is at least the number of $)$'s. \square

In the rest of Section 5 we complete the complexity picture for the simulation problem on finite graphs with edge labels. In Section 5.1 we show that P -completeness for the simulation problem already holds for graphs of bounded path-width. In Section 5.2 we show that simulation between a tree and an arbitrary graph, as well as simulation between an arbitrary graph and a tree is LogCFL -complete.

5.1. Simulation on graphs of bounded path-width. In this section we show that the simulation problem is P -complete on graphs of bounded path-width, and hence also on graphs of bounded tree-width. It remains open whether the bisimulation problem for graphs of bounded tree-width belongs to NC or remains P -complete. For integers i, j we use the abbreviation $[i, j] = \{k \in \mathbb{N} \mid i \leq k \leq j\}$. Let $\text{PW}_{\leq k}$ denote the class of all graphs with path-width at most k .

Theorem 5.3. *There is a number k such that $\text{SIM}(\text{PW}_{\leq k})$ P -complete.*

Proof. Fix a P -complete language $L \subseteq \{0, 1\}^*$ and a deterministic polynomial time bounded Turing machine $M = (Q, \Gamma, \{0, 1\}, q_0, q_f, \delta)$ that accepts L . Here Q is the set of states, $\Gamma \supseteq \{0, 1, \square\}$ is the tape alphabet (\square is the blank symbol), q_0 is the initial state, q_f is the final state, and $\delta: Q \times \Gamma \rightarrow Q \times \Gamma \times \{\rightarrow, \leftarrow\}$ is the transition function, where \rightarrow and \leftarrow indicate the head direction. The machine has a single two-way infinite tape, whose cells are indexed with integers. Initially, the input x is written in cells $0, \dots, |x| - 1$ and the tape head scans cell 0. We can assume that there is a polynomial $p(n)$ such that for every input $x \in \{0, 1\}^*$ we have: $x \in L$ if and only if after $p(|x|)$ many transitions the machine is in state q_f , cell 0 contains \square , and the tape head scans cell 0.

We can view configurations of M as words from $\Gamma^*(Q \times \Gamma)\Gamma^*$. Let $\Omega = \Gamma \cup (Q \times \Gamma)$. We define a partial mapping $\Delta: \Omega^3 \rightarrow \Omega$ which computes from the three symbols at positions $i - 1, i, i + 1$ in a configuration the symbol at position i in the successor configuration. Formally, for all $a, a', b, c \in \Gamma, p, q \in Q$ we define

- $\Delta(a, b, c) = b$
- $\Delta(b, c, (q, a)) = (p, c)$ if $\delta(q, a) = (p, a', \leftarrow)$
- $\Delta(b, c, (q, a)) = c$ if $\delta(q, a) = (p, a', \rightarrow)$
- $\Delta(b, (q, a), c) = a'$ if $\delta(q, a) = (p, a', d)$ for some $d \in \{\rightarrow, \leftarrow\}$
- $\Delta((q, a), b, c) = b$ if $\delta(q, a) = (p, a', \leftarrow)$
- $\Delta((q, a), b, c) = (p, b)$ if $\delta(q, a) = (p, a', \rightarrow)$

In all other cases, Δ is undefined.

Let us fix an input $x = a_0 a_1 \cdots a_{n-1}$ of length $n > 0$ for the machine M and let $N = p(n) + 1$. Then there exists a unique computation of M on input x . We denote with C

the corresponding computation table. Formally, it is a mapping $C: [-N, N] \times [0, N-1] \rightarrow \Omega$, where $C(i, t)$ is the symbol at cell i in the t -th configuration. It can be defined by the following properties:

- $C(0, 0) = (a_0, q_0)$, $C(i, 0) = a_i$ for $i \in [1, n-1]$, $C(i, 0) = \square$ for $i \in [-N, N] \setminus [0, n-1]$,
- $C(-N, t) = C(N, t) = \square$ for all $t \in [0, N-1]$
- $C(i, t) = \Delta(C(i-1, t-1), C(i, t-1), C(i+1, t-1))$ for all $t \in [1, N-1]$, $i \in [-N+1, N-1]$.

Let us fix the set of edge labels $A = \{-1, 0, 1, \alpha\} \uplus \Omega$. We define two edge-labelled graphs P (for position) and T (for time) with edge labels from A and the node sets

$$V(P) = [-N, N] \times \{0, 1\}, \quad V(T) = [0, N-1] \cup [0, N-1] \times \Omega \cup [0, N-1] \times \Omega^3.$$

For better readability, we write edges of P (resp., T) as $x \xrightarrow{a}_P y$ (resp., $x \xrightarrow{a}_T y$). Then, P and T contain the following edges:

$$\begin{aligned} & (i, 0) \xrightarrow{\alpha}_P (i, 1) \text{ for all } i \in [-N, N] \\ & (i, 1) \xrightarrow{\delta}_P (i + \delta, 0) \text{ for all } i \in [-N, N], \delta \in \{-1, 0, 1\} \text{ with } i + \delta \in [-N, N] \\ & (i, 0) \xrightarrow{C(i,0)}_P (i, 0) \text{ for all } i \in [-N, N] \\ & (t, a) \xrightarrow{\alpha}_T (t-1, b, c, d) \text{ for all } a \in \Omega, t \in [1, N-1], (b, c, d) \in \Delta^{-1}(a) \\ & (t, a_{-1}, a_0, a_1) \xrightarrow{\delta}_T (t, a_\delta) \text{ for all } t \in [0, N-2], a_{-1}, a_0, a_1 \in \Omega, \delta \in \{-1, 0, 1\} \\ & (t, a) \xrightarrow{b}_T t \text{ for all } a, b \in \Omega, t \in [1, N-1] \\ & (0, a) \xrightarrow{b}_T 0 \text{ for all } a \in \Omega, b \in \{a, \alpha\} \\ & t \xrightarrow{a}_T t \text{ for all } t \in [0, N-1], a \in A \end{aligned}$$

An example of the construction is shown in Figure 8, where we assume $N = 3$ for simplicity. It is easy to see that both P and T (and hence also the disjoint union of P and T) have bounded path-width. More precisely, P has path-width 3 (the bags are the sets $\{(i, 0), (i, 1), (i+1, 0), (i+1, 1)\}$ for $-N \leq i \leq N-1$), whereas the path-width of T is bounded by $|\Omega|^3 + |\Omega|$ (the bags are the set $\{t, (t, a), (t-1, b, c, d) \mid a, b, c, d \in \Omega\}$ for $1 \leq t \leq N-1$ and $\{(t, a), (t, b, c, d) \mid a, b, c, d \in \Omega\}$ for $0 \leq t \leq N-2$). Recall that $|\Omega|$ is a fixed constant since the machine M is fixed.

The following claim proves the theorem, since M accepts x if and only if $C(0, N-1) = (q_f, \square)$ by our assumptions on M .

Claim. For all $t \in [0, N-1]$, $i \in [-N, N]$ and $a \in \Omega$ such that $i+t \leq N$ and $i-t \geq -N$ we have: $C(i, t) = a$ if and only if $(i, 0) \sqsubseteq (t, a)$.

We prove the claim by induction on t . First, note that for every $v \in V(P)$ and every $t \in [0, N-1] \subseteq V(T)$ we have $v \sqsubseteq t$, since at t we can loop with every $a \in A$. For $t = 0$ note that indeed $(i, 0) \sqsubseteq (0, C(i, 0))$: The only outgoing edges for $(i, 0)$ are labelled with $C(i, 0)$ and α . From $(0, C(i, 0))$ these labels lead to node 0, which simulates every node of P . On the other hand, if $a \neq C(i, 0)$, then $(i, 0)$ has a $C(i, 0)$ -labelled outgoing edge, whereas $(0, a)$ has no such outgoing edge. This implies $(i, 0) \not\sqsubseteq (0, a)$.

Now assume that $t \in [1, N-1]$, $i \in [-N, N]$, $i+t \leq N$, and $i-t \geq -N$ and that the claim holds for $t-1$. First assume that $C(i, t) = a$. Since $t \geq 1$, we have $i+1 \leq N$ and $i-1 \geq -N$, i.e., $i \in [-N+1, N-1]$. We have to show that $(i, 0) \sqsubseteq (t, a)$. The

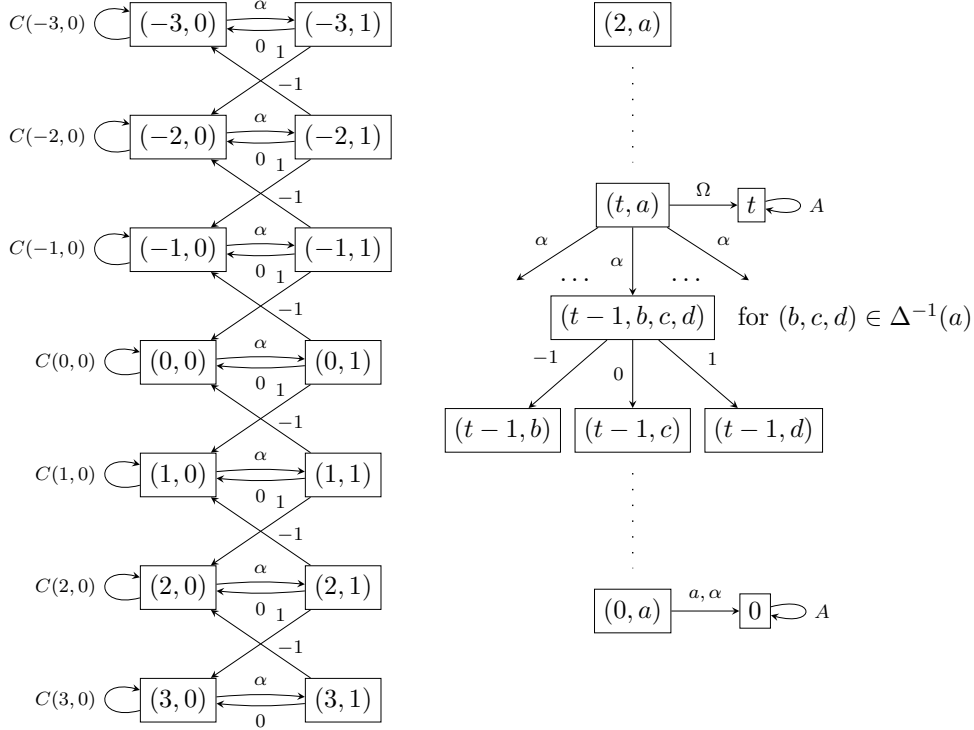


FIGURE 8. The edge-labelled graphs P and T , where we assume $N = 3$ for simplicity.

edge $(i, 0) \xrightarrow{C(i,0)}_P (i, 0)$ can be simulated by the edge $(t, a) \xrightarrow{C(i,0)}_T t$ (recall that $(i, 0) \sqsubseteq t$). Now consider the other possible edge $(i, 0) \xrightarrow{\alpha}_P (i, 1)$. Since $C(i, t) = a$, there must exist $(b, c, d) \in \Delta^{-1}(a)$ such that $b = C(i-1, t-1)$, $c = C(i, t-1)$, and $d = C(i+1, t-1)$. Also note that $i + \delta + (t-1) \leq N$ and $i + \delta - (t-1) \geq -N$ for all $\delta \in \{-1, 0, 1\}$. Hence, by induction $(i-1, 0) \sqsubseteq (t-1, b)$, $(i, 0) \sqsubseteq (t-1, c)$, and $(i+1, 0) \sqsubseteq (t-1, d)$. But this implies that $(i, 1) \sqsubseteq (t-1, b, c, d)$. Hence, we can choose the edge $(t, a) \xrightarrow{\alpha}_T (t-1, b, c, d)$ in order to simulate the edge $(i, 0) \xrightarrow{\alpha}_P (i, 1)$.

Finally, assume that $C(i, t) \neq a$. We have to show that $(i, 0) \not\sqsubseteq (t, a)$. Let us choose the edge $(i, 0) \xrightarrow{\alpha}_P (i, 1)$. We have to show that for every $(b, c, d) \in \Delta^{-1}(a)$, $(i, 1) \not\sqsubseteq (t, b, c, d)$. Let us fix a triple $(b, c, d) \in \Delta^{-1}(a)$. Since $C(i, t) \neq a$, one of the following three statements holds: $C(i-1, t-1) \neq b$, $C(i, t-1) \neq c$, $C(i+1, t-1) \neq d$. Hence, by induction, $(i-1, 0) \not\sqsubseteq (t-1, b)$ or $(i, 0) \not\sqsubseteq (t-1, c)$ or $(i+1, 0) \not\sqsubseteq (t-1, d)$. This implies that, indeed, $(i, 1) \not\sqsubseteq (t, b, c, d)$. \square

It seems to be difficult to modify the above proof so that it shows P-hardness for bisimulation on graphs of bounded path-width or bounded tree-width. One might try to restrict the choices of the players in the bisimulation game (see e.g. [AI07]) so that they are forced to play as in the simulation game. There is a technique to achieve this (defenders forcing) but the problem is that it yields grid-like graph structures and hence graphs of unbounded tree-width.

5.2. Simulation between trees and general graphs. In this section we show that the problems $\text{SIM}(\text{Trees}, \text{Graphs})$ and $\text{SIM}(\text{Graphs}, \text{Trees})$ are both **LogCFL**-complete. Since simulation is not symmetric, it is not obvious that these two problems are equivalent.

Theorem 5.4. $\text{SIM}(\text{Trees}, \text{Graphs})$ is **LogCFL**-complete.

Proof. Again we can restrict ourselves to the case that the input vertices are the roots of the graph and the tree, respectively. We start with showing that the simulation problem in the theorem belongs to **LogCFL**. For this we outline a straightforward alternating logspace algorithm with polynomial proof tree size. Given a tree T and a graph G , the algorithm stores a pair $(u, w) \in V(T) \times V(G)$ and continues as follows:

- Branch universally to all edges $u \xrightarrow{a} u'$ in T .
- Branch existentially to all edges $w \xrightarrow{a} w'$ in G .
- Continue with the pair (u', w') .

The algorithm starts with the pair (r, v) . It is obvious that there exists a proof tree if and only if $r \sqsubseteq v$. Moreover, a proof tree results from T by inserting on each edge a node of G with a unique child.

Let us now prove the lower bound. Fix a **LogCFL**-complete problem L and let w be an input word of length n . Let $(C_n)_{n \geq 0}$ be a logspace-uniform SAC^1 circuit family for L . We first construct in logspace the SAC^1 -circuit C_n for input length n . We set the input gates to the bits in w ; let $C = (G_0, \beta)$ be the resulting Boolean circuit. Recall from Section 2.2 that G_0 is a directed graph $G_0 = (V_0, E_0)$ and that edges from E_0 are directed towards the input gates.

Our goal is to reduce the question whether C evaluates to 1 to a simulation problem. By [GLS01, Lemma 4.6] we can assume that C is layered and alternating, i.e., the set of gates is partitioned into $m \in O(\log n)$ many layers L_1, L_2, \dots, L_m , where $m > 1$ is odd, L_1 contains the input gates, all wires go from layer L_i to layer L_{i-1} for some $2 \leq i \leq m$, and $L_m = \{g_0\}$ contains the output gate of the circuit. Moreover, layers $2i$ ($2i+1$, resp.) contain only \wedge -gates (\vee -gates, resp.) for $1 \leq i \leq \lfloor m/2 \rfloor$. Since m is odd, the output gate g_0 is a \vee -gate. Recall that every \wedge -gate u has exactly two input nodes that we denote with u_1 and u_2 (we allow $u_1 = u_2$).

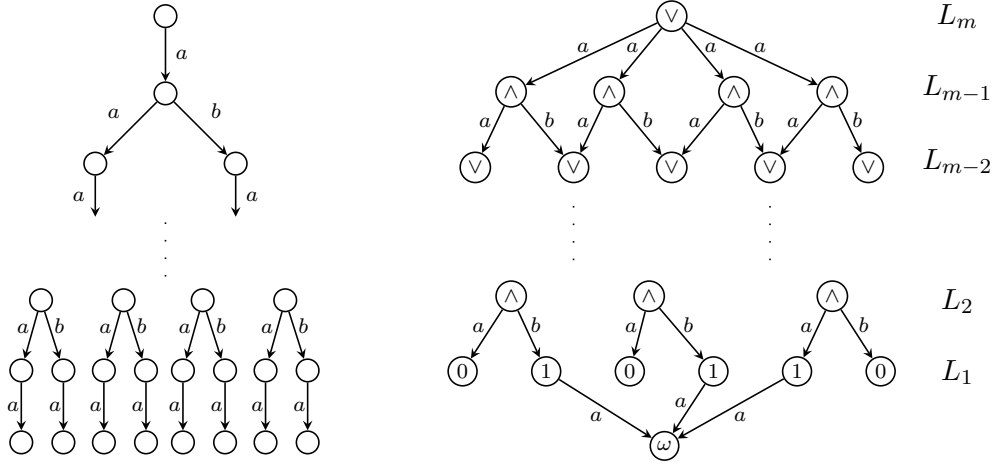
Define the edge-labelled graph $G = (V_0 \uplus \{\omega\}, E)$, where E consists of the following set of $\{a, b\}$ -labelled edges.

- $v \xrightarrow{a} \omega$ for all $v \in L_1$ with $\beta(v) = 1$.
- $v \xrightarrow{a} v'$ for all $v \in L_i$ with $i > 1$ odd and $(v, v') \in E_0$.
- $v \xrightarrow{a} v_1, v \xrightarrow{b} v_2$ for all $v \in L_i$ with $i > 1$ even, where v_1, v_2 are the left and right child of v in G_0 .

The tree T is a full binary tree of height $\lfloor m/2 \rfloor$ with intermediate layers of unary nodes inserted. More precisely, we define trees T_i ($1 \leq i \leq m$) recursively as follows:

- If $i = 1$ then T_i consists of a root node r with a unique outgoing edge $r \xrightarrow{a} r'$, where r' is a leaf.
- If $i > 1$ is odd, then T_i consists of a root node r with a unique outgoing edge $r \xrightarrow{a} r'$, where r' is the root of a copy of T_{i-1} .
- If $i > 1$ is even, then T_i consists of a root node r with two outgoing edges $r \xrightarrow{a} r_1$ and $r \xrightarrow{b} r_2$. Each of the nodes r_1 and r_2 is the root of a copy of T_{i-1} .

The tree T is defined as T_m . Both T and G are displayed in Figure 9.

FIGURE 9. Reduction from SAC^1 -circuits to $\text{SIM}(\text{Trees}, \text{Graphs})$

Claim. If r denotes the root of T_i and $v \in V(G)$ belongs to layer L_i then $r \sqsubseteq v$ if and only if gate v evaluates to 1 in the circuit C .

We prove the claim by induction on i . The case $i = 1$ is clear: In the tree T_1 the root node r has a unique outgoing edge $r \xrightarrow{a} r'$, where r' has no outgoing edges. Moreover, layer L_1 contains the input gates of the circuit C . If v is labelled with 0, then v has no outgoing edges in G , and hence $r \not\sqsubseteq v$. On the other hand, if v is labelled with 1, then v has a single outgoing a -labelled edge to a node without outgoing edges, which implies $r \sqsubseteq v$.

Now assume that $i > 1$. Let us first assume that i is odd. Then, the root r of T_i has a single outgoing edge $r \xrightarrow{a} r'$, where r' is a root of T_{i-1} . Moreover, v is a \vee -gate. If v evaluates to 1, then v has an outgoing edge $v \xrightarrow{a} v'$, where $v' \in L_{i-1}$ evaluates to 1. By induction, we get $r' \sqsubseteq v'$, which implies $r \sqsubseteq v$. On the other hand, if $r \sqsubseteq v$, then v must have an outgoing edge (v, a, v') such that $r' \sqsubseteq v'$. By induction, it follows that v' evaluates to 1. Hence, the \vee -gate v evaluates to 1 too.

Now assume that i is even. Then, the root r of T_i has two outgoing edges $r \xrightarrow{a} r_1$ and $r \xrightarrow{b} r_2$ where r_i is a root of a copy of T_{i-1} . Moreover, v is a \wedge -gate, which has two input gates v_1 and v_2 (we might have $v_1 = v_2$). In G , v has the two outgoing edges $v \xrightarrow{a} v_1$ and $v \xrightarrow{b} v_2$. If v evaluates to 1, then v_1 and v_2 evaluate to 1. By induction, we get $r_1 \sqsubseteq v_1$ and $r_2 \sqsubseteq v_2$, which implies $r \sqsubseteq v$. On the other hand, if $r \sqsubseteq v$, then we must have $r_1 \sqsubseteq v_1$ and $r_2 \sqsubseteq v_2$. Hence, by induction, it follows that v_1 and v_2 both evaluate to 1. Thus, v evaluates to 1. \square

Theorem 5.5. $\text{SIM}(\text{Graphs}, \text{Trees})$ is **LogCFL**-complete.

Proof. The proof of the upper bound is very similar to the previous proof, using an alternating logspace machine with a polynomial proof tree size. Since **LogCFL** is closed under complement [BCD⁺89], we can consider the problem of deciding whether a graph node $u \in V(G)$ is not simulated by a tree node $w \in V(T)$. The algorithm stores the pair $(u, w) \in V(G) \times V(T)$ and continues as follows:

- Branch existentially to all edges $u \xrightarrow{a} u'$ in G .
- Branch universally to all edges $w \xrightarrow{a} w'$ in T .

6. CONCLUSION

We proved the following results:

- (1) The bisimulation problem for trees that are given by pointer structures (resp., in term representation) is complete for deterministic logspace (resp., NC^1). These results also hold for the simulation problem for trees.
- (2) Already for graphs of bounded path-width (a subclass of the graphs of bounded tree-width), the simulation problem becomes P-complete.
- (3) The bisimulation problem between a tree and a dag (or arbitrary graph) belongs to AC^1 and is NL-hard.
- (4) Simulation of a tree by a dag as well as simulation of a dag by a tree is **LogCFL**-complete.

As an application of (1), we showed that equality of hereditarily finite sets is NC^1 -complete. For the proof of (1) we introduced the new class of tree-shaped circuits and proved that the circuit evaluation problem for tree-shaped circuits of bounded width is in logspace or NC^1 , depending on the representation of the circuit. It would be nice to find further applications of tree-shaped circuits.

The main open problem that remains is whether the bisimulation problem for graphs of bounded tree-width is in NC or P-complete. Another open problem is the exact complexity of the bisimulation problem for a tree and a dag. Our bounds (in AC^1 and NL-hard) are not matching. In many situations (e.g. for various classes of infinite state systems), simulation is harder than bisimulation. We proved that simulation of a tree by a dag (or vice versa) is LogCFL-complete. Hence, one might conjecture that also the bisimulation problem for a tree and a dag belongs to LogCFL.

REFERENCES

- [AI07] L. Aceto and A. Ingólfssdóttir. *Reactive Systems: Modelling, Specification and Verification*. Cambridge University Press, 2007.
- [BC89] D. A. M. Barrington and J. Corbet. On the relative complexity of some languages in NC^1 . *Information Processing Letters*, 32:251–256, 1989.
- [BCD⁺89] A. Borodin, S. A. Cook, P. W. Dymond, W. L. Ruzzo, and M. Tompa. Two applications of inductive counting for complementation problems. *SIAM Journal on Computing*, 18(3):559–578, 1989.
- [BGS92] J. L. Balcázar, J. Gabarró, and M. Sántha. Deciding bisimilarity is P-complete. *Formal Aspects of Computing*, 4(6A):638–648, 1992.
- [BIS90] D. A. M. Barrington, N. Immerman, and H. Straubing. On uniformity within NC^1 . *Journal of Computer and System Sciences*, 41:274–306, 1990.
- [BM95] M. Beaudry and P. McKenzie. Circuits, matrices, and nonassociative computation. *Journal of Computer and System Sciences*, 50(3):441 – 455, 1995.
- [Bus97] S. R. Buss. Alogtime algorithms for tree isomorphism, comparison, and canonization. In *Kurt Gödel Colloquium 97*, pages 18–33, 1997.
- [CS01] R. Cleaveland and O. Sokolsky. Chapter 6 - equivalence and preorder checking for finite-state systems. In J.A. Bergstra, A. Ponse, and S.A. Smolka, editors, *Handbook of Process Algebra*, pages 391 – 424. Elsevier Science, Amsterdam, 2001.
- [DF13] R. G. Downey and M. R. Fellows. *Fundamentals of Parameterized Complexity*. Texts in Computer Science. Springer, 2013.
- [EJT10] M. Elberfeld, A. Jakoby, and T. Tantau. Logspace versions of the theorems of Bodlaender and Courcelle. In *Proceedings of FOCS 2010*, pages 143–152. IEEE, 2010.
- [EJT12] M. Elberfeld, A. Jakoby, and T. Tantau. Algorithmic meta theorems for circuit classes of constant and logarithmic depth. In *Proceedings of STACS 2012*, volume 14 of *LIPICs*, pages 66–77. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2012.

- [ES16] M. Elberfeld and P. Schweitzer. Canonizing graphs of bounded tree width in logspace. In *Proceedings of STACS 2016*, volume 47 of *LIPICs*, pages 32:1–32:14. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2016.
- [Ete97] K. Etessami. Counting quantifiers, successor relations, and logarithmic space. *Journal of Computer and System Sciences*, 54(3):400–411, 1997.
- [GGHL12] M. Grohe, B. Grüßen, A. Hernich, and B. Laubner. L-recursion and a new logic for logarithmic space. *Logical Methods in Computer Science*, 9(1), 2012.
- [GGL16] M. Ganardi, S. Göller, and M. Lohrey. On the parallel complexity of bisimulation on finite systems. In *Proceedings of CSL 2016*, volume 62 of *LIPICs*, pages 12:1–12:17. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2016.
- [GHR95] R. Greenlaw, H. J. Hoover, and W. L. Ruzzo. *Limits to Parallel Computation: P-Completeness Theory*. Oxford University Press, 1995.
- [GLS01] G. Gottlob, N. Leone, and F. Scarcello. The complexity of acyclic conjunctive queries. *Journal of the ACM*, 48(3):431–498, 2001.
- [GV06] M. Grohe and O. Verbitsky. Testing graph isomorphism in parallel by playing a game. In *Proceedings of ICALP 2006, Part I*, volume 4051 of *Lecture Notes in Computer Science*, pages 3–14. Springer, 2006.
- [GW00] F. Gurski and E. Wanke. The tree-width of clique-width bounded graphs without $K_{n,n}$. In *Proceedings of WG 2000*, volume 1928 of *Lecture Notes in Computer Science*, pages 196–205. Springer, 2000.
- [HAB02] W. Hesse, E. Allender, and D. A. M. Barrington. Uniform constant-depth threshold circuits for division and iterated multiplication. *Journal of Computer and System Sciences*, 65:695–716, 2002.
- [Imm88] N. Immerman. Nondeterministic space is closed under complementation. *SIAM Journal on Computing*, 17(5):935–938, 1988.
- [JKMT03] B. Jenner, J. Köbler, P. McKenzie, and J. Torán. Completeness results for graph isomorphism. *Journal of Computer and System Sciences*, 66(3):549–566, 2003.
- [JW96] D. Janin and I. Walukiewicz. On the expressive completeness of the propositional mu-calculus with respect to monadic second order logic. In *Proceedings of CONCUR '96*, volume 1119 of *Lecture Notes in Computer Science*, pages 263–277. Springer, 1996.
- [KLM09] M. Kaminski, V. V. Lozin, and M. Milanic. Recent developments on graphs of bounded clique-width. *Discrete Applied Mathematics*, 157(12):2747–2761, 2009.
- [KS90] P. C. Kanellakis and S. A. Smolka. CCS expressions, finite state processes, and three problems of equivalence. *Information and Computation*, 86(1):43–68, 1990.
- [Lin92] S. Lindell. A logspace algorithm for tree canonization (extended abstract). In *Proceedings of STOC 1992*, pages 400–404. ACM, 1992.
- [LM98] K.-J. Lange and P. McKenzie. On the Complexity of Free Monoid Morphisms. In *Proceedings of ISAAC '98*, volume 1533 of *Lecture Notes in Computer Science*, pages 247–256. Springer, 1998.
- [PT87] R. Paige and R. E. Tarjan. Three partition refinement algorithms. *SIAM Journal on Computing*, 16(6):973–989, 1987.
- [Ruz80] W. L. Ruzzo. Tree-size bounded alternation. *Journal of Computer and System Sciences*, 21:218–235, 1980.
- [Ruz81] W. L. Ruzzo. On uniform circuit complexity. *Journal of Computer and System Sciences*, 22:365–383, 1981.
- [SJ05] Z. Sawa and P. Jancar. Behavioural equivalences on finite-state systems are PTIME-hard. *Computers and Artificial Intelligence*, 24(5):513–528, 2005.
- [Srb01] J. Srba. On the power of labels in transition systems. In *Proceedings of CONCUR 2001*, volume 2154 of *Lecture Notes in Computer Science*, pages 277–291. Springer, 2001.
- [Srb04] J. Srba. Roadmap of infinite results. In *Current Trends In Theoretical Computer Science, The Challenge of the New Century*, volume Vol 2: Formal Models and Semantics, pages 337–350. World Scientific Publishing Co., 2004.
- [Sze88] R. Szelepcsényi. The method of forced enumeration for nondeterministic automata. *Acta Informatica*, 26(3):279–284, 1988.
- [Ven91] H. Venkateswaran. Properties that characterize LOGCFL. *Journal of Computer and System Sciences*, 43(2):380–404, 1991.
- [Vol99] H. Vollmer. *Introduction to Circuit Complexity*. Springer, 1999.