Charles University in Prague

Faculty of Mathematics and Physics

DOCTORAL THESIS

Jakub Bulín

Constraint satisfaction, graphs and algebras

Department of Algebra

Supervisor of the doctoral thesis:	Libor Barto
Study programme:	Mathematics
Specialization:	Algebra, Theory of Numbers and Mathematical Logic

Prague 2014

Above all I wish to thank my advisor Libor Barto for his support, guidance and unlimited patience. The way I understand and enjoy mathematics was also significantly influenced by the following people (ordered alphabetically): Erhard Aichinger, Marcel Jackson, Marcin Kozik, Petar Marković, Miklós Maróti, Peter Mayr, Todd Niven and David Stanovský as well as my fellow students Alexandr Kazda and Jakub Opršal. A significant amount of credit is due to my family and friends for their continuous support during my graduate studies.

I declare that I carried out this doctoral thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Coll., the Copyright Act, as amended, in particular the fact that the Charles University in Prague has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 paragraph 1 of the Copyright Act.

In Prague, 10th July 2014 Jakub Bulín

Název práce: CSP, grafy a algebry

Autor: Jakub Bulín

Katedra: Katedra algebry

Vedoucí disertační práce: Mgr. Libor Barto, Ph.D., Katedra algebry

Abstrakt: Tato práce sestává ze tří článků v oblasti algebraického přístupu k problému splňování podmínek (CSP). V prvním článku, se spoluautory Delićem, Jacksonem a Nivenem, studujeme redukci CSP na orientované grafy. Pro každou relační strukturu $\mathbb{A}$ konstruujeme orientovaný graf $\mathcal{D}(\mathbb{A})$ takový, že CSP($\mathbb{A}$) a CSP($\mathcal{D}(\mathbb{A})$) jsou logspace ekvivalentní a většina relevantních vlastností se přenáší z $\mathbb{A}$ na $\mathcal{D}(\mathbb{A})$. Důsledkem je, že algebraické hypotézy charekterizující CSP řešitelné v **P**, **NL** a **L** jsou ekvivalentní jejich restrikcím na orientované grafy. Ve druhém článku dokazujeme, že pro danou core relační strukturu $\mathbb{A}$ s konečnou šířkou a $B \subseteq A$ lze algoritmicky rozhodnout, zda je B absorbující podalgebra algebry polymorfismů $\mathbb{A}$. Jako vedlejší produkt získáváme, že Jónssonova absorpce se v tomto případě shoduje s obvyklou absorpcí. Ve třetím článku, za použití moderních algebraických nástrojů (např. teorie absorpce a pointující operace), potvrzujeme hypotézu o dichotomii CSP pro tzv. speciální orientované stromy. Konkrétně, core speciální stromy řešitelné v **P** mají konečnou šířku.

Klíčová slova: problém splňování omezení, algebra polymorfismů, absorbující podalgebra, konečná šířka, orientovaný strom

Title: Constraint satisfaction, graphs and algebras

Author: Jakub Bulín

Department: Department of Algebra

Supervisor: Mgr. Libor Barto, Ph.D., Department of Algebra

Abstract: This thesis consists of three papers in the area of algebraic approach to the constraint satisfaction problem. In the first paper, a joint work with Delić, Jackson and Niven, we study the reduction of fixed template CSPs to digraphs. We construct, for every relational structure $\mathbb{A}$, a digraph $\mathcal{D}(\mathbb{A})$ such that CSP($\mathbb{A}$) and CSP($\mathcal{D}(\mathbb{A})$) are logspace equivalent and most relevant properties carry over from $\mathbb{A}$ to $\mathcal{D}(\mathbb{A})$. As a consequence, the algebraic conjectures characterizing CSPs solvable in **P**, **NL** and **L** are equivalent to their restrictions to digraphs. In the second paper we prove that, given a core relational structure $\mathbb{A}$ of bounded width and $B \subseteq A$, it is decidable whether B is an absorbing subuniverse of the algebra of polymorphisms of $\mathbb{A}$. As a by-product, we show that Jónsson absorption coincides with the usual absorption in this case. In the third paper we establish, using modern algebraic tools (e.g. absorption theory and pointing operations), the CSP dichotomy conjecture for so-called special oriented trees and in particular prove that all tractable core special trees have bounded width.

Keywords: constraint satisfaction problem, algebra of polymorphisms, absorbing subuniverse, bounded width, oriented tree

Contents

Introduction

Many decision problems from diverse areas of computer science (eg. graph theory, database theory, artificial intelligence, scheduling) can be naturally expressed in a common framework, as Constraint Satisfaction Problems (CSPs). In a CSP, one seeks to assign values to variables subject to constraints on possible evaluations of tuples of variables. The history of this problem dates back to 1970s and it has been central to the development of theoretical computer science in the past few decades.

Constraint satisfaction problems are **NP**-complete in general; it is natural to impose restrictions on the constraints allowed to appear in an instance. The so-called fixed template CSP can then be expressed as the homomorphism problem for a fixed relational structure:

Definition. For a fixed relational structure $\mathbb{A}$[1], the *constraint satisfaction problem with template* $\mathbb{A}$ is the membership problem for the class

$$\mathrm{CSP}(\mathbb{A}) = \{\mathbb{X} \mid \mathbb{X} \to \mathbb{A}\}$$

of structures (of the same *type* as $\mathbb{A}$) admitting a homomorphism to $\mathbb{A}$. For a (directed) graph $\mathbb{H}$, $\mathrm{CSP}(\mathbb{H})$ is also known as the $\mathbb{H}$-*coloring problem.*

A lot of interest in this class of problems was sparked by a seminal work of Feder and Vardi [18], in which the authors conjectured fixed template CSPs to be the (in some sense) "largest natural" class of **NP** decision problems avoiding the complexity classes strictly between **P** and **NP**-complete (assuming that **P**$\neq$**NP**).

The CSP dichotomy conjecture. *For every relational structure* $\mathbb{A}$*,* $\mathrm{CSP}(\mathbb{A})$ *is in* ***P*** *or* ***NP****-complete.*

A major breakthrough towards this conjecture followed the discovery of a rather intimate connection of decision CSPs to universal algebra [23, 13]. Each structure $\mathbb{A}$ can be associated with a finite algebra, built up from operations preserving the relations of $\mathbb{A}$, the so called "polymorphisms". The variety generated by this algebra of polymorphisms of $\mathbb{A}$ then controls complexity (as well as other properties) of $\mathrm{CSP}(\mathbb{A})$ in the following fashion: Either the variety satisfies some "nice" identities (or *Maltsev conditions*) which manifest themselves (usually in a highly non-trivial way) in structural properties and can be exploited algorithmically, or there is a "bad" member in the variety which can be used to show that $\mathrm{CSP}(\mathbb{A})$ can encode a "hard" problem (for various meanings of "hardness").

Examples include the conjectured classifications of CSPs solvable in **P** [13], **NL** and **L** [24], or classes of problems solvable by certain types of algorithms, e.g.

[1] We only consider finite relational structures.

local consistency checking [6] (so-called *bounded width*; the characterization was conjectured, and the "hardness" part proved, in [25]) or "generalized Gaussian elimination" [10]. Other important results built upon this algebraic approach include dichotomies for three-element templates [12] (generalizing Schaefer's dichotomy theorem for the boolean case [26]), for conservative structures [11] (see also [3]) and for digraphs with no sources and no sinks [8] (generalizing the pre-algebraic dichotomy for undirected graphs [21]).

This thesis consists of three contributions in the area of algebraic approach to the constraint satisfaction problem, based on the following papers:

(1) Bulín, J., Delić, D., Jackson, M., Niven, T.: A finer reduction of constraint problems to digraphs. Preprint (2014),
`http://arxiv.org/abs/1406.6413`

A subset of the results is already published, in proceedings of the 19th International Conference on Principles and Practice of Constraint Programming [16].

(2) Bulín, J.: Decidability of absorption in relational structures of bounded width. Algebra Universalis, published electronically (2014),
`http://dx.doi.org/10.1007/s00012-014-0283-2`

(3) Bulín, J.: On the complexity of $\mathbb{H}$-coloring for special oriented trees. Preprint (2014),
`http://arxiv.org/abs/1407.1779`

Below we briefly present contents of these papers. All three parts are more or less self-contained, inside the reader will find more motivation and discussion of the results as well as definitions of the notions used in this introduction.

Part I – Reduction to digraphs

This part is based on the paper [17]. Feder and Vardi in their paper [18] not only conjectured the CSP dichotomy, but also reduced the conjecture to the particular case of digraphs. Specifically, for every structure $\mathbb{A}$ they constructed a digraph $\mathcal{D}(\mathbb{A})$ such that $\mathrm{CSP}(\mathbb{A})$ and $\mathrm{CSP}(\mathcal{D}(\mathbb{A}))$ are polynomial-time equivalent.

In this paper, a joint work with Delić, Jackson and Niven, we present a simple variant of such a construction and prove that (under our construction[2]) $\mathrm{CSP}(\mathbb{A})$ and $\mathrm{CSP}(\mathcal{D}(\mathbb{A}))$ are, in fact, logspace equivalent and most properties relevant to (algebraic approach to) the CSP carry over from $\mathbb{A}$ to $\mathcal{D}(\mathbb{A})$. The main results from this paper are summarized in the following theorem:

Theorem. *For every relational structure $\mathbb{A}$ there exists a digraph $\mathcal{D}(\mathbb{A})$ such that the following holds:*

(i) $\mathrm{CSP}(\mathbb{A})$ *and* $\mathrm{CSP}(\mathcal{D}(\mathbb{A}))$ *are logspace equivalent.*

(ii) $\mathbb{A}$ *is a core if and only if* $\mathcal{D}(\mathbb{A})$ *is a core.*[3]

[2]It is shown in [22] that there are other constructions of $\mathcal{D}(\mathbb{A})$ under which the reduction from $\mathrm{CSP}(\mathcal{D}(\mathbb{A}))$ to $\mathrm{CSP}(\mathbb{A})$ is still polynomial-time, but which are not as well behaved as ours.

[3]A finite relational structure $\mathbb{A}$ is a *core*, if every endomorphism of $\mathbb{A}$ is an automorphism. The algebraic approach works well only for cores, but it is easy to prove that for every $\mathbb{A}$ there exists a core $\mathbb{A}'$ such that $\mathrm{CSP}(\mathbb{A}) = \mathrm{CSP}(\mathbb{A}')$.

(iii) *If Σ is a linear idempotent set of identities such that the algebra of polymorphisms of the oriented path $\bullet \to \bullet \leftarrow \bullet \to \bullet$ satisfies Σ and each identity in Σ is either balanced or contains at most two variables, then*

$$\mathbb{A} \models \Sigma \text{ if and only if } \mathcal{D}(\mathbb{A}) \models \Sigma.$$

The condition on Σ in item (iii) is not very restrictive: it includes almost all of the commonly encountered properties relevant to the CSP. Indeed, to date, these include all Maltsev conditions that are conjectured to divide differing levels of tractability and hardness, as well as all the main tractable algorithmic classes (e.g. few subpowers, bounded width, bounded strict width, etc.). In particular, it follows that the algebraic conjectures characterizing CSPs in **P** [13], **NL** and **L** [24] are equivalent to their restrictions to the case of digraphs.

In the conference version [16] we proved that the conjecture characterizing CSPs solvable in **P** (the so-called Algebraic CSP dichotomy conjecture) is equivalent to its restriction to digraphs. This was established by showing that our construction preserves a particular Maltsev condition, namely existence of a *weak near-unanimity* polymorphism. In this paper the result is extended to include many more Maltsev conditions; and the logspace reduction from CSP($\mathcal{D}(\mathbb{A})$) to CSP($\mathbb{A}$) is also new.

Part II – Decidability of absorption

This part is based on the paper [14]. An essential idea of the proof of the so-called Bounded Width Theorem [6] is that an instance of CSP($\mathbb{A}$) can be reduced to certain subsets of the template $\mathbb{A}$, the *absorbing subuniverses* of its algebra of polymorphisms.

Definition. Let $\mathbf{A}$ be an algebra and $B \leq \mathbf{A}$. We say that B is an *absorbing subuniverse* of $\mathbf{A}$, if there exists an idempotent term operation t of $\mathbf{A}$ such that

$$\begin{gathered} t(A, B, B, \dots, B, B) \subseteq B, \\ t(B, A, B, \dots, B, B) \subseteq B, \\ \vdots \\ t(B, B, B, \dots, B, A) \subseteq B. \end{gathered}$$

The notion of absorbing subuniverse was motivated by algebras with a *near unanimity* operation. Jónsson terms (characterizing congruence distributivity) generalize in a similar fashion to *Jónsson-absorbing subuniverses*.

The idea of absorption has proven to be very useful in a number of other problems related to the CSP and structure of finite algebras in general. We refer the reader to [5, 9] for some of the applications.

In [3], Barto used absorption to provide a new algorithm for the CSP for conservative templates (i.e., relational structures containing all subsets as unary relations), significantly simplifying the result of Bulatov [11]. The new algorithm uses knowledge of absorbing subuniverses of the algebra of polymorphisms as a black box, which led Barto to formulate the following problem.

Problem (Problem 24 in [1]). *Given a finite relational structure $\mathbb{A}$ and a subset $B \subseteq A$, is it decidable whether B is an absorbing subuniverse of the algebra of polymorphisms of $\mathbb{A}$?*

The main result of this part is the following theorem:

Theorem. *Let $\mathbb{A}$ be a core relational structure of bounded width and $B \subseteq A$.*

(i) If B is a Jónsson-absorbing subuniverse of the algebra of polymorphisms of $\mathbb{A}$, then it is an absorbing subuniverse.

(ii) If B is an absorbing subuniverse of the algebra of polymorphisms of $\mathbb{A}$, then this absorption is realized via some polymorphism of arity at most $4^{8|A|^k}+1$, where k is the maximum arity of a relation of $\mathbb{A}$.

The proof is based on techniques developed for the proof of the Zádori conjecture by Barto [2]. As a corollary, we provide a partial solution to the above problem.

Corollary. *Given a core relational structure $\mathbb{A}$ of bounded width and $B \subseteq A$, there is a* ***co-NEXPTIME*** *algorithm that checks whether B is an absorbing subuniverse of the algebra of polymorphisms of $\mathbb{A}$.*

In the rest of the paper [14] we show that some questions about absorption in relational structures can be reduced to digraphs. For that, we use the construction from Part I and prove that it preserves the property of being an absorbing subuniverse as well as the arity of absorbing terms and also, in a sense, it preserves the absorption-free subuniverses. We conclude by discussing the problem of characterizing finite algebras which generate a pseudovariety containing no absorption-free members, another open problem in absorption theory.

Part III – Oriented trees

This part is based on the paper [15], and continues a line of research started in a joint work with Barto [4]. Using the algebraic approach, Barto, Kozik and Niven confirmed the conjecture of Bang-Jensen and Hell and established the $\mathbb{H}$-coloring dichotomy for *smooth digraphs* (i.e., digraphs with no sources and no sinks) [8].

Our paper is concerned with $\mathbb{H}$-coloring for oriented trees. In the class of all digraphs, oriented trees are in some sense very far from smooth digraphs, and the algebraic tools seem to be not yet developed enough to deal with them. Hence oriented trees serve as a good field-test for new methods.

Except the oriented paths (which are all tractable), the simplest class of oriented trees are *triads* (i.e., oriented trees with one vertex of degree 3 and all other vertices of degree 2 or 1). Unfortunately, the CSP dichotomy conjecture remains open even for triads. Among the triads, Hell, Nešetřil and Zhu [19, 20] identified a (fairly restricted) subclass, the *special triads*, which allowed them to handle at least some examples.

In [7], Barto et al used algebraic methods to prove that every special triad has **NP**-complete $\mathbb{H}$-coloring or a compatible majority operation (so-called *strict width 2*) or compatible totally symmetric idempotent operations of all arities (so-called *width 1*). In [4], the author and Barto established the CSP dichotomy conjecture for *special polyads*, a generalization of special triads where the one vertex of degree > 2 is allowed to have an arbitrary degree. In particular, every tractable core special polyad has bounded width. However, there are special polyads which have bounded width, but neither bounded strict width nor width 1.

In this paper we study *special trees*, a fairly broad generalization of special triads and special polyads. Special trees have an underlying structure of a height 1 oriented tree and while for special triads it has only 7 vertices and for special polyads it has radius 2, for general special trees it can be arbitrary.

Definition. A *special tree* (of height h) is an oriented tree obtained from some oriented tree $\mathbb{T} = (T; E)$ of height 1 by replacing every edge $(a, b) \in E$ with some *minimal*[4] path $\mathbb{P}_{(a,b)}$ of height h, preserving orientation. (That is, identifying the initial vertex of $\mathbb{P}_{(a,b)}$ with a and the terminal vertex with b.)

We confirm the CSP dichotomy conjecture for special trees and, moreover, prove that every tractable core special tree has bounded width:

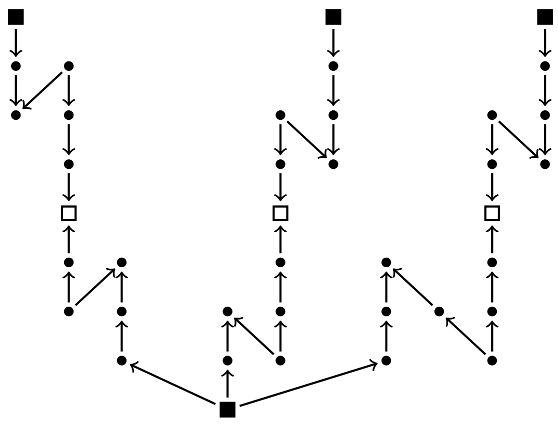
Theorem. *Let $\mathbb{H}$ be a special tree. If the algebra of idempotent polymorphisms of $\mathbb{H}$ is Taylor, then it is congruence meet-semidistributive.*

Corollary. *The CSP dichotomy conjecture holds for special trees. For every core special tree $\mathbb{H}$,* CSP($\mathbb{H}$) *is **NP**-complete or $\mathbb{H}$ has bounded width.*

The proof uses modern tools from the algebraic approach to the CSP (in particular, absorption and pointing operations [9]) and is somewhat simpler and more natural than the proofs in [7] and [4]. Therefore we believe that there is hope for further generalization. In particular, we conjecture that tractability implies bounded width for all oriented trees.

In our terminology, the digraphs $\mathcal{D}(\mathbb{A})$ constructed in Part I are *special balanced digraphs* and the reader may notice similarities with some of the proofs from Part I. We hope that some of our techniques can be adapted to obtain interesting results about special balanced digraphs as well.

We conclude this introduction with an illustration of a special triad from [7], which has **NP**-complete $\mathbb{H}$-coloring and is conjectured to be the smallest oriented tree with this property (vertices from the bottom and top level are marked by ■ and □, respectively).



[4]An oriented path is *minimal*, if its initial vertex has level 0, its terminal vertex has the maximum level and all other vertices lie strictly in between those two levels.

References

[1] Open problems from the Workshop on Algebra and CSPs, Fields Institute for Research in Mathematical Sciences, Toronto (2011), `http://www.karlin.mff.cuni.cz/~barto/articles/fields-2011-problems.pdf`

[2] Barto, L.: Finitely related algebras in congruence distributive varieties have near unanimity terms. Canadian Journal of Mathematics 65(1), 3–21 (2013)

[3] Barto, L.: The dichotomy for conservative constraint satisfaction problems revisited. Proceedings of the 26th Annual IEEE Symposium on Logic in Computer Science, LICS 2011, pp. 21–24 (2011)

[4] Barto, L., Bulín, J.: CSP dichotomy for special polyads. International Journal of Algebra and Computation 23(5), 1151–1174 (2013)

[5] Barto, L., Kozik, M.: Absorbing subalgebras, cyclic terms, and the constraint satisfaction problem. Logical Methods in Computer Science 8(1:07), 1–26 (2012)

[6] Barto, L., Kozik, M.: Constraint satisfaction problems solvable by local consistency methods. Journal of the ACM 61(1), Art. 3, 1–19 (2014)

[7] Barto, L., Kozik, M., Maróti, M., Niven, T.: CSP dichotomy for special triads. Proceedings of the American Mathematical Society 137(9), 2921–2934 (2009)

[8] Barto, L., Kozik, M., Niven, T.: The CSP dichotomy holds for digraphs with no sources and no sinks (a positive answer to a conjecture of Bang-Jensen and Hell). SIAM Journal on Computing 38(5), 1782–1802 (2008)

[9] Barto, L., Kozik, M., Stanovský, D.: Mal'tsev conditions, lack of absorption, and solvability. Manuscript (2013), `http://www.karlin.mff.cuni.cz/~stanovsk/math/maltsev_absorption_solvability.pdf`

[10] Berman, J., Idziak, P., Marković, P., McKenzie, R., Valeriote, M., Willard, R.: Varieties with few subalgebras of powers. Transactions of the American Mathematical Society 362(3), 1445–1473 (2010)

[11] Bulatov, A.: Tractable conservative constraint satisfaction problems. Proceedings of the 18th Annual IEEE Symposium on Logic in Computer Science, LICS '03, p. 321 (2003)

[12] Bulatov, A.: A dichotomy theorem for constraint satisfaction problems on a 3-element set. J. ACM 53(1), 66–120 (2006)

[13] Bulatov, A., Jeavons, P., Krokhin, A.: Classifying the complexity of constraints using finite algebras. SIAM Journal on Computing 34(3), 720–742 (2005)

[14] Bulín, J.: Decidability of absorption in relational structures of bounded width. Algebra universalis, published electronically (2014), `http://dx.doi.org/10.1007/s00012-014-0283-2`

[15] Bulín, J.: On the complexity of H-coloring for special oriented trees. Preprint (2014), `http://arxiv.org/abs/1407.1779`

[16] Bulín, J., Delić, D., Jackson, M., Niven, T.: On the reduction of the CSP dichotomy conjecture to digraphs. Principles and Practice of Constraint Programming, pp. 184–199. Lecture Notes in Computer Science 8124 (2013)

[17] Bulín, J., Delić, D., Jackson, M., Niven, T.: A finer reduction of constraint problems to digraphs. Preprint (2014), `http://arxiv.org/abs/1406.6413`

[18] Feder, T., Vardi, M.Y.: The computational structure of monotone monadic SNP and constraint satisfaction: a study through datalog and group theory. SIAM Journal on Computing 28(1), 57–104 (electronic) (1999)

[19] Hell, P., Nešetřil, J., Zhu, X.: Complexity of tree homomorphisms. Discrete Applied Mathematics 70(1), 23–36 (1996)

[20] Hell, P., Nešetřil, J., Zhu, X.: Duality and polynomial testing of tree homomorphisms. Transactions of the American Mathematical Society 348(4), 1281–1297 (1996)

[21] Hell, P., Nešetřil, J.: On the complexity of H-coloring. Journal of Combinatorial Theory, Series B 48(1), 92–110 (1990)

[22] Jackson, M., Kowalski, T., Niven, T.: Digraph related constructions and the complexity of digraph homomorphism problems. Manuscript (2013), `http://arxiv.org/abs/1304.4986`

[23] Jeavons, P., Cohen, D., Gyssens, M.: Closure properties of constraints. Journal of the ACM 44(4), 527–548 (1997)

[24] Larose, B., Tesson, P.: Universal algebra and hardness results for constraint satisfaction problems. Theoretical Computer Science 410(18), 1629–1647 (2009)

[25] Larose, B., Zádori, L.: Bounded width problems and algebras. Algebra universalis 56(3-4), 439–466 (2007)

[26] Schaefer, T.J.: The complexity of satisfiability problems. Proceedings of the Tenth Annual ACM Symposium on Theory of Computing, STOC '78, pp. 216–226 (1978)

Part I

Reduction to digraphs

A finer reduction of constraint problems to digraphs

J. Bulín, D. Delić, M. Jackson and T. Niven

Abstract

It is well known that the constraint satisfaction problem over a general relational structure $\mathbb{A}$ is polynomial time equivalent to the constraint problem over some associated digraph. We present a variant of this construction and show that the corresponding constraint satisfaction problem is logspace equivalent to that over $\mathbb{A}$. Moreover, we show that almost all of the commonly encountered polymorphism properties are held equivalently on the $\mathbb{A}$ and the constructed digraph. As a consequence, the Algebraic CSP dichotomy conjecture as well as the conjectures characterizing CSPs solvable in logspace and in nondeterministic logspace are equivalent to their restriction to digraphs.

Introduction

A fundamental problem in constraint programming is to understand the computational complexity of constraint satisfaction problems (CSPs). While it is well known that constraint satisfaction problems can be NP-complete in general, there are many subclasses of problems for which there are efficient solving methods. One way to restrict the instances is to only allow a fixed set of constraint relations, often referred to as a *constraint language* [9] or *fixed template.* Classifying the computational complexity of fixed template CSPs has been a major focus in the theoretical study of constraint satisfaction. In particular it is of interest to know which templates produce polynomial time solvable problems to help provide more efficient solution techniques.

The study of fixed template CSPs dates back to the 1970's with the work of Montanari [35] and Schaefer [39]. A standout result from this era is that of Schaefer who showed that the CSPs arising from constraint languages over 2-element domains satisfy a *dichotomy.* The decision problems for fixed template CSPs over finite domains belong to the class NP, and Schaefer showed that in the 2-element domain case, a constraint language is either solvable in polynomial time or NP-complete. Dichotomies cannot be expected for decision problems in general, since (under the assumption that P$\neq$NP) there are many problems in NP that are neither solvable in polynomial time, nor NP-complete [31]. Another important dichotomy was proved by Hell and Nešetřil [19]. They showed that if a fixed template is a finite simple graph (the vertices make up the domain and the edge relation is the only allowed constraint), then the corresponding CSP is either polynomial time solvable or NP-complete. The decision problem for a graph constraint language can be rephrased as a graph homomorphism problem (a graph homomorphism is a function from the vertices of one graph to another such that the edges are preserved). Specifically, given a fixed graph $\mathcal{H}$ (the constraint language), an instance is a graph $\mathcal{G}$ together with the question "Is there a graph homomorphism from $\mathcal{G}$ to $\mathcal{H}$?". In this sense, 3-colorability corresponds to $\mathcal{H}$ being the complete graph on 3 vertices. The notion of graph homomorphism problems naturally extends to directed graph (digraph) homomorphism problems and to relational structure homomorphism problems.

These early examples of dichotomies, by Schaefer, Hell and Nešetřil, form the basis of a larger project of classifying the complexity of fixed template CSPs. Of particular importance in this project is to prove the so-called *CSP dichotomy conjecture* of Feder and Vardi [18] dating back to 1993. It states that the CSPs related to a fixed constraint language over a finite domain are either polynomial time solvable or NP-complete. To date this conjecture remains unanswered, but it has driven major advances in the study of CSPs.

One such advance is the algebraic connection revealed by Jeavons, Cohen and Gyssens [27] and later refined by Bulatov, Jeavons and Krokhin [9]. This connection associates with each finite domain constraint language $\mathbb{A}$ a finite algebraic structure, the so-called *algebra of polymorphisms*. The properties of this algebraic structure are deeply linked with the computational complexity of the constraint language. In particular, for a fixed core constraint language $\mathbb{A}$, if the algebra of polymorphisms of $\mathbb{A}$ does not satisfy a certain natural property, sometimes called being *Taylor*, then the class of problems determined by $\mathbb{A}$ is NP-complete. Bulatov, Jeavons and Krokhin [9] go on to conjecture that all constraint languages (over finite domains) whose algebras of polymorphisms are Taylor determine polynomial time CSPs (a stronger form of the CSP dichotomy conjecture, since it describes where the split between polynomial time and NP-completeness lies). This conjecture is often referred to as the *Algebraic CSP dichotomy conjecture*.

Many important results have been built upon this algebraic connection. Bulatov [10] extended Schaefer's [39] result on 2-element domains to prove the CSP dichotomy conjecture for 3-element domains. Barto, Kozik and Niven [4] extended Hell and Nešetřil's result [19] on simple graphs to constraint languages consisting of a finite digraph with no sources and no sinks. Barto and Kozik [3] gave a complete algebraic description of the constraint languages over finite domains that are solvable by local consistency methods (these problems are said to be of *bounded width*) and as a consequence it is decidable to determine whether a constraint language can be solved by such methods.

The algebraic approach was also succesfully applied to study finer complexity classification of CSPs. Larose and Tesson [33] conjectured a natural algebraic characterization of templates giving rise to CSPs solvable in logspace (L) and in nondeterministic logspace (NL). In both cases they established the hardness part of the conjecture.

In their seminal paper, Feder and Vardi [18] not only conjectured a P vs. NP-complete dichotomy, they also reduced the problem of proving the dichotomy conjecture to the particular case of digraph homomorphism problems, and even to digraph homomorphism problems where the digraph is balanced (here balanced means that its vertices can be partitioned into levels). Specifically, for every template $\mathbb{A}$ (a finite relational structure of finite type) there is a balanced digraph

$\mathcal{D}(\mathbb{A})$ such that the CSP over $\mathbb{A}$ is polynomial time equivalent to that over $\mathcal{D}(\mathbb{A})$.

In this paper we present a variant of such a construction and prove that (under our construction) CSP over $\mathcal{D}(\mathbb{A})$ is *logspace* equivalent to CSP over $\mathbb{A}$ and that the algebra of polymorphisms of the digraph $\mathcal{D}(\mathbb{A})$ retains almost all relevant properties. For example, $\mathcal{D}(\mathbb{A})$ has bounded width, if and only if $\mathbb{A}$ does. In particular, it follows that the Algebraic CSP dichotomy conjecture, the conjectures characterizing CSPs in L and NL as well as other open questions reduce to the case of digraphs.

In a conference version of this article [12], the authors showed that the Algebraic CSP dichotomy conjecture is equivalent to its restriction to the case of digraphs. This was established by showing that our construction preserves a particular kind of algebraic property, namely existence of a *weak near-unanimity* polymorphism.

Organization of the paper

In Section 1 we present the main results of this paper. Section 2 introduces our notation and the necessary notions concerning relational structures, digraphs and the algebraic approach to the CSP. In Section 3 we describe the construction of $\mathcal{D}(\mathbb{A})$. Sections 4 and 5 are devoted to proving that the construction preserves cores and a large part of the equational properties satisfied by the algebra of polymorphisms. Section 6 contains the logspace reduction of $\mathrm{CSP}(\mathcal{D}(\mathbb{A}))$ to $\mathrm{CSP}(\mathbb{A})$. In Section 7 we discuss a few applications of our result and related open problems.

1 The main results

In general, fixed template CSPs can be modelled as relational structure homomorphism problems [18]. For detailed definitions of relational structures, homomorphisms and other notions used in this section, see Section 2.

Let $\mathbb{A}$ be a finite structure with signature $\mathcal{R}$ (the fixed template). Then the *constraint satisfaction problem for* $\mathbb{A}$ is the following decision problem.

Constraint satisfaction problem for $\mathbb{A}$.

CSP($\mathbb{A}$)
INSTANCE: A finite $\mathcal{R}$-structure $\mathbb{X}$.
QUESTION: Is there a homomorphism from $\mathbb{X}$ to $\mathbb{A}$?

The dichotomy conjecture [18] can be stated as follows:

CSP dichotomy conjecture. *Let $\mathbb{A}$ be a finite relational structure. Then* $\mathrm{CSP}(\mathbb{A})$ *is solvable in polynomial time or NP-complete.*

Every finite relational structure $\mathbb{A}$ has a unique *core* substructure $\mathbb{A}'$ (see Section 2.4 for the precise definition) such that $\mathrm{CSP}(\mathbb{A})$ and $\mathrm{CSP}(\mathbb{A}')$ are identical problems, i.e., the "yes" and "no" instances are precisely the same. The algebraic dichotomy conjecture [9] is the following:

Algebraic CSP dichotomy conjecture. *Let $\mathbb{A}$ be a finite relational structure that is a core. If the algebra of polymorphisms of $\mathbb{A}$ is Taylor, then* $\mathrm{CSP}(\mathbb{A})$ *is solvable in polynomial time, otherwise* $\mathrm{CSP}(\mathbb{A})$ *is NP-complete.*

Indeed, perhaps the above conjecture should be called the *algebraic tractability conjecture* since it is known that if the algebra of polymorphisms of a core $\mathbb{A}$ is not Taylor, then CSP($\mathbb{A}$) is NP-complete [9].

Larose and Tesson [33] conjectured a similar characterization of finite relational structures with the corresponding CSP solvable in L and in NL. In the same paper they also proved the hardness part of boths claims. Their conjecture is widely discussed in the following slightly stronger form (equivalent modulo reasonable complexity-theoretic assumptions; see the discussion in [25]).

Finer CSP complexity conjectures. *Let $\mathbb{A}$ be a finite relational structure that is a core. Then the following hold.*

(i) CSP($\mathbb{A}$) *is solvable in nondeterministic logspace, if and only if the algebra of polymorphisms of $\mathbb{A}$ is congruence join-semidistributive.*

(ii) CSP($\mathbb{A}$) *is solvable in logspace, if and only if the algebra of polymorphisms of $\mathbb{A}$ is congruence join-semidistributive and congruence n-permutable for some n.*

Feder and Vardi [18] proved that every fixed template CSP is polynomial time equivalent to a digraph CSP. Thus the CSP dichotomy conjecture is equivalent to its restriction to digraphs. In this paper we investigate a construction similar to theirs. The main results of this paper are summarized in the following theorem.

Theorem 1.1. *For every finite relational structure $\mathbb{A}$ there exists a finite digraph $\mathcal{D}(\mathbb{A})$ such that the following holds:*

(i) CSP($\mathbb{A}$) *and* CSP($\mathcal{D}(\mathbb{A})$) *are logspace equivalent.*

(ii) $\mathbb{A}$ *is a core if and only if* $\mathcal{D}(\mathbb{A})$ *is a core.*

(iii) If Σ is a linear idempotent set of identities such that the algebra of polymorphisms of the oriented path $\bullet \to \bullet \leftarrow \bullet \to \bullet$ satisfies Σ and each identity in Σ is either balanced or contains at most two variables, then

$$\mathbb{A} \models \Sigma \text{ if and only if } \mathcal{D}(\mathbb{A}) \models \Sigma.$$

Proof. Item (i) is Theorem 6.1, (ii) is Corollary 4.2 and (iii) is Theorem 5.1. □

The construction of $\mathcal{D}(\mathbb{A})$ is described in Section 3, for a bound on the size of $\mathcal{D}(\mathbb{A})$ see Remark 3. The condition on Σ in item (iii) is not very restrictive: it includes almost all of the commonly encountered properties relevant to the CSP. A number of these are listed in Corollary 5.2. Note that the list includes the properties of being Taylor, congruence join-semidistributive and congruence n-permutable (for $n \geq 3$); hence we have the following corollary.

Corollary 1.2. *The Algebraic CSP dichotomy conjecture and the Finer CSP complexity conjectures are also equivalent to their restrictions to digraphs.*

2 Background and definitions

We approach fixed template constraint satisfaction problems from the "homomorphism problem" point of view. For background on the homomorphism approach to CSPs, see [18], and for background on the algebraic approach to CSPs, see [9].

A *relational signature* $\mathcal{R}$ is a (in our case finite) set of *relation symbols* R_i, each with an associated arity k_i. A (finite) *relational structure* $\mathbb{A}$ *over relational signature* $\mathcal{R}$ (called an $\mathcal{R}$*-structure*) is a finite set A (the *domain*) together with a relation $R_i \subseteq A^{k_i}$, for each relation symbol R_i of arity k_i in $\mathcal{R}$. A *CSP template* is a fixed finite $\mathcal{R}$-structure, for some signature $\mathcal{R}$.

For simplicity we do not distinguish the relation with its associated relation symbol. However, to avoid ambiguity, we sometimes write $R^{\mathbb{A}}$ to indicate that R is interpreted in $\mathbb{A}$. We will often refer to the domain of a relational structure $\mathbb{A}$ simply by A. When referring to a fixed relational structure, we may simply specify it as $\mathbb{A} = (A; R_1, R_2, \dots, R_n)$. For technical reasons we require that signatures are nonempty and that all the relations of a relational structure are nonempty.

2.1 Notation

For a positive integer n we denote the set $\{1, 2, \dots, n\}$ by $[n]$. We write tuples using boldface notation, e.g. $\mathbf{a} = (a_1, a_2, \dots, a_k) \in A^k$ and when ranging over tuples we use superscript notation, e.g. $(\mathbf{r}^1, \mathbf{r}^2, \dots, \mathbf{r}^l) \in R^l \subseteq (A^k)^l$, where $\mathbf{r}^i = (r_1^i, r_2^i, \dots, r_k^i)$, for $i = 1, \dots, l$.

Let $R_i \subseteq A^{k_i}$ be relations of arity k_i, for $i = 1, \dots, n$. Let $k = \sum_{i=1}^{n} k_i$ and $l_i = \sum_{j<i} k_j$. We write $R_1 \times \cdots \times R_n$ to mean the k-ary relation

$$\{(a_1, \dots, a_k) \in A^k \mid (a_{l_i+1}, \dots, a_{l_i+k_i}) \in R_i \text{ for } i = 1, \dots, n\}.$$

An *n-ary operation* on a set A is simply a mapping $f : A^n \to A$; the number n is the *arity* of f. Let f be an n-ary operation on A and let $k > 0$. We write $f^{(k)}$ to denote the n-ary operation obtained by applying f coordinatewise on A^k. That is, we define the n-ary operation $f^{(k)}$ on A^k by

$$f^{(k)}(\mathbf{a}^1, \dots, \mathbf{a}^n) = (f(a_1^1, \dots, a_1^n), \dots, f(a_k^1, \dots, a_k^n)),$$

for $\mathbf{a}^1, \dots, \mathbf{a}^n \in A^k$.

We will be particularly interested in so-called idempotent operations. An n-ary operation f is said to be *idempotent* if it satisfies the equation $f(x, x, \dots, x) = x$.

2.2 Homomorphisms, cores and polymorphisms

We begin with the notion of a relational structure homomorphism.

Definition 2.1. *Let $\mathbb{A}$ and $\mathbb{B}$ be relational structures in the same signature $\mathcal{R}$. A* homomorphism *from $\mathbb{A}$ to $\mathbb{B}$ is a mapping φ from A to B such that for each k-ary relation symbol R in $\mathcal{R}$ and each k-tuple $\mathbf{a} \in A^k$, if $\mathbf{a} \in R^{\mathbb{A}}$, then $\varphi^{(k)}(\mathbf{a}) \in R^{\mathbb{B}}$.*

We write $\varphi : \mathbb{A} \to \mathbb{B}$ to mean that φ is a homomorphism from $\mathbb{A}$ to $\mathbb{B}$, and $\mathbb{A} \to \mathbb{B}$ to mean that there exists a homomorphism from $\mathbb{A}$ to $\mathbb{B}$.

An *isomorphism* is a bijective homomorphism φ such that φ^{-1} is also a homomorphism. A homomorphism $\mathbb{A} \to \mathbb{A}$ is called an *endomorphism.* An isomorphism from $\mathbb{A}$ to $\mathbb{A}$ is an *automorphism.* It is an easy fact that if $\mathbb{A}$ is finite, then every surjective endomorphism is an automorphism.

A finite relational structure $\mathbb{A}'$ is a *core* if every endomorphism $\mathbb{A}' \to \mathbb{A}'$ is surjective (and therefore an automorphism). For every $\mathbb{A}$ there exists a relational structure $\mathbb{A}'$ such that $\mathbb{A} \to \mathbb{A}'$ and $\mathbb{A}' \to \mathbb{A}$ and $\mathbb{A}'$ is minimal with respect to these properties; that structure $\mathbb{A}'$ is called the *core of* $\mathbb{A}$. The core of $\mathbb{A}$ is unique (up to isomorphism) and $\mathrm{CSP}(\mathbb{A})$ and $\mathrm{CSP}(\mathbb{A}')$ are the same decision problems. Equivalently, the core of $\mathbb{A}$ can be defined as a minimal induced substructure that $\mathbb{A}$ retracts onto. (See [20] for details on cores for graphs, cores for relational structures are a natural generalization.)

The notion of *polymorphism* is central in the so-called algebraic approach to CSP. Polymorphisms are a natural generalization of endomorphisms to higher arity operations.

Definition 2.2. *Given an $\mathcal{R}$-structure $\mathbb{A}$, an n-ary* polymorphism *of $\mathbb{A}$ is an n-ary operation f on A such that f preserves the relations of $\mathbb{A}$. That is, if $\mathbf{a}^1, \ldots, \mathbf{a}^n \in R$, for some k-ary relation R in $\mathcal{R}$, then $f^{(k)}(\mathbf{a}^1, \ldots, \mathbf{a}^n) \in R$.*

Thus, an endomorphism is a unary polymorphism. Polymorphisms satisfying certain identities has been used extensively in the algebraic study of CSPs.

2.3 Algebra

Given a finite relational structure $\mathbb{A}$, let $\operatorname{Pol} \mathbb{A}$ denote the set of all polymorphisms of $\mathbb{A}$. The *algebra of polymorphisms* of $\mathbb{A}$ is simply the algebra with the same universe whose operations are all polymorphisms of $\mathbb{A}$. A subset $B \subseteq A$ is a *subuniverse* of $\mathbb{A}$, denoted by $B \leq \mathbb{A}$, if it is a subuniverse of the algebra of polymorphisms of $\mathbb{A}$, i.e., it is closed under all $f \in \operatorname{Pol} \mathbb{A}$.

An *(operational) signature* is a (possibly infinite) set of operation symbols with arities assigned to them. By an *identity* we mean an expression $u \approx v$ where u and v are terms in some signature. An identity $u \approx v$ is *linear* if both u and v involve at most one occurrence of an operation symbol (e.g. $f(x, y) \approx g(x)$, or $h(x, y, x) \approx x$); and *balanced* if the sets of variables occuring in u and in v are the same (e.g. $f(x, x, y) \approx g(y, x, x)$).

A set of identities Σ is *linear* if it contains only linear identities; *balanced* if all the identities in Σ are balanced; and *idempotent* if for each operation symbol f appearing in an identity of Σ, the identity $f(x, x, \ldots, x) \approx x$ is in Σ. [1] For example, the identities $p(y, x, x) \approx y,\ p(x, x, y) \approx y,\ p(x, x, x) \approx x$ (defining the so-called *Maltsev* operation) form a linear idempotent set of identities which is not balanced.

The *strong Maltsev condition*, a notion usual in universal algebra, can be defined in this context as a finite set of identities. A *Maltsev condition* is an increasing chain of strong Maltsev conditions, ordered by syntactical consequence. In all results from this paper, "set of identities" can be replaced with "Maltsev condition".

[1] We can relax this condition and require the identity $f(x, x, \ldots, x) \approx x$ only to be a *syntactical consequence* of identities in Σ.

Let Σ be a set of identities in a signature with operation symbols $\mathcal{F} = \{f_\lambda \mid \lambda \in \Lambda\}$. We say that a relational structure $\mathbb{A}$ *satisfies* Σ (and write $\mathbb{A} \models \Sigma$), if for every $\lambda \in \Lambda$ there is a polymorphism $f_\lambda^{\mathbb{A}} \in \operatorname{Pol} \mathbb{A}$ such that the identities in Σ hold universally in $\mathbb{A}$ when for each $\lambda \in \Lambda$ the symbol f_λ is interpreted as $f_\lambda^{\mathbb{A}}$.

For example, a *weak near-unanimity* (*WNU*) is an n-ary ($n \geq 2$) idempotent operation ω satisfying the identities

$$\omega(x, \dots, x, y) = \omega(x, \dots, x, y, x) = \cdots = \omega(y, x, \dots, x).$$

Thus, having an n-ary weak near-unanimity is definable by a linear balanced idempotent set of identities. Existence of WNU polymorphisms influences $\mathrm{CSP}(\mathbb{A})$ to a great extent. The following characterization was discovered in [34]: a finite algebra (or relational structure) is called

- *Taylor*, if it has a weak near-unanimity operation of some arity, and
- *congruence meet-semidistributive* if it has WNU operations of all but finitely many arities.

The Algebraic CSP dichotomy conjecture asserts that being Taylor is what distinguishes tractable (core) relational structures from the NP-complete ones, and a similar split is known for congruence meet-semidistributivity and solvability by local consistency checking (the so-called *bounded width*):

Bounded width theorem. [3] *Let $\mathbb{A}$ be a finite relational structure that is a core. Then* $\mathrm{CSP}(\mathbb{A})$ *is solvable by local consistency checking, if and only if the algebra of polymorphisms of $\mathbb{A}$ is congruence meet-semidistributive.*

The properties of *congruence join-semidistributivity* as well as *congruence n-permutability*, which appear in the finer CSP complexity conjectures, are also definable by linear idempotent sets of identities, albeit more complicated ones; we refer the reader to [23]. We will introduce more Maltsev conditions and their connection to the CSP in Section 5.

2.4 Primitive positive definability

A first order formula is called *primitive positive* if it is an existential conjunction of atomic formulæ. Since we only refer to relational signatures, a primitive positive formula is simply an existential conjunct of formulæ of the form $x = y$ or $(x_1, x_2, \dots, x_k) \in R$, where R is a relation symbol of arity k.

For example, if we have a binary relation symbol E in our signature, then the formula

$$\psi(x, y) = (\exists z)((x, z) \in E \ \wedge \ (z, y) \in E)$$

pp-defines a binary relation in which elements a, b are related if there is a directed path of length 2 from a to b in E.

Definition 2.3. *A relational structure $\mathbb{B}$ is* primitive positive definable *in $\mathbb{A}$ (or $\mathbb{A}$* pp-defines *$\mathbb{B}$) if*

(i) the set B is a subset of A and is definable by a primitive positive formula interpreted in $\mathbb{A}$, and

(ii) each relation R in the signature of $\mathbb{B}$ is definable on the set B by a primitive positive formula interpreted in $\mathbb{A}$.

The following result relates the above definition to the complexity of CSPs. The connection is originally due to Jeavons, Cohen and Gyssens [27], though the logspace form stated and used here can be found in Larose and Tesson [33, Theorem 2.1].

Lemma 2.4. *Let $\mathbb{A}$ be a finite relational structure that pp-defines $\mathbb{B}$. Then,* $\mathrm{CSP}(\mathbb{B})$ *is logspace reducible to* $\mathrm{CSP}(\mathbb{A})$*.*

It so happens that, if $\mathbb{A}$ pp-defines $\mathbb{B}$, then $\mathbb{B}$ inherits the polymorphisms of $\mathbb{A}$. See [9] for a detailed explanation.

Lemma 2.5. [9] *Let $\mathbb{A}$ be a finite relational structure that pp-defines $\mathbb{B}$. If φ is a polymorphism of $\mathbb{A}$, then its restriction to B is a polymorphism of $\mathbb{B}$.*

In particular, as an easy consequence of this lemma, if $\mathbb{A}$ pp-defines $\mathbb{B}$ and $\mathbb{A}$ satisfies a set of identities Σ, then $\mathbb{B}$ also satisfies Σ.

In the case that $\mathbb{A}$ pp-defines $\mathbb{B}$ and $\mathbb{B}$ pp-defines $\mathbb{A}$, we say that $\mathbb{A}$ and $\mathbb{B}$ are *pp-equivalent*. In this case, $\mathrm{CSP}(\mathbb{A})$ and $\mathrm{CSP}(\mathbb{B})$ are essentially the same problems (they are logspace equivalent) and $\mathbb{A}$ and $\mathbb{B}$ have the same polymorphisms.

Example 2.6. *Let $\mathbb{A} = (A; R_1, \ldots, R_n)$, where each R_i is k_i-ary, and define $R = R_1 \times \cdots \times R_n$. Then the structure $\mathbb{A}' = (A; R)$ is pp-equivalent to $\mathbb{A}$.*

Indeed, let $k = \sum_{i=1}^{n} k_i$ be the arity of R and $l_i = \sum_{j<i} k_j$ for $i = 1, \ldots, n$. The relation R is pp-definable from $R_1, \ldots, R_n$ using the formula

$$\Psi(x_1, \ldots, x_k) = \bigwedge_{i=1}^{n} (x_{l_i+1}, \ldots, x_{l_i+k_i}) \in R_i.$$

The relation R_1 can be defined from R by the primitive positive formula

$$\Psi(x_1, \ldots, x_{k_1}) = (\exists y_{k_1+1}, \ldots, \exists y_k)((x_1, \ldots, x_{k_1}, y_{k_1+1}, \ldots, y_k) \in R)$$

and the remaining R_i's can be defined similarly.

Example 2.6 shows that when proving Theorem 1.1 we can restrict ourselves to relational structures with a single relation.

2.5 Digraphs

A *directed graph*, or *digraph*, is a relational structure $\mathbb{G}$ with a single binary relation symbol E as its signature. We typically call the members of G and $E^{\mathbb{G}}$ *vertices* and *edges*, respectively. We usually write $a \to b$ to mean $(a, b) \in E^{\mathbb{G}}$, if there is no ambiguity.

A special case of relational structure homomorphism (see Definition 2.1), is that of digraph homomorphism. That is, given digraphs $\mathbb{G}$ and $\mathbb{H}$, a function $\varphi : G \to H$ is a homomorphism if $(\varphi(a), \varphi(b)) \in E^{\mathbb{H}}$ whenever $(a, b) \in E^{\mathbb{G}}$.

Definition 2.7. *For $i = 1, \ldots, n$, let $\mathbb{G}_i = (G_i, E_i)$ be digraphs. The* direct product *of $\mathbb{G}_1, \ldots, \mathbb{G}_n$, denoted by $\prod_{i=1}^n \mathbb{G}_i$, is the digraph with vertices $\prod_{i=1}^n G_i$ (the cartesian product of the sets G_i) and edge relation*

$$\{(\mathbf{a}, \mathbf{b}) \in (\prod_{i=1}^{n} G_i)^2 \mid (a_i, b_i) \in E_i \text{ for } i = 1 \ldots, n\}.$$

If $\mathbb{G}_1 = \cdots = \mathbb{G}_n = \mathbb{G}$ then we write $\mathbb{G}^n$ to mean $\prod_{i=1}^n \mathbb{G}_i$.

With the above definition in mind, an n-ary polymorphism on a digraph $\mathbb{G}$ is simply a digraph homomorphism from $\mathbb{G}^n$ to $\mathbb{G}$.

Definition 2.8. *A digraph $\mathbb{P}$ is an* oriented path *if it consists of a sequence of vertices $v_0, v_1, \ldots, v_k$ such that precisely one of $(v_{i-1}, v_i), (v_i, v_{i-1})$ is an edge, for each $i = 1, \ldots, k$. We require oriented paths to have a direction; we denote the* initial *vertex v_0 and the* terminal *vertex v_k by $\iota\mathbb{P}$ and $\tau\mathbb{P}$, respectively.*

Given a digraph $\mathbb{G}$ and an oriented path $\mathbb{P}$, we write $a \xrightarrow{\mathbb{P}} b$ to mean that we can walk in $\mathbb{G}$ from a following $\mathbb{P}$ to b, i.e., there exists a homomorphism $\varphi : \mathbb{P} \to \mathbb{G}$ such that $\varphi(\iota\mathbb{P}) = a$ and $\varphi(\tau\mathbb{P}) = b$. Note that for every $\mathbb{P}$ there exists a primitive positive formula $\psi(x, y)$ such that $a \xrightarrow{\mathbb{P}} b$ if and only if $\psi(a, b)$ is true in $\mathbb{G}$. If there exists an oriented path $\mathbb{P}$ such that $a \xrightarrow{\mathbb{P}} b$, we say that a and b are *connected*. If vertices a and b are connected, then the *distance* from a to b is the number of edges in the shortest oriented path connecting them. Connectedness forms an equivalence relation on G; its classes are called the *connected components* of $\mathbb{G}$. We say that a digraph is connected if it consists of a single connected component.

A connected digraph is *balanced* if it admits a *level function* $\mathrm{lvl} : G \to \mathbb{N} \cup \{0\}$, where $\mathrm{lvl}(b) = \mathrm{lvl}(a) + 1$ whenever (a, b) is an edge, and the minimum level is 0. The maximum level is called the *height* of the digraph. Oriented paths are natural examples of balanced digraphs.

By a *zigzag* we mean the oriented path $\bullet \to \bullet \leftarrow \bullet \to \bullet$ and a *single edge* is the path $\bullet \to \bullet$. For oriented paths $\mathbb{P}$ and $\mathbb{P}'$, the *concatenation of* $\mathbb{P}$ *and* $\mathbb{P}'$, denoted by $\mathbb{P} \dotplus \mathbb{P}'$, is the oriented path obtained by identifying $\tau\mathbb{P}$ with $\iota\mathbb{P}'$.

Our digraph reduction as described in Section 3 relies on oriented paths obtained by concatenation of zigzags and single edges. For example, the path in Figure 1 is a concatenation of a single edge followed by two zigzags and two more single edges (for clarity, we organize its vertices into levels).

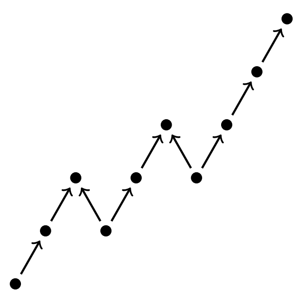


Figure 1: A minimal oriented path

3 The reduction to digraphs

In this section we take an arbitrary finite relational structure $\mathbb{A}$ and construct a balanced digraph $\mathcal{D}(\mathbb{A})$ such that CSP($\mathbb{A}$) and CSP($\mathcal{D}(\mathbb{A})$) are logspace equivalent.

Let $\mathbb{A} = (A; R_1, \dots, R_n)$ be a finite relational structure, where R_i is of arity k_i, for $i = 1, \dots, n$. Let $k = \sum_{i=1}^{n} k_i$ and let R be the k-ary relation $R_1 \times \cdots \times R_n$. For $\mathcal{I} \subseteq [k]$ define $\mathbb{Q}_{\mathcal{I},l}$ to be a single edge if $l \in \mathcal{I}$, and a zigzag if $l \in [k] \setminus \mathcal{I}$.

We define the oriented path $\mathbb{Q}_{\mathcal{I}}$ (of height $k+2$) by

$$\mathbb{Q}_{\mathcal{I}} = \bullet \to \bullet \dotplus \mathbb{Q}_{\mathcal{I},1} \dotplus \mathbb{Q}_{\mathcal{I},2} \dotplus \dots \dotplus \mathbb{Q}_{\mathcal{I},k} \dotplus \bullet \to \bullet$$

Instead of $\mathbb{Q}_{\emptyset}, \mathbb{Q}_{\emptyset,l}$ we write just $\mathbb{Q}, \mathbb{Q}_l$, respectively. For example, the oriented path in Figure 1 is $\mathbb{Q}_{\mathcal{I}}$ where $k = 3$ and $\mathcal{I} = \{3\}$. We will need the following observation.

Observation. *Let $\mathcal{I}, \mathcal{J} \subseteq [k]$. A homomorphism $\varphi : \mathbb{Q}_{\mathcal{I}} \to \mathbb{Q}_{\mathcal{J}}$ exists, if and only if $\mathcal{I} \subseteq \mathcal{J}$. In particular $\mathbb{Q} \to \mathbb{Q}_{\mathcal{I}}$ for all $\mathcal{I} \subseteq [k]$. Moreover, if φ exists, it is unique and surjective.*

We are now ready to define the digraph $\mathcal{D}(\mathbb{A})$.

Definition 3.1. *For every $e = (a, \mathbf{r}) \in A \times R$ we define $\mathbb{P}_e$ to be the path $\mathbb{Q}_{\{i \,|\, a = r_i\}}$. The digraph $\mathcal{D}(\mathbb{A})$ is obtained from the digraph $(A \cup R; A \times R)$ by replacing every $e = (a, \mathbf{r}) \in A \times R$ by the oriented path $\mathbb{P}_e$ (identifying $\iota\mathbb{P}_e$ with a and $\tau\mathbb{P}_e$ with $\mathbf{r}$).*

(We often write $\mathbb{P}_{e,l}$ to mean $\mathbb{Q}_{\mathcal{I},l}$ where $\mathbb{P}_e = \mathbb{Q}_{\mathcal{I}}$.)

Example 3.2. *Consider the relational structure $\mathbb{A} = (\{0,1\}; R)$ where $R = \{(0,1), (1,0)\}$, i.e., $\mathbb{A}$ is the directed 2-cycle. Figure 2 is a visual representation of $\mathcal{D}(\mathbb{A})$.*

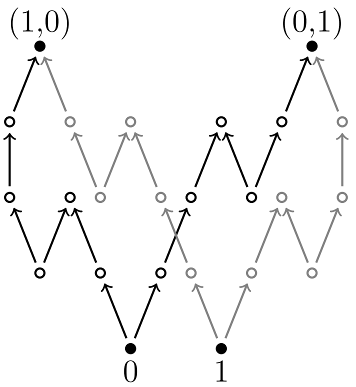


Figure 2: $\mathcal{D}(\mathbb{A})$ where $\mathbb{A}$ is the directed 2-cycle

Remark. *The number of vertices in $\mathcal{D}(\mathbb{A})$ is $(3k+1)|R||A| + (1-2k)|R| + |A|$ and the number of edges is $(3k+2)|R||A| - 2k|R|$. The construction of $\mathcal{D}(\mathbb{A})$ can be performed in logspace (under any reasonable encoding).*

Proof. The vertices of $\mathcal{D}(\mathbb{A})$ consist of the elements of $A \cup R$, along with vertices from the connecting paths. The number of vertices lying strictly within the connecting paths would be $(3k+1)|R||A|$ if every $\mathbb{P}_e$ was $\mathbb{Q}$. We need to deduct 2 vertices whenever there is a single edge instead of a zigzag and there are $\sum_{(a,\mathbf{r})\in A\times R} |\{i \mid a = r_i\}| = k|R|$ such instances. The number of edges is counted very similarly. $\square$

Remark. *Note that if we apply this construction to itself (that is, $\mathcal{D}(\mathcal{D}(\mathbb{A}))$) then we obtain balanced digraphs of height* 4. *When applied to digraphs, the $\mathcal{D}$ construction is identical to that given by Feder and Vardi [18, Theorem 13].*

The following lemma, together with Lemma 2.4, shows that CSP($\mathbb{A}$) reduces to CSP($\mathcal{D}(\mathbb{A})$) in logspace.

Lemma 3.3. $\mathbb{A}$ *is pp-definable from* $\mathcal{D}(\mathbb{A})$.

Proof. Example 2.6 demonstrates that $\mathbb{A}$ is pp-equivalent to $(A; R)$. We now show that $\mathcal{D}(\mathbb{A})$ pp-defines $(A; R)$, from which it follows that $\mathcal{D}(\mathbb{A})$ pp-defines $\mathbb{A}$.

Note that $\mathbb{Q} \to \mathbb{P}_e$ for all $e \in A \times R$, and $\mathbb{Q}_{\{i\}} \to \mathbb{P}_{(a,\mathbf{r})}$ if and only if $a = r_i$. The set A is pp-definable in $\mathcal{D}(\mathbb{A})$ by $A = \{x \mid (\exists y)(x \xrightarrow{\mathbb{Q}} y)\}$ and the relation R can be defined as the set $\{(x_1, \dots, x_k) \mid (\exists y)(x_i \xrightarrow{\mathbb{Q}_{\{i\}}} y \text{ for all } i \in [k])\}$, which is also a primitive positive definition. $\square$

It is not, in general, possible to pp-define $\mathcal{D}(\mathbb{A})$ from $\mathbb{A}$.[1] Nonetheless the following lemma is true.

Lemma 3.4. CSP($\mathcal{D}(\mathbb{A})$) *reduces in logspace to* CSP($\mathbb{A}$).

The proof of Lemma 3.4 is rather technical, though broadly follows the polynomial process described in the proof of [18, Theorem 13] (as mentioned, our construction coincides with theirs in the case of digraphs). Details of the argument are provided in Section 6.

4 Preserving cores

In what follows, let $\mathbb{A}$ be a fixed finite relational structure. Without loss of generality we may assume that $\mathbb{A} = (A; R)$, where R is a k-ary relation (see Example 2.6).

Lemma 4.1. *The endomorphisms of* $\mathbb{A}$ *and* $\mathcal{D}(\mathbb{A})$ *are in one-to-one correspondence.*

Proof. We first show that every endomorphism φ of $\mathbb{A}$ can be extended to an endomorphism $\overline{\varphi}$ of $\mathcal{D}(\mathbb{A})$. Let $\overline{\varphi}(a) = \varphi(a)$ for $a \in A$, and let $\overline{\varphi}(\mathbf{r}) = \varphi^{(k)}(\mathbf{r})$ for $\mathbf{r} \in R$. Note that $\varphi^{(k)}(\mathbf{r}) \in R$ since φ is an endomorphism of $\mathbb{A}$.

Let $c \in \mathcal{D}(\mathbb{A}) \setminus (A \cup R)$ and let $e = (a, \mathbf{r})$ be such that $c \in \mathbb{P}_e$. Define $e' = (\varphi(a), \varphi^{(k)}(\mathbf{r}))$. If $\mathbb{P}_{e,l}$ is a single edge for some $l \in [k]$, then $r_l = a$ and

[1] Using the definition of pp-definability as described in this paper, this is true for cardinality reasons. However, a result of Kazda [28] can be used to show that the statement remains true even for more general definitions of pp-definability.

$\varphi(r_l) = \varphi(a)$, and therefore $\mathbb{P}_{e',l}$ is a single edge. Thus there exists a (unique) homomorphism $\mathbb{P}_e \to \mathbb{P}_{e'}$. Define $\overline{\varphi}(c)$ to be the image of c under this homomorphism, completing the definition of $\overline{\varphi}$.

We now show that every endomorphism Φ of $\mathcal{D}(\mathbb{A})$ is of the form $\overline{\varphi}$, for some endomorphism φ of $\mathbb{A}$. Let Φ be an endomorphism of $\mathcal{D}(\mathbb{A})$. Let φ be the restriction of Φ to A. By Lemma 2.5 and Lemma 3.3, φ is an endomorphism of $\mathbb{A}$. For every $e = (a, \mathbf{r})$, the endomorphism Φ maps $\mathbb{P}_e$ onto $\mathbb{P}_{(\varphi(a),\Phi(\mathbf{r}))}$. If we set $a = r_l$, then $\mathbb{P}_{e,l}$ is a single edge. In this case it follows that $\mathbb{P}_{(\varphi(a),\Phi(\mathbf{r})),l}$ is also a single edge. Thus, by the construction of $\mathcal{D}(\mathbb{A})$ the l^{th} coordinate of $\Phi(\mathbf{r})$ is $\varphi(a) = \varphi(r_l)$. This proves that the restriction of Φ to R is $\varphi^{(k)}$ and therefore $\Phi = \overline{\varphi}$. □

The following corollary is Theorem 1.1 (ii).

Corollary 4.2. *$\mathbb{A}$ is a core if and only if $\mathcal{D}(\mathbb{A})$ is a core.*

Proof. To prove the corollary we need to show that an endomorphism φ of $\mathbb{A}$ is surjective if and only if $\overline{\varphi}$ (from Lemma 4.1) is surjective. Clearly, if $\overline{\varphi}$ is surjective then so is φ.

Assume φ is surjective (and therefore an automorphism of $\mathbb{A}$). It follows that $\varphi^{(k)}$ is surjective on R and therefore $\overline{\varphi}$ is a bijection when restricted to the set $A \cup R$. Let $a \in A$ and $\mathbf{r} \in R$. By definition we know that $\overline{\varphi}$ maps $\mathbb{P}_{(a,\mathbf{r})}$ homomorphically onto $\mathbb{P}_{(\varphi(a),\varphi^{(k)}(\mathbf{r}))}$. Since φ has an inverse φ^{-1}, it follows that $\overline{\varphi^{-1}}$ maps $\mathbb{P}_{(\varphi(a),\varphi^{(k)}(\mathbf{r}))}$ homomorphically onto $\mathbb{P}_{(a,\mathbf{r})}$. Thus $\mathbb{P}_{(a,\mathbf{r})}$ and $\mathbb{P}_{(\varphi(a),\varphi^{(k)}(\mathbf{r}))}$ are isomorphic, completing the proof. □

Using similar arguments it is not hard to prove a bit more, namely that the *monoids of endomorphisms* of $\mathbb{A}$ and $\mathcal{D}(\mathbb{A})$ are isomorphic. Since endormorphisms are just the unary part of the algebra of polymorphisms, this section can be viewed as a "baby case" to the more involved proof in the next section.

5 Preserving Maltsev conditions

Given a finite relational structure $\mathbb{A}$, we are interested in the following question: How similar are the algebras of polymorphisms of $\mathbb{A}$ and $\mathcal{D}(\mathbb{A})$? More precisely, which equational properties (or *Maltsev conditions*) do they share? In this section we provide a quite broad range of Maltsev conditions that hold equivalently in $\mathbb{A}$ and $\mathcal{D}(\mathbb{A})$. Indeed, to date, these include all Maltsev conditions that are conjectured to divide differing levels of tractability and hardness, as well as all the main tractable algorithmic classes (e.g. few subpowers and bounded width).

5.1 The result

We start by an overview and statement of the main result of this section. Since $\mathbb{A}$ is pp-definable from the digraph $\mathcal{D}(\mathbb{A})$ (see Lemma 3.3), it follows that A and R are subuniverses of $\mathcal{D}(\mathbb{A})$ and for any $f \in \operatorname{Pol}\mathcal{D}(\mathbb{A})$, the restriction $f|_A$ is a polymorphism of $\mathbb{A}$. Consequently, for any set of identities Σ,

$$\mathcal{D}(\mathbb{A}) \models \Sigma \text{ implies that } \mathbb{A} \models \Sigma.$$

The theorem below, which is a restatement of Theorem 1.1 (iii), provides a partial converse of the above implication.

Theorem 5.1. *Let $\mathbb{A}$ be a finite relational structure. Let Σ be a linear idempotent set of identities such that the algebra of polymorphisms of the zigzag satisfies Σ and each identity in Σ is either balanced or contains at most two variables. Then*

$$\mathcal{D}(\mathbb{A}) \models \Sigma \text{ if and only if } \mathbb{A} \models \Sigma.$$

The following corollary lists some popular properties that can be expressed as sets of identities satisfying the above assumptions. Indeed, they include many commonly encountered Maltsev conditions.

Corollary 5.2. *Let $\mathbb{A}$ be a finite relational structure. Then each of the following hold equivalently on (the polymorphism algebra of) $\mathbb{A}$ and $\mathcal{D}(\mathbb{A})$.*

(1) Being Taylor or equivalently having a weak near-unanimity (WNU) operation [34] or equivalently a cyclic operation [2] (conjectured to be equivalent to being in P if $\mathbb{A}$ is a core [9]);

(2) Congruence join-semidistributivity (SD($\vee$)) (conjectured to be equivalent to NL if $\mathbb{A}$ is a core [33]);

(3) (For $n \geq 3$) congruence n-permutability (CnP) (together with (2) conjectured to be equivalent to L if $\mathbb{A}$ is a core [33]).

(4) Congruence meet-semidistributivity (SD($\wedge$)) (equiv. to bounded width [3]);

(5) (For $k \geq 4$) k-ary edge operation (equivalent to few subpowers [5], [24]);

(6) k-ary near-unanimity operation (equivalent to strict width [18]);

(7) Totally symmetric idempotent (TSI) operations of all arities (equivalent to width 1 [16], [18]);

(8) Hobby-McKenzie operations (equivalent to the corresponding variety satisfying a non-trivial congruence lattice identity);

(9) Congruence modularity (CM);

(10) Congruence distributivity (CD);

Items (2) and (3) above, together with Theorem 1.1 (i) and (ii), show that the Finer CSP complexity conjectures need only be established in the case of digraphs to obtain a resolution in the general case.

Note that the above list includes all six conditions for omitting types in the sense of Tame Congruence Theory [23]. Figure 3, taken from [25], presents a diagram of what might be called the "universal algebraic geography of CSPs".

We will prove Theorem 5.1 and Corollary 5.2 in subsection 5.3.

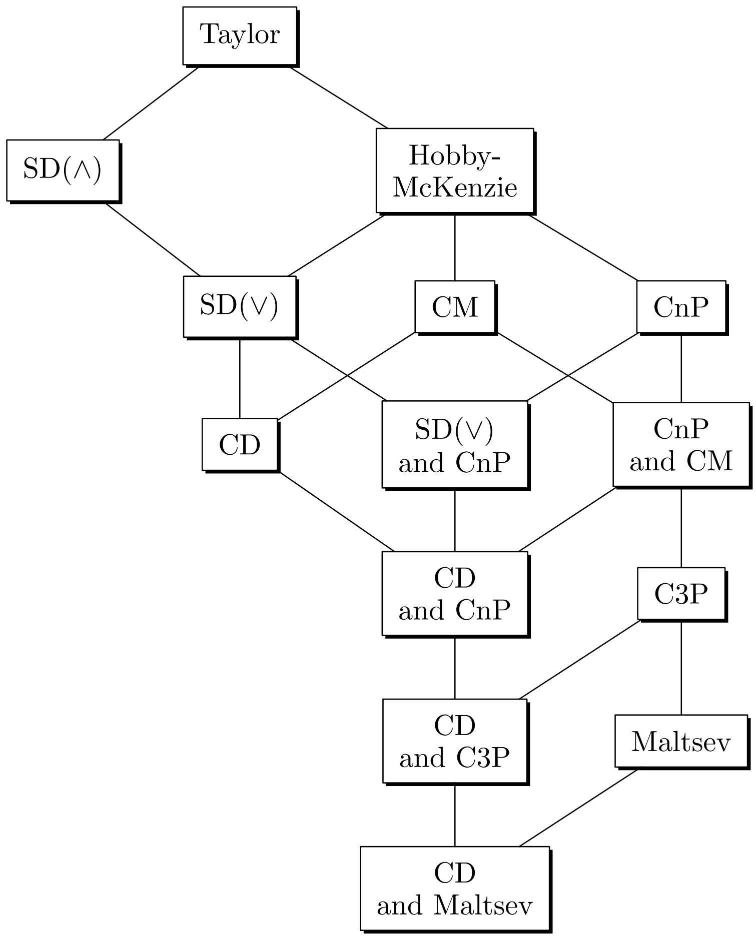


Figure 3: The universal algebraic geography of tractable CSPs.

5.2 Polymorphisms of the zigzag

In the following, let $\mathbb{Z}$ be a zigzag with vertices 00, 01, 10 and 11 (i.e., the oriented path $00 \to 01 \leftarrow 10 \to 11$). Let us denote by $\leq_{\mathbb{Z}}$ the linear order on $\mathbb{Z}$ given by $00 <_{\mathbb{Z}} 01 <_{\mathbb{Z}} 10 <_{\mathbb{Z}} 11$.

Note that the subset $\{00, 10\}$ is closed under all polymorphisms of $\mathbb{Z}$ (as it is pp-definable using the formula $(\exists y)(x \to y)$, see Lemma 2.5). The same holds for $\{01, 11\}$. We will use this fact later in our proof.

The digraph $\mathbb{Z}$ satisfies most of the important Maltsev conditions (an exception being congruence 2-permutability, i.e., having a Maltsev polymorphism). We need the following.

Lemma 5.3. *The following holds.*

(i) $\mathbb{Z}$ *has a majority polymorphism,*

(ii) $\mathbb{Z}$ *satisfies any balanced set of identities,*

(iii) $\mathbb{Z}$ *is congruence 3-permutable.*

Proof. Let $x \wedge y$ and $x \vee y$ denote the binary operations of minimum and maximum with respect to $\leq_{\mathbb{Z}}$, respectively. That is, $x \wedge y$ is the vertex from $\{x, y\}$ closer to 00 and $x \vee y$ the vertex closer to 11. It can be easily seen that $\wedge, \vee$ are polymorphisms of $\mathbb{Z}$ and form a distributive lattice. Note that it follows that $\mathbb{Z}$ satisfies any set of identities which holds in the variety of distributive lattices (equivalently, in the two-element lattice).

In particular, to prove (i), note that the ternary operation defined by

$$m(x, y, z) = (x \wedge y) \vee (x \wedge z) \vee (y \wedge z)$$

(the *median*) is a majority polymorphism. To prove (ii), let Σ be a balanced set of identities. For every operation symbol f (say k-ary) occurring in Σ, we define $f^{\mathbb{Z}}(x_1, \dots, x_k) = \bigwedge_{i=1}^{k} x_i$. It is easy to check that $f^{\mathbb{Z}}$ is a polymorphism and that such a construction satisfies any balanced identity.

To prove (iii), we directly construct the ternary polymorphisms p_1 and p_2 witnessing 3-permutability:

$$p_1(x, y, z) = \begin{cases} 01 & \text{if } y \neq z \text{ and } 01 \in \{x, y, z\}, \\ 10 & \text{if } y \neq z \text{ and } 10 \in \{x, y, z\} \text{ and } 01 \notin \{x, y, z\}, \\ x & \text{otherwise,} \end{cases}$$

$$p_2(x, y, z) = \begin{cases} 01 & \text{if } x \neq y \text{ and } 01 \in \{x, y, z\}, \\ 10 & \text{if } x \neq y \text{ and } 10 \in \{x, y, z\} \text{ and } 01 \notin \{x, y, z\}, \\ z & \text{if } x = y. \\ x & \text{otherwise} \end{cases}$$

The identities $p_1(x, y, y) \approx x$ and $p_2(x, x, y) \approx y$ follow directly from the construction. To verify $p_1(x, x, y) \approx p_2(x, y, y)$ we can assume that $x \neq y$. If 01 or 10 are in $\{x, y\}$, then p_1 and p_2 agree (the result is 01 if $01 \in \{x, y\}$ and 10 else). If not, then $p_1(x, x, y) = p_2(x, y, y) = x$.

Finally, we prove that p_1 is a polymorphism of $\mathbb{Z}$; a similar argument works for p_2. If we have triples $\mathbf{a}, \mathbf{b} \in Z^3$ such that $a_i \to b_i$, for $i = 1, 2, 3$, then $\{a_1, a_2, a_3\} \subseteq \{00, 10\}$ and $\{b_1, b_2, b_3\} \subseteq \{01, 11\}$. Thus also $p_1(\mathbf{a}) \in \{00, 10\}$ and $p_1(\mathbf{b}) \in \{01, 11\}$. If $p_1(\mathbf{a}) = 10$, then $p_1(\mathbf{a}) \to p_1(\mathbf{b})$ follows immediately. If $p_1(\mathbf{a}) = 00$, then $\mathbf{a} = (00, 10, 10)$ or $\mathbf{a} = (00, 00, 00)$. In both cases $b_1 = 01$ which gives $p_1(\mathbf{b}) = 01$ and $p_1(\mathbf{a}) \to p_1(\mathbf{b})$. □

5.3 The proof

In this subsection we prove Theorem 5.1 and Corollary 5.2. Fix a finite relational structure; without loss of generality we can assume that $\mathbb{A} = (A; R)$, where R is a k-ary relation (see Example 2.6).

First we need to gather a few facts about connected components of powers of $\mathcal{D}(\mathbb{A})$. This is because when constructing an m-ary polymorphism, one can define it independently on different connected components of $\mathcal{D}(\mathbb{A})^m$ without violating the polymorphism condition.

We start with the diagonal component: since $\mathcal{D}(\mathbb{A})$ is connected, it follows that for every $m > 0$ the *diagonal* (i.e., the set $\{(c, c, \dots, c) \mid c \in \mathcal{D}(\mathbb{A})\}$) is connected in $\mathcal{D}(\mathbb{A})^m$. We denote by Δ_m the connected component of $\mathcal{D}(\mathbb{A})^m$ containing the diagonal.

Lemma 5.4. *For every $m > 0$, both $A^m \subseteq \Delta_m$ and $R^m \subseteq \Delta_m$.*

Proof. Fix an arbitrary element $a \in A$. Let $(\mathbf{r}^1, \dots, \mathbf{r}^m) \in R^m$ and for every $i \in [m]$ let $\varphi_i : \mathbb{Q} \to \mathbb{P}_{(a,\mathbf{r}^i)}$. The homomorphism defined by $x \mapsto (\varphi_1(x), \dots, \varphi_m(x))$ witnesses $(a, \dots, a) \xrightarrow{\mathbb{Q}} (\mathbf{r}^1, \dots, \mathbf{r}^m)$ in $\mathcal{D}(\mathbb{A})^m$. This proves that $R^m \subseteq \Delta_m$; a similar argument gives $A^m \subseteq \Delta_m$. $\square$

The next lemma shows that there is only one non-trivial connected component of $\mathcal{D}(\mathbb{A})^m$ that contains tuples (whose entries are) on the same level in $\mathcal{D}(\mathbb{A})$; namely Δ_m. All other such components are singleton.

Lemma 5.5. *Let $m > 0$ and let Γ be a connected component of $\mathcal{D}(\mathbb{A})^m$ containing an element $\mathbf{c}$ such that $\mathrm{lvl}(c_1) = \cdots = \mathrm{lvl}(c_m)$. Then every element $\mathbf{d} \in \Gamma$ is of the form $\mathrm{lvl}(d_1) = \cdots = \mathrm{lvl}(d_m)$ and the following hold.*

(i) If $\mathbf{c} \to \mathbf{d}$ is an edge in Γ such that $\mathbf{c} \notin A^m$ and $\mathbf{d} \notin R^m$, then there exist $e_1, \dots, e_m \in A \times R$ and $l \in [k]$ such that $\mathbf{c}, \mathbf{d} \in \prod_{i=1}^m \mathbb{P}_{e_i,l}$.

(ii) Either $\Gamma = \Delta_m$ or Γ is one-element.

Proof. First observe that if an element $\mathbf{d}$ is connected in $\mathcal{D}(\mathbb{A})^m$ to an element $\mathbf{c}$ with $\mathrm{lvl}(c_1) = \cdots = \mathrm{lvl}(c_m)$, then there is an oriented path $\mathbb{Q}'$ such that $\mathbf{c} \xrightarrow{\mathbb{Q}'} \mathbf{d}$ from which it follows that $\mathrm{lvl}(d_1) = \cdots = \mathrm{lvl}(d_m)$. To prove (i), let $\mathbf{c} \to \mathbf{d}$ be an edge in Γ such that $\mathbf{c} \notin A^m$ and $\mathbf{d} \notin R^m$. For $i = 1, \dots, m$ let e_i be such that $c_i \in \mathbb{P}_{e_i}$ and let $l = \mathrm{lvl}(c_1)$. The claim now follows immediately from the construction of $\mathcal{D}(\mathbb{A})$.

It remains to prove (ii). If $|\Gamma| > 1$, then there is an edge $\mathbf{c} \to \mathbf{d}$ in Γ. If $\mathbf{c} \in A^m$ or $\mathbf{d} \in R^m$, then the claim follows from Lemma 5.4. Otherwise, from (i), there exists $l \in [k]$ and $e_i = (a_i, \mathbf{r}^i)$ such that $\mathbf{c}, \mathbf{d} \in \prod_{i=1}^m \mathbb{P}_{e_i,l}$. For every $i \in [m]$ we can walk from c_i to $\iota\mathbb{P}_{e_i,l}$ following the path $\bullet \to \bullet \leftarrow \bullet$; and so $\mathbf{c}$ and $(\iota\mathbb{P}_{e_1,l}, \dots, \iota\mathbb{P}_{e_m,l})$ are connected. For every $i \in [m]$ there exists a homomorphism $\varphi_i : \mathbb{Q} \to \mathbb{P}_{e_i}$ such that $\varphi_i(\iota\mathbb{Q}) = a_i$ and $\varphi_i(\iota\mathbb{Q}_l) = \iota\mathbb{P}_{e_i,l}$. The homomorphism $\mathbb{Q} \to \mathcal{D}(\mathbb{A})^m$ defined by $x \mapsto (\varphi_1(x), \dots, \varphi_m(x))$ shows that $(a_1, \dots, a_m)$ and $(\iota\mathbb{P}_{e_1,l}, \dots, \iota\mathbb{P}_{e_m,l})$ are connected. By transitivity, $(a_1, \dots, a_m)$ is connected to $\mathbf{c}$ and therefore $(a_1, \dots, a_m) \in \Gamma$. Using (i) we obtain $\Gamma = \Delta_m$. $\square$

In order to deal with connected components that contain tuples of varying levels, we need to define two linear orders $\sqsubseteq, \sqsubseteq^\star$ on $\mathcal{D}(\mathbb{A})$. These linear orders will then be used to choose elements from input tuples of the polymorphisms under construction in a "uniform" way.

Fix an arbitrary linear order $\preceq$ on A. It induces lexicographic orders on relations on A. We will use $\preceq_{\mathrm{LEX}}$ on R, $A \times R$ and also on $R \times A$. (Note the difference!) We define the linear order $\sqsubseteq$ on $\mathcal{D}(\mathbb{A})$ by putting $x \sqsubset y$ if either of the following five conditions holds:

(1) $x, y \in A$ and $x \prec y$, or

(2) $x, y \in R$ and $x \prec_{\mathrm{LEX}} y$, or

(3) $\mathrm{lvl}(x) < \mathrm{lvl}(y)$,

or $\mathrm{lvl}(x) = \mathrm{lvl}(y)$, $x, y \notin A \cup R$, say $x \in \mathbb{P}_{(a,\mathbf{r})}$, $y \in \mathbb{P}_{(b,\mathbf{s})}$, and

(4) $(a, \mathbf{r}) = (b, \mathbf{s})$ and x is closer to $\iota\mathbb{P}_{(a,\mathbf{r})}$ than y, or

(5) $(a, \mathbf{r}) \prec_{\mathrm{LEX}} (b, \mathbf{s})$.

We also define the linear order $\sqsubseteq^\star$, which will serve as a "dual" to $\sqsubseteq$ in some sense. The definition is almost identical, we put $x \sqsubset^\star y$ if one of (1), (2), (3), (4) or ($5^\star$) holds, where

($5^\star$) $(\mathbf{r}, a) \prec_{\mathrm{LEX}} (\mathbf{s}, b)$.

The last ingredient is the following lemma; $\sqsubseteq$ and $\sqsubseteq^\star$ were taylored to satisfy it.

Lemma 5.6. *Let C and D be subsets of $\mathcal{D}(\mathbb{A})$ such that*

- *for every $x \in C$ there exists $y' \in D$ such that $x \to y'$, and*
- *for every $y \in D$ there exists $x' \in C$ such that $x' \to y$.*

Then the following is true.

(i) If $D \not\subseteq R$ and c and d are the $\sqsubseteq$-minimal elements of C and D, respectively, then $c \to d$.

(ii) If $C \not\subseteq A$ and c and d are the $\sqsubseteq^\star$-maximal elements of C and D, respectively, then $c \to d$.

Proof. We will prove item (ii); the proof of (i) is similar. Let c', d' be such that $c \to d'$ and $c' \to d$. There exist $(a, \mathbf{r}), (b, \mathbf{s}) \in A \times R$ such that $c, d' \in \mathbb{P}_{(a,\mathbf{r})}$ and $c', d \in \mathbb{P}_{(b,\mathbf{s})}$. Suppose for contradiction that $c \not\to d$. In particular, $c \neq c'$ and $d \neq d'$. Note that the assumptions of c, c', d, d' and item (3) of the definition of $\sqsubseteq^\star$ give $\mathrm{lvl}(c') + 1 = \mathrm{lvl}(d) \geq \mathrm{lvl}(d') = \mathrm{lvl}(c) + 1 \geq \mathrm{lvl}(c') + 1$, so that $\mathrm{lvl}(c) = \mathrm{lvl}(c')$ and $\mathrm{lvl}(d) = \mathrm{lvl}(d')$. So, the reason for $d' \sqsubset^\star d$ must be one of items (2), (4) or ($5^\star$).

If it is (2), then $d' = \mathbf{r}$ and $d = \mathbf{s}$ with $\mathbf{r} \prec_{LEX} \mathbf{s}$. Therefore $(\mathbf{r}, a) \prec_{\mathrm{LEX}} (\mathbf{s}, b)$ and ($5^\star$) gives us $c \sqsubset^\star c'$, a contradiction with the maximality of c. If it is (4), then $(a, \mathbf{r}) = (b, \mathbf{s})$ and $c \to d' \leftarrow c' \to d$ form a zigzag. By (4) we again get $c \sqsubset^\star c'$. In case the reason for $d' \sqsubset^\star d$ is ($5^\star$), the same item gives $c \sqsubset^\star c'$. (Here we need the assumption that $C \not\subseteq A$, otherwise we could have $c = a$, $c' = b$, $b \prec a$ and $c' \sqsubset^\star c$ by (1) even though $(\mathbf{r}, a) \prec_{\mathrm{LEX}} (\mathbf{s}, b)$.) □

Proof of Theorem 5.1

Let Σ be a set of identities in operation symbols $\{f_\lambda : \lambda \in \Lambda\}$ satisfying the assumptions. Let $\{f_\lambda^{\mathbb{A}} \mid \lambda \in \Lambda\}$ and $\{f_\lambda^{\mathbb{Z}} \mid \lambda \in \Lambda\}$ be interpretations of the operation symbols witnessing $\mathbb{A} \models \Sigma$ and $\mathbb{Z} \models \Sigma$, respectively.

We will now define polymorphisms $\{f_\lambda^{\mathcal{D}(\mathbb{A})} \mid \lambda \in \Lambda\}$ witnessing that $\mathcal{D}(\mathbb{A}) \models \Sigma$. Fix $\lambda \in \Lambda$ and assume that f_λ is m-ary. We split the definition of $f_\lambda^{\mathcal{D}(\mathbb{A})}$ into several cases and subcases. Let $\mathbf{c} \in \mathcal{D}(\mathbb{A})^m$ be an input tuple.

Case 1. $\mathbf{c} \in A^m \cup R^m$.

<u>1a</u> If $\mathbf{c} \in A^m$, we define $f_\lambda^{\mathcal{D}(\mathbb{A})}(\mathbf{c}) = f_\lambda^{\mathbb{A}}(\mathbf{c})$.

<u>1b</u> If $\mathbf{c} \in R^m$, we define $f_\lambda^{\mathcal{D}(\mathbb{A})}(\mathbf{c}) = (f_\lambda^{\mathbb{A}})^{(k)}(\mathbf{c})$.

Case 2. $\mathbf{c} \in \Delta_m \setminus (A^m \cup R^m)$.

Let $c_i \in \mathbb{P}_{e_i}$ and define $e = (f_\lambda^{\mathbb{A}})^{(k+1)}(e_1, \ldots, e_m)$. Let $l \in [k]$ be minimal such that $c_i \in \mathbb{P}_{e_i,l}$ for all $i \in [m]$. (Its existence is guaranteed by Lemma 5.5 (i).)

2a If $\mathbb{P}_{e,l}$ is a single edge, then we define $f_\lambda^{\mathcal{D}(\mathbb{A})}(\mathbf{c})$ to be the vertex from $\mathbb{P}_{e,l}$ having the same level as all the c_i's.

If $\mathbb{P}_{e,l}$ is a zigzag, then at least one of the $\mathbb{P}_{e_i,l}$'s is a zigzag as well. (This follows from the construction of $\mathcal{D}(\mathbb{A})$ and the fact that $f_\lambda^{\mathbb{A}}$ preserves R.) For every $i \in [m]$ such that $\mathbb{P}_{e_i,l}$ is a zigzag let $\Phi_i : \mathbb{P}_{e_i,l} \to \mathbb{Z}$ be the (unique) isomorphism. Let Φ denote the isomorphism from $\mathbb{P}_{e,l}$ to $\mathbb{Z}$.

2b If all of the $\mathbb{P}_{e_i,l}$'s are zigzags, then the value of $f_\lambda^{\mathcal{D}(\mathbb{A})}$ is defined as follows:

$$f_\lambda^{\mathcal{D}(\mathbb{A})}(\mathbf{c}) = \Phi^{-1}(f_\lambda^{\mathbb{Z}}(\Phi_1(c_1), \ldots, \Phi_m(c_m))).$$

2c Otherwise, we define $f_\lambda^{\mathcal{D}(\mathbb{A})}(\mathbf{c})$ to be the $\sqsubseteq$-minimal element from the set

$$\{\Phi^{-1}(\Phi_i(c_i)) \mid \mathbb{P}_{e_i,l} \text{ is a zigzag}\}.$$

(Equivalently, $f_\lambda^{\mathcal{D}(\mathbb{A})}(\mathbf{c}) = \Phi^{-1}(z)$, where z is the $\leq_{\mathbb{Z}}$-minimal element from the set $\{(\Phi_i(c_i) \mid \mathbb{P}_{e_i,l} \text{ is a zigzag}\}$.)

Case 3. $\mathbf{c} \notin \Delta_m$.

3a If $|\{\mathrm{lvl}(c_i) \mid i \in [m]\}| = 1$ and the c_i's lie on precisely two paths (say, $\{c_1, \ldots, c_m\} \subseteq \mathbb{P}_e \cup \mathbb{P}_{e'}$ with $e \prec_{LEX} e'$, the lexicographic order of $A \times R$), then we define the mapping $\Psi : \{c_1, \ldots, c_m\} \to \{00, 10\}$ as follows:

$$\Psi(c_i) = \begin{cases} 00 & \text{if } c_i \in \mathbb{P}_e, \\ 10 & \text{if } c_i \in \mathbb{P}_{e'}. \end{cases}$$

We define $f_\lambda^{\mathcal{D}(\mathbb{A})}(\mathbf{c})$ to be the $\sqsubseteq$-minimal element from the set

$$\{c_i : \Psi(c_i) = f_\lambda^{\mathbb{Z}}(\Psi(c_1), \ldots, \Psi(c_m))\}.$$

3b If $|\{\mathrm{lvl}(c_i) \mid i \in [m]\}| = 2$ (say, $\mathrm{lvl}(c_i) \in \{l, l'\}$ for all $i \in [m]$ and $l < l'$), then we define the mapping $\Theta : \{c_1, \ldots, c_m\} \to \{00, 10\}$ as follows:

$$\Theta(c_i) = \begin{cases} 00 & \text{if } \mathrm{lvl}(c_i) = l \\ 10 & \text{if } \mathrm{lvl}(c_i) = l'. \end{cases}$$

We set $z = f_\lambda^{\mathbb{Z}}(\Theta(c_1), \ldots, \Theta(c_m))$ and $C' = \{c_i : \Theta(c_i) = z\}$ and define

$$f_\lambda^{\mathcal{D}(\mathbb{A})}(\mathbf{c}) = \begin{cases} \text{the } \sqsubseteq\text{-minimal element from } C' & \text{if } z = 00 \\ \text{the } \sqsubseteq^\star\text{-maximal element from } C' & \text{if } z = 10. \end{cases}$$

3c In all other cases we define $f_\lambda^{\mathcal{D}(\mathbb{A})}(\mathbf{c})$ to be the $\sqsubseteq$-minimal element from the set $\{c_1, \ldots, c_m\}$.

While the construction is a bit technical, the ideas behind it are not so complicated. Case 1 gives us no choice. In Case 2 we use $f^{\mathbb{A}}$ to determine on which path $\mathbb{P}_e$ should the result lie, and we are left with a choice of at most two possible elements (when $\mathbb{P}_{e,l}$ is a zigzag). In Case 3 we cannot use $f^{\mathbb{A}}$ anymore. Instead, we choose the result as a minimal element from (a subset of) the input elements under a suitable linear order $\sqsubseteq$. This choice typically does not depend on order or repetition of the input elements, which allows us to satisfy balanced identities "for free". The trickiest part is to deal with connected components which can contain tuples with just two distinct elements, as these can play a role in some non-balanced identity (in two variables) which we need to satisfy. We need to employ $f^{\mathbb{Z}}$ to choose from two possibilities: a result which is the right element (in subcase 2c), from the right path (in 3a) or from the right level (in 3b). We then use $\sqsubseteq$ to choose the result from the "good" elements (and as a technical nuisance, to maintain the polymorphism property, in 3b we sometimes need to use $\sqsubseteq^\star$-maximal elements instead).

We need to verify that the operations we constructed are polymorphisms and that they satisfy all identities from Σ. We divide the proof into three claims.

Claim 5.7. *For every $\lambda \in \Lambda$, $f_\lambda^{\mathcal{D}(\mathbb{A})}$ is a polymorphism of $\mathcal{D}(\mathbb{A})$.*

Proof. Let $\mathbf{c} \to \mathbf{d}$ be an edge in $\mathcal{D}(\mathbb{A})^m$. Note that $\mathbf{c} \in \Delta_m$ if and only if $\mathbf{d} \in \Delta_m$. The tuple $\mathbf{c}$ cannot fall under subcase 1b or under 3a, because these cases both prevent an outgoing edge from $\mathbf{c}$ (see Lemma 5.5 (ii) for why this is true for 3a).

We first consider the situation where $\mathbf{c}$ falls under subcase 1a of the definition. Then $\mathbf{d}$ falls under case 2 and, moreover, $d_i = \iota\mathbb{P}_{e_i,1}$ for all $i \in [m]$. It is not hard to verify that $f_\lambda^{\mathcal{D}(\mathbb{A})}(\mathbf{d}) = \iota\mathbb{P}_{e,1}$. (In subcase 2b we need the fact that $f_\lambda^{\mathbb{Z}}$ is idempotent.) Therefore $f_\lambda^{\mathcal{D}(\mathbb{A})}(\mathbf{c}) = \iota\mathbb{P}_e \to \iota\mathbb{P}_{e,1} = f_\lambda^{\mathcal{D}(\mathbb{A})}(\mathbf{d})$ and the polymorphism condition holds. The argument is similar when $\mathbf{d}$ falls under subcase 1b (and so $\mathbf{c}$ under case 2).

Consider now that $\mathbf{c}$ falls under case 2. Then $\mathbf{d}$ falls either under subcase 1b, which was handled in the above paragraph, or also under case 2. The elements $e_1, \ldots, e_m$ and e are the same for both $\mathbf{c}$ and $\mathbf{d}$. By Lemma 5.5 (i), there exists $l \in [k]$ such that $c_i, d_i \in \mathbb{P}_{e_i,l}$ for all $i \in [m]$.

If the value of l is also the same for both $\mathbf{c}$ and $\mathbf{d}$, then $f_\lambda^{\mathcal{D}(\mathbb{A})}(\mathbf{c}) \to f_\lambda^{\mathcal{D}(\mathbb{A})}(\mathbf{d})$ follows easily; in subcase 2a trivially, in 2b from the fact that $f_\lambda^{\mathbb{Z}}$ is a polymorphism of $\mathbb{Z}$ and in 2c from Lemma 5.6.

It may be the case that this l is not minimal for the tuple $\mathbf{c}$, that is, that $c_i \in \mathbb{P}_{e_i,l-1} \cap \mathbb{P}_{e_i,l}$ for all $i \in [m]$. But then $c_i = \tau\mathbb{P}_{e_i,l-1} = \iota\mathbb{P}_{e_i,l}$ and thus $f_\lambda^{\mathcal{D}(\mathbb{A})}(\mathbf{c}) = \iota\mathbb{P}_{e_i,l}$ (again, using idempotency of $f_\lambda^{\mathbb{Z}}$ in subcase 2b). Knowing this allows for the same argument as in the above paragraph.

If $\mathbf{c}$ falls under one of the subcases 3b or 3c, then $\mathbf{d}$ falls under the same subcase. In subcase 3c we apply 5.6 (i) with $\{c_1, \ldots, c_m\}$ and $\{d_1, \ldots, d_m\}$ in the roles of C and D, respectively. In subcase 3b our construction "chooses" either the lower or the higher level, and it is easy to see that this choice (i.e., the element z) is the same for both $\mathbf{c}$ and $\mathbf{d}$. We then apply Lemma 5.6 (i) or (ii) (depending on z, note that the assumptions are satisfied) with $C' = \{c_i : \Theta(c_i) = z\}$ and $D' = \{d_i : \Theta(d_i) = z\}$ in the role of C and D, respectively. In both cases we get $f_\lambda^{\mathcal{D}(\mathbb{A})}(\mathbf{c}) \to f_\lambda^{\mathcal{D}(\mathbb{A})}(\mathbf{d})$. $\square$

Claim 5.8. *The $f_\lambda^{\mathcal{D}(\mathbb{A})}$'s satisfy every balanced identity from Σ.*

Proof. Let $f_\lambda(\mathbf{u}) \approx f_\mu(\mathbf{v}) \in \Sigma$ be a balanced identity in s distinct variables $\{x_1, \ldots, x_s\}$. Let $\mathcal{E} : \{x_1, \ldots, x_s\} \to \mathcal{D}(\mathbb{A})$ be some evaluation of the variables. Let $\mathbf{u}^{\mathcal{E}}$ and $\mathbf{v}^{\mathcal{E}}$ denote the corresponding evaluation of these tuples.

Note that both $f_\lambda^{\mathcal{D}(\mathbb{A})}(\mathbf{u}^\mathcal{E})$ and $f_\mu^{\mathcal{D}(\mathbb{A})}(\mathbf{v}^\mathcal{E})$ fall under the same subcase of the definition. The subcase to be applied depends only on the set of elements occuring in the input tuple, except for case two, where the choice of e matters as well. However, since the identity $f_\lambda(\mathbf{u}) \approx f_\mu(\mathbf{v})$ holds in $\mathbb{A}$, this e is the same for both $\mathbf{u}^\mathcal{E}$ and $\mathbf{v}^\mathcal{E}$. Therefore, to verify that $f_\lambda^{\mathcal{D}(\mathbb{A})}(\mathbf{u}^\mathcal{E}) = f_\mu^{\mathcal{D}(\mathbb{A})}(\mathbf{v}^\mathcal{E})$, it is enough to consider the individual subcases separately.

In case 1 it follows immediately from the fact that the identity holds in $\mathbb{A}$. In case 2 it is easily seen that both $f_\lambda^{\mathcal{D}(\mathbb{A})}(\mathbf{u}^\mathcal{E})$ and $f_\mu^{\mathcal{D}(\mathbb{A})}(\mathbf{v}^\mathcal{E})$ have the same level, and since the identity holds in $\mathbb{A}$, they also lie on the same path $\mathbb{P}_{e,l}$. To see that these two elements are equal, note that in subcase 2a it is trivial, in 2b it follows directly from the fact that the identity holds in $\mathbb{Z}$, and in 2c we use the fact that the identity is balanced: they are both the $\sqsubseteq$-minimal element of the same set.

Similar arguments can be used in case 3. In 3a we choose one of the paths $\mathbb{P}_e$, $\mathbb{P}_{e'}$; the choice is the same because $f_\lambda(\mathbf{u}) \approx f_\mu(\mathbf{v})$ holds in $\mathbb{Z}$. Both $f_\lambda^{\mathcal{D}(\mathbb{A})}(\mathbf{u}^\mathcal{E})$ and $f_\mu^{\mathcal{D}(\mathbb{A})}(\mathbf{v}^\mathcal{E})$ then evaluate to the same element, namely the $\sqsubseteq$-minimal element from $\{\mathcal{E}(x_1), \dots, \mathcal{E}(x_s)\}$ intersected with the chosen path. In 3b the chosen level is the same for both of them (since the identity holds in $\mathbb{Z}$) and they are both the $\sqsubseteq$-minimal, or $\sqsubseteq^\star$-maximal, element of the set of elements from $\{\mathcal{E}(x_1), \dots, \mathcal{E}(x_s)\}$ lying on that level. In 3c both are the $\sqsubseteq$-minimal element of the same set $\{\mathcal{E}(x_1), \dots, \mathcal{E}(x_s)\}$. $\square$

Claim 5.9. *The $f_\lambda^{\mathcal{D}(\mathbb{A})}$'s satisfy every identity from Σ in at most two variables.*

Proof. Balanced identities fall under the scope of the previous claim. Since Σ is idempotent, we may without loss of generality consider only identities of the form $f_\lambda(\mathbf{u}) \approx x$, where $\mathbf{u} \in \{x, y\}^m$. Suppose that x and y evaluate to c and d in $\mathcal{D}(\mathbb{A})$, respectively, and let $\mathbf{c} \in \{c, d\}^n$ be the corresponding evaluation of $\mathbf{u}$. We want to prove that $f_\lambda^{\mathcal{D}(\mathbb{A})}(\mathbf{c}) = c$.

The tuple $\mathbf{c}$ cannot fall into subcase 3c of the definition of $f_\lambda^{\mathcal{D}(\mathbb{A})}$. If it falls into case 1, the equality follows from the fact that the identity holds in $\mathbb{A}$, while in subcases 3a and 3b we use the fact that it holds in $\mathbb{Z}$. (The linear orders $\sqsubseteq, \sqsubseteq^\star$ do not matter, since we only choose elements from singleton sets.)

In case 2 it is easily seen that $f_\lambda^{\mathcal{D}(\mathbb{A})}(\mathbf{c})$ lies on the same path $\mathbb{P}_{e,l}$ as c (using that the identity holds in $\mathbb{A}$) as well as on the same level of this path. In 2a it is trivial that $f_\lambda^{\mathcal{D}(\mathbb{A})}(\mathbf{c}) = c$ while in 2b it follows from the fact that the identity holds in $\mathbb{Z}$. If $\mathbf{c}$ falls under subcase 2c, then $c \in \mathbb{P}_{e,l}$, which is a zigzag, and $d \in \mathbb{P}_{e',l}$, which must be a single edge. Therefore $f_\lambda^{\mathcal{D}(\mathbb{A})}(\mathbf{c})$ is defined to be the $\sqsubseteq$-minimal element from the singleton set $\{c\}$. $\square$

Proof of Corollary 5.2

All items are expressible by linear idempotent sets of identities. In all items except (7) they are in at most two variables, in item (7) the defining identities are balanced. It remains to check that all these conditions are satisfied in the zigzag, which follows from Lemma 5.3 (iii) for item (3), Lemma 5.3 (ii) for item (7) and Lemma 5.3 (i) for all other items.

6 The logspace reduction

In this section, we give the proof of Lemma 3.4, by showing that CSP($\mathcal{D}(\mathbb{A})$) reduces in logspace to CSP($\mathbb{A}$). A sketch of a *polymomial time* reduction is given in the proof of [18, Theorem 13]; technically, that argument is for the special case where $\mathbb{A}$ is itself already a digraph, but the arguments can be broadened to cover our case. To perform this process in logspace is rather technical, with many of the difficulties lying in details that are omitted in the polymomial time description in the proof of [18, Theorem 13]. We wish to thank Barnaby Martin for encouraging us to pursue Lemma 3.4.

The following theorem is an immediate consequence of Lemmata 3.3 and 3.4. As this improves the oft-mentioned polynomial time equivalence of general CSPs with digraph CSPs, we now present it as stand-alone statement.

Theorem 6.1. *Every fixed finite template CSP is logspace equivalent to the CSP over some finite digraph.*

6.1 Outline of the algorithm

We first assume that CSP($\mathbb{A}$) is itself not trivial (that is, that there is at least one no instance and one yes instance): this uninteresting restriction is necessary because CSP($\mathcal{D}(\mathbb{A})$) will have no instances always. Now let $\mathbb{G}$ be an instance of CSP($\mathcal{D}(\mathbb{A})$). Also, let n denote the height of $\mathcal{D}(\mathbb{A})$ and k the arity of the single fundamental relation R of $\mathbb{A}$: so, $n = k + 2$. Recall that the vertices of $\mathcal{D}(\mathbb{A})$ include those of A as well as the elements in R. The rough outline of the algorithm is as follows.

(Stage 1.) Some initial analysis of $\mathbb{G}$ is performed to decide if it is broadly of the right kind of digraph to be a possible yes instance. If not, some fixed no instance of CSP($\mathbb{A}$) is output.

(Stage 2.) It is convenient to remove any components of $\mathbb{G}$ that are too small. These are considered directly, and in logspace we determine whether or not they are YES or NO instances of CSP($\mathcal{D}(\mathbb{A})$). If all are YES we ignore them. If one returns NO we reject the entire instance and return some fixed NO instance of CSP($\mathbb{A}$).

(Stage 3.) Now it may be assumed that $\mathbb{G}$ is roughly similar to a digraph of the form $\mathcal{D}(\mathbb{B})$ (for some structure $\mathbb{B}$), but where some vertices at level 0 have been lost, and other vertices at this level and at level n have been split into numerous copies, with each possibly containing different parts of the information in the connecting edges of $\mathcal{D}(\mathbb{B})$. Essentially, the required $\mathbb{B}$ is a kind of quotient of an object definable from $\mathbb{G}$, though some extra vertices must be added (this is similar to the addition of vertices to account for existentially quantified variables in a primitive positive definition of a relation: only new vertices are added, and they are essentially unconstrained beyond the specific purpose for which they are added). To construct $\mathbb{B}$ in logspace, we work in two steps: we describe a logspace construction of some intermediate information. Then we describe a logspace reduction from strings of suitable information of this kind to $\mathbb{B}$. The overall process is logspace because a composition of two logspace processes is logspace.

(Stage 3A.) From $\mathbb{G}$ we output a list of "generalized hyperedges". These are k-tuples consisting of sets of vertices of $\mathbb{G}$ plus some newly added vertices. Moreover, they sometimes include a labelling to record how they were created.

(Step 3B.) The actual structure $\mathbb{B}$ is constructed from the generalized hyperedges in the previous step. The input consists of generalized hyperedges. To create $\mathbb{B}$, numerous undirected graph reachability checks are performed. The final "vertices" of $\mathbb{B}$ are in fact sets of vertices of $\mathbb{G}$, so that the generalized hyperedges become actual hyperedges in the conventional sense (k-tuples of "vertices", now consisting of sets of vertices of $\mathbb{G}$). This may be reduced to an adjacency matrix description as a separate logspace process, but that is routine.

Stage 1 is described in Subsection 6.3, while Stage 2 is described in Subsection 6.4. The most involved part of the algorithm is stage 3A. In Subsection 6.5 we give some preliminary discussion on how the process is to proceed: an elaboration on the item listed in the present subsection. In particular a number of definitions are introduced to aid the description of Stage 3A. The actual algorithm is detailed in Subsection 6.6. Step 3B is described in Subsection 6.7. After a brief discussion of why the algorithm is a valid reduction from CSP($\mathcal{D}(\mathbb{A})$) to CSP($\mathbb{A}$), we present an example of Stages 3A and 3B in action. This example may be a useful reference while reading Subsection 6.6 and 6.7.

Before we begin describing the algorithm we recall some basic logspace process that we will use frequently.

6.2 Subroutines

The algorithm we describe makes numerous calls on other logspace computable processes. Our algorithm may be thought of as running on an oracle machine, with several query tapes. Each query tape verifies membership in some logspace solvable problem. It is well known that $\mathtt{L}^{\mathtt{L}} = \mathtt{L}$, and this enables all of the query tapes to be eliminated within logspace. For the sake of clarity, we briefly recall some basic information on logspace on an oracle machine. An oracle program with logspace query language U has access to an input tape, a working tape (or tapes) and an output tape. Unlimited reading may be done from the input, but no writing. Unlimited writing may be done to the output tape, but no reading. Unlimited writing may be done to the query tape, but no reading. Once the query state is reached however, the current word written to the query tape is tested for membership in the language U (at the cost of one step of computation), and a (correct) answer of either yes or no is received by the program, and the query tape is immediately erased. The space used is measured *only from the working tape*, where both reading and writing is allowed. If such a program runs in logspace, then it can be emulated by an actual logspace program (with no oracle), so that $\mathtt{L}^{\mathtt{L}} = \mathtt{L}$. The argument is essentially the fact that a composition of logspace reductions is a logspace reduction: each query to the oracle (of a string w for instance) during the computation is treated as a fresh instance of a reduction to the membership problem of U, which is then composed with the logspace algorithm for U (which is, as usual, done without ever writing any more than around one symbol of w at a time—plus a short counter—which is why space used on the oracle tape does not matter in the oracle formulation of logspace, and why we may assume that the query tape may be erased after completion of the query).

In this subsection we describe the basic checks that are employed during our algorithm.

Undirected reachability. Given an undirected graph and two vertices u, v, there is a logspace algorithm to determine if u is reachable from v (Reingold [36, 37]). In the case of a directed graph we may use this to determine if two vertices are connected by some oriented path (simply treat the digraph as an undirected graph, and use undirected reachability). This means, for example, that we can construct, in logspace, the *smallest equivalence relation containing some input binary relation.*

A second process we frequently perform is reachability checks involving edges that are not precisely those of the current input digraph. A typical instance might be where we have some fixed vertex u in consideration, and we wish to test if some vertex v can be reached from u by an oriented path consisting only of vertices satisfying some property $\mathcal{Q}$, where $\mathcal{Q}$ is a logspace testable property. This is undirected graph reachability, except that as well as ignoring the edge direction, we must also ignore any vertex failing property $\mathcal{Q}$. This can be performed in logspace on an oracle machine running an algorithm for undirected graph reachability and whose query tape tests property $\mathcal{Q}$.

Component checking. Undirected graph reachability is also fundamental to checking properties of induced subgraphs. In a typical situation we have some induced subgraph C of $\mathbb{G}$ (containing some vertex u, say) and we want to test if it satisfies some property $\mathcal{P}$. Membership of vertices in C is itself determined by some property $\mathcal{Q}$, testable in logspace. It is convenient to assume that the query tape for $\mathcal{P}$ expects inputs that consist of a list of directed edges (if adjacency matrix is preferred, then this involves one further nested logspace process, but the argument is routine). We may construct a list of the directed edges in the component C on a logspace machine with a query tape for $\mathcal{P}$, for undirected graph reachability and for $\mathcal{Q}$. We write C to the query tape for $\mathcal{P}$ as follows. Systematically enumerate pairs of vertices v_1, v_2 of $\mathbb{G}$ (re-using some fixed portion of work tape for each pair), in each case testing for undirected reachability of both v_1 and v_2 from u, and also for satisfaction of property $\mathcal{Q}$. If both are reachable, and if (v_1, v_2) is an edge of $\mathbb{G}$ then we output the edge (v_1, v_2) to the query tape for $\mathcal{P}$. After the last pair has been considered, we may finally query $\mathcal{P}$.

Testing for interpretability in paths. By an *interpretation* of a digraph C in another digraph $\mathbb{Q}$ we mean simply a graph homomorphism from C to $\mathbb{Q}$. The basic properties we wish to test of components usually concern interpretability within some fixed finite family of directed paths. We consider the paths $\mathbb{Q}_S$, where S is some subset of $[k] = \{1, \dots, k\}$: recall (Section 3) that these have zigzags in a position i when $i \notin S$ (so that a small S corresponds to a large number of zigzags, while $\mathbb{Q}_{[k]}$ itself is simply the directed path on $k + 3$ vertices, with no zigzags).

It is not hard to see that a balanced digraph of height $n = k + 2$ admits a homomorphism into $\mathbb{Q}_S$ if and only if it admits a homomorphism into each of $\mathbb{Q}_{[k]\setminus\{i\}}$ for $i \notin S$ (this is discussed further in the proof of the next lemma). For balanced digraphs of smaller height this may fail, as the interpretations in the various $\mathbb{Q}_{[k]\backslash\{i\}}$ need not be at the same levels. To circumvent this, we say that (for $S \subseteq [k]$) a balanced connected digraph $\mathbb{H}$ is interpretable in $\mathbb{Q}_S$ *at level* i, if it is interpretable in $\mathbb{Q}_S$ with a vertex of height 0 in $\mathbb{H}$ taking the value of a height i vertex of $\mathbb{Q}_S$.

Lemma 6.2. *Let $\mathbb{H}$ be a connected balanced digraph. Then $\mathbb{H}$ is interpretable in $\mathbb{Q}_S$ at height i if and only if for each $j \notin S$ we have $\mathbb{H}$ interpretable in $\mathbb{Q}_{[k]\backslash\{j\}}$ at height i. For fixed i and connected balanced digraph $\mathbb{H}$ of height at most $n-i$, there is a unique minimum set $S \subseteq [k]$ with $\mathbb{H}$ interpretable in $\mathbb{Q}_S$ at height i.*

Proof. The second statement follows immediately from the first, as we may successively test for interpretability (at height i) in $\mathbb{Q}_{[k]\backslash\{j\}}$ for $j = 1, \dots, k$. The bound $n-i$ is simply to account for the fact that if $\mathbb{H}$ has height greater than $n-i$, then it is not even interpretable in $\mathbb{Q}_{[k]}$ at height i (yet this does not imply that $S = [k]$). For the first statement, observe that if $j \notin S$, then there is a height-preserving homomorphism from $\mathbb{Q}_S$ onto $\mathbb{Q}_{[k]\backslash\{j\}}$ (as $S \subseteq [k]\backslash\{j\}$). So it suffices to show that if $\mathbb{H}$ is interpretable in $\mathbb{Q}_{[k]\backslash\{j\}}$ at height i for each $j \notin S$ then it is interpretable in $\mathbb{Q}_S$ at height i. This is routine, because the single zigzag in $\mathbb{Q}_{[k]\backslash\{j\}}$ (based at height j) for $j \notin S$ matches the corresponding zigzag based at height j in $\mathbb{Q}_S$. More formally, in the direct product $\prod_{j\notin S} \mathbb{Q}_{[k]\backslash\{j\}}$, the component connecting the tuple of initial vertices to terminal vertices maps homomorphically onto $\mathbb{Q}_S$. $\square$

Definition 6.3. *The smallest set $S \subseteq [k]$ for which a connected balanced digraph $\mathbb{H}$ is interpretable in $\mathbb{Q}_S$ at height i is denoted by $\Gamma(\mathbb{H})^{(i)}$. When i is implicit, then we write simply $\Gamma(\mathbb{H})$.*

Lemma 6.4.

1. $\mathrm{CSP}(\mathbb{Q}_{[k]})$ *is solvable in logspace, even with singleton unary relations added.*
2. *If $\mathbb{H}$ is connected and balanced of height at most n, then for any vertex u and v, the height of v relative to that of u may be computed in logspace.*
3. $\mathrm{CSP}(\mathbb{Q}_{[k]\backslash\{i\}})$ *is solvable in logspace for any $i \in \{1, \dots, k\}$, even when singleton unary relations are added.*
4. *For any $S \subseteq \{1, \dots, k\}$ the problem* $\mathrm{CSP}(\mathbb{Q}_S)$ *is solvable in logspace, even when singleton unary relations are added.*
5. *For a balanced connected digraph $\mathbb{H}$ of height at most n, we may test membership of numbers j in the set $\Gamma(\mathbb{H})^{(i)}$ in logspace.*
6. *If $\mathbb{Q}_{S_1}, \mathbb{Q}_{S_2}, \dots, \mathbb{Q}_{S_\ell}$ is a family of connecting paths in $\mathcal{D}(\mathbb{A})$, then the CSP over the digraph formed by amalgamating the $\mathbb{Q}_{S_i}$ at either all the initial points, or at all the terminal points is logspace solvable.*

Proof. (1) Note that $\mathbb{Q}_{[k]}$ has both a Maltsev polymorphism and a majority, hence is solvable in logspace even when unary singleton relations are added [15].

(2) For each $0 \leq i, j \leq n$ (the possible heights) we may test for interpretability of $\mathbb{H}$ in $\mathbb{Q}_{[k]}$ with u constrained to lie at height i and v constrained to lie at height j. As $\mathbb{H}$ is balanced of height at most n, at least one such instance is interpretable, and the number $j-i$ is the relative height of v above u.

(3) Note that as $\mathbb{Q}_{[k]\backslash\{i\}}$ is a core, we have CSP($\mathbb{Q}_{[k]\backslash\{i\}}$) logspace equivalent to the CSP over $\mathbb{Q}_{[k]\backslash\{i\}}$ with all unary singletons added (see [9]).

Given an input digraph $\mathbb{H}$, we first test if $\mathbb{H}$ is interpretable in $\mathbb{Q}_{[k]}$ (which verifies that $\mathbb{H}$ is balanced, and of sufficiently small height). Reject if NO. Otherwise we may assume that $\mathbb{H}$ is a single component (because it suffices to interpret each component, and component checking has been described as logspace in earlier discussion).

We successively search for an interpretation of $\mathbb{H}$ in $\mathbb{Q}_{[k]\backslash\{i\}}$ at heights $0, 1, \ldots, n$; in each case, item (2) shows that we have access to a suitable notion of height for the vertices of $\mathbb{H}$. The remaining part of this proof concerns an attempt at interpretation at one particular height. Now, if there is to be an interpretation of $\mathbb{H}$ in $\mathbb{Q}_{[k]\backslash\{i\}}$ at the given height, then any vertices of the same height $j \notin \{i, i+1\}$ will be identified. We need only ensure there is no directed path of vertices of heights $i-1, i, i+1, i+2$. So it suffices to enumerate all 4-tuples of vertices $u_1, \ldots, u_4$, check if $u_1 \to u_2 \to u_3 \to u_4$, and if so, check that the height of u_1 is not $i-1$. If it is, then reject. Otherwise accept.

(4) & (5) These follow immediately from Lemma 6.2, and part (3) of the present lemma.

(6) We refer to a digraph formed by amalgamating paths in one of the two described fashions, a *fan*. Consider some instance $\mathbb{H}$. As above, we may assume that $\mathbb{H}$ is connected, balanced and is of sufficiently small height. We may first use item (4) to test if $\mathbb{H}$ is interpretable in one of the individual paths $\mathbb{Q}_{S_1}, \mathbb{Q}_{S_2}, \ldots$. If one of these returns a positive answer, then $\mathbb{H}$ is a YES instance. Otherwise, remove all level 0 vertices of $\mathbb{H}$, and successively test each individual component C of the resulting digraph for interpretability in $\mathbb{Q}_{S_1}, \mathbb{Q}_{S_2}, \ldots$, with an additional condition: the vertices of C which were adjacent to a level 0 vertex in $\mathbb{H}$ must be interpreted at the level 1 vertex of $\mathbb{Q}_{S_i}$ adjacent to the initial vertex. Provided each such C is interpretable in at least one of these paths in the described way, then $\mathbb{H}$ is interpretable in the fan (with the level 0 vertices of $\mathbb{H}$ interpreted at the amalgamated initial vertices). Otherwise, $\mathbb{H}$ is not interpretable in the fan and is a NO instance. $\square$

6.3 Stage 1: Verification that $\mathbb{G}$ is balanced and a test for height

If $\mathbb{G}$ is not balanced of height at most n, then we can output some fixed NO instance. The logspace test for this property is Lemma 6.4 part (1). From this point on, we will assume that $\mathbb{G}$ is balanced and of height at most n.

6.4 Stage 2: Elimination of "short components"

If $\mathbb{G}$ contains some component of height strictly less than n, then we will test directly whether or not this component is a YES or NO instance of CSP($\mathcal{D}(\mathbb{A})$) (this is explained in the next paragraph). If any are NO instances, then so is $\mathbb{G}$ and we can output some fixed NO instance of CSP($\mathbb{A}$). Otherwise (if all are YES instances), we may simply ignore these short components. If $\mathbb{G}$ itself has height less than n, then instead of ignoring all components of $\mathbb{G}$ we can output some fixed YES instance of CSP($\mathbb{A}$), completing the reduction.

The process for testing membership of a short component $\mathbb{H}$ in CSP($\mathcal{D}(\mathbb{A})$) is as follows. In any satisfying interpretation of $\mathbb{H}$ in $\mathcal{D}(\mathbb{A})$, we must either interpret within some single path $\mathbb{Q}_S$ connecting A to R in $\mathcal{D}(\mathbb{A})$, or at some fan of such paths emanating from some vertex in A or some vertex in R. There are a fixed finite number of such subgraphs of $\mathcal{D}(\mathbb{A})$, and we may use Lemma 6.4(6) for each one.

For the remainder of the algorithm we will assume that *all connected components have height n.*

6.5 Stage 3: All components of $\mathbb{G}$ have height n

In this case we will eventually output an actual structure $\mathbb{B}$ with the property that $\mathbb{G}$ is a YES instance of CSP($\mathcal{D}(\mathbb{A})$) if and only if $\mathbb{B}$ is a YES instance of CSP($\mathbb{A}$). In fact we focus on the production of a preliminary construction $\mathbb{B}'$ that is not specifically a relational structure, but holds all the information for constructing $\mathbb{B}$ using undirected graph reachability checks. The output $\mathbb{B}'$ will consist of a list of "generalized hyperedges", that will (in step 3B) eventually become the actual hyperedges of $\mathbb{B}$.

For the remainder of the argument, an *internal component* of $\mathbb{G}$ means a connected component of the induced subgraph of $\mathbb{G}$ obtained by removing all vertices of height 0 and n. Note that we have already described that testing for height can be done in logspace. A *base vertex* for such a component C is a vertex at height 0 that is adjacent to C, and a *top vertex* for C is a level n vertex adjacent to C. Note that an internal component may have none, one, or more than one base vertices, and similarly for top vertices. Every internal component must have at least one of a base vertex or a top vertex however, because we have already considered the case of "short" components in Stage 2.

Let C be an internal component. In a satisfying interpretation of $\mathbb{G}$ in $\mathcal{D}(\mathbb{A})$, the component C must be satisfied within some single connecting path (of the form $\mathbb{Q}_S$ for some $S \subseteq [k]$), with any vertices adjacent to the base of C (or to a top of C) being interpreted adjacent to the initial point of the path (or adjacent to the terminal point of the path, respectively). We may identify the smallest set $S \subseteq [k]$ for which C is interpretable in $\mathbb{Q}_S$: it is $\Gamma(C)$ introduced in Definition 6.3, and by Lemma 6.4(5) we can, in logspace, verify membership of numbers up to k in the set $\Gamma(C)$. (Note that we omit the superscript "$^{(i)}$" in the notation, as there is no ambiguity as to what level C is to be satisfied at: it is either $i = 1$, or dually, measured one down from the top, $i = n - 1$.) These internal components of $\mathcal{D}(\mathbb{A})$ are in essence encoding positions of base level vertices in hyperedges, and Lemma 6.4(5) supplies, in logspace, the positions which are being asserted as "filled" by a given internal component C. If $\mathbb{G}$ itself is the path $\mathbb{Q}_I$ for example, then the single internal component C has $\Gamma(C) = I$.

6.6 Stage 3A: Constructing the approximation $\mathbb{B}'$ to $\mathbb{B}$

To being with we do not output $\mathbb{B}$ itself, but rather some approximation $\mathbb{B}'$ to $\mathbb{B}$. This piece of information consists of a list of "generalized hyperedges" plus a list of equalities. These generalized hyperedges consist of k-tuples of lists of vertex names: vertices in the same list will later be identified to create $\mathbb{B}$, but this is a separate construction. Some hyperedges also encode some extra vertex of $\mathbb{G}$ from which they were created. So a typical generalized hyperedge may look like $[V_1, V_2, \ldots, V_k]_e$, where e is some vertex of $\mathbb{G}$ (at height n) used for book-keeping purposes and the V_i are lists consisting of some vertices of $\mathbb{G}$ (of height 0) and some new vertices we create during the algorithm. Other hyperedges may not require the special book-keeping subscript.

Note that any *new* vertices created during the algorithm should be different each time (even though we often use x to denote such a vertex): we should use some counter on a fixed spare piece of tape for the entire algorithm; this counter is incremented at each creation of a new variable, and its value recorded within the new vertex name. (There will be only polynomially many new variables created, so only logspace used to store this one counter.)

1 To output the generalized hyperedges. There are two causes for writing generalized hyperedges to the output: the first is due to vertices at height n, and the second is due to vertices at height 0 that are the base vertex for some internal component with no top vertices. The generalized hyperedges will be written in such a way as to record some extra information that will be used for identifications.

For each vertex e at height n we will need to output a generalized hyperedge, however there may be many different vertices placed at a given position: these vertices will later be identified. We will also record in the encoding that the generalized hyperedge is created from vertex e. The following process is performed for each height n vertex e and in each case, we perform the following process for $i = 1$ to k.

1.1 Systematically search for an internal component C in which $i \in \Gamma(C)$ and for which e is a top vertex. These searches involve the following: we systematically search through all vertices of $\mathbb{G}$ until some u is found to be undirected-reachable from e amongst vertices not at height 0 or n. To avoid unnecessary duplication, we may also check that u does not lie in the same internal component as some earlier vertex (in which case we may ignore u: this internal component has already been considered). Then we proceed to systematically search through all vertices of $\mathbb{G}$ to identify the internal component C_u of $\mathbb{G}$ containing u. This component is then checked using Lemma 6.4(5) for whether $i \in \Gamma(C_u)$. If $i \in \Gamma(C)$ we go to substep 1.1.1. If $i \notin \Gamma(C_u)$ we increment u and continue our search for an internal component C with e as top and with $i \in \Gamma(C)$. If no such components are encountered we proceed to substep 1.1.2.

1.1.1 We have identified an internal component C with $i \in \Gamma(C)$ and for which e is a top vertex. If C has base vertices $b_1, \ldots, b_j$ then these will be written to the vertex set for the i^{th} coordinate of the output hyperedge. If C does not have base vertices, then we will create some *new* vertex x and write the vertex set $\{x\}$ to the i^{th} coordinate.

1.1.2 No internal component C is found with $i \in \Gamma(C)$ and for which e is a top vertex. In this case only one vertex will appear in the vertex set for coordinate i of this generalized hyperedge: a *new* vertex x.

1.2 Generalized hyperedges may also be created because of level 0 vertices. The following is performed for each level 0 vertex b and for each internal component C for which b is the base vertex and such that C has no top vertex. (If none are found there is nothing to do and no generalized hyperedge is written at step 1.2 for b.) We create a generalized hyperedge by performing the following checks for $i = 1, \dots, k$.

   1.2.1 If $i \in \Gamma(C)$ then $\{b\}$ is placed in position i of the generalized hyperedge,

   1.2.2 If $i \notin \Gamma(C)$ then a new vertex x is created and $\{x\}$ is placed in position i of the generalized hyperedge.

2 Finally we output information that will later be used to find certain vertices that will be forced to be identified in any satisfying interpretation of $\mathbb{G}$.

   2.1 For each pair of distinct height n vertices e, f, if e and f are the top vertex for the same internal component, then we write $e = f$ to the output tape.

   2.2 For each pair of distinct height 0 vertices b, c, if b and c are base vertices for the same internal component we write $b = c$.

This completes the construction of $\mathbb{B}'$. There are clearly further identifications that will be forced: for example, if b appears in the list of position i vertices for some generalized hyperedge e, and c appears in the list of position i vertices for some generalized hyperedge f, and if $e = f$ has been output, then we must have b and c identified. Accounting for these is stage 3B.

6.7 Stage 3B: construction of $\mathbb{B}$

We now need to construct $\mathbb{B}$ from the list of generalized hyperedges and equalities. The actual vertices of $\mathbb{B}$ will consist of sets of the vertices currently stated. If desired, this could be simplified as a later separate logspace process (such as by using only the earliest vertex from each set). Currently the input consists of generalized hyperedges where the entry in a given position is a set of vertices of $\mathbb{G}$ or new vertices. To create $\mathbb{B}$ we only need to amalgamate these sets, also taking into account the equality constraints.

In the following, a "vertex" refers to an element of some set within the position of some hyperedge. A "vertex set" consists of a set of vertices. The actual vertices of $\mathbb{B}$ will be vertex sets, produced from those appearing within $\mathbb{B}'$ by amalgamation.

The amalgamation process involves considering an undirected graph on the vertices, which we refer to as the *equality graph*. The undirected edges of the equality graph arise in several different ways.

(i) There will be an undirected edge from a vertex a to a vertex b if a and b lie within the same vertex set somewhere in the input list.

(ii) There will be an undirected edge from a vertex a to a vertex b if $a = b$ has been written as an equality constraint.

(iii) Recall that a hyperedge created from a height n vertex e records the vertex e in its description. We include an undirected edge between a and b if they appear in vertex sets at position i of two generalized hyperedges, either with the same label e, or with different labels e, f but where $e = f$ appears in the input.

There are logspace checks to recognise these undirected edges, and using the logspace solvability of undirected graph reachability we may determine if two vertices identified within our list of generalized hyperedges are connected in the equality graph, all within logspace.

For each vertex u we first check if there is some lexicographically earlier vertex v for which u and v are connected in the equality graph. If an earlier vertex is discovered, then we ignore u and continue to the next vertex. Otherwise, if no earlier vertex is discovered, we proceed to write down the vertex set of the component of the equality graph containing u. For each v lexicographically later than u, we check whether v is reachable from u (in the equality graph) and if so include it in vertex set of u. For the actual hyperedges of $\mathbb{B}$ we may simply write the existing generalized hyperedges (removing the book-keeping subscript), which can be read in the following way. A vertex set U appears in the i^{th} position of a hyperedge E if the intersection of U with the vertices listed for position i in E is nontrivial. Some hyperedges may be repeated in this output and obviously this could also be neatened by following with a totally new logspace reduction (even to an adjacency matrix).

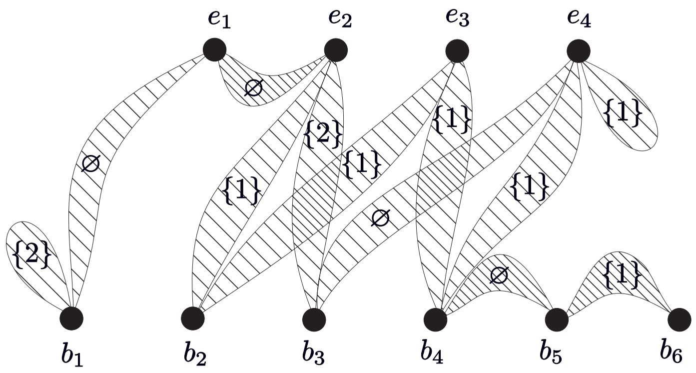
6.8 If and only if

Any homomorphism ϕ from $\mathbb{B}$ (the amalgamated "vertex sets") into $\mathbb{A}$, determines a function Φ from the height 0 vertices of $\mathbb{G}$ to the height 0 vertices of $\mathcal{D}(\mathbb{A})$. The construction of the hyperedges of $\mathbb{B}$ exactly reflects the satisfiability of the internal components of $\mathbb{G}$, so that the function Φ extends to cover all of $\mathbb{G}$ (this ignores any "short components" that we considered directly in stage 2A: but they cannot have been NO instances, as otherwise $\mathbb{B}$ was already created in stage 2 to be some fixed NO instance).

The converse is also true. First, we only grouped vertices together in vertex sets and their later amalgamation if they were forced to be identified in any possible interpretation in $\mathcal{D}(\mathbb{A})$. So any homomorphism f from $\mathbb{G}$ to $\mathcal{D}(\mathbb{A})$ determines a function F from the "vertices" of $\mathbb{B}$ to the vertices of $\mathbb{A}$. The hyperedges of $\mathbb{B}$ were determined by the internal components of $\mathbb{G}$, which f is satisfying within the encoding (in $\mathcal{D}(\mathbb{A})$) of the hyperedges of $\mathbb{A}$. So the $\mathbb{B}$ hyperedges are preserved by F.

6.9 An example

The following diagram depicts a reasonably general instance $\mathbb{G}$ of $\mathrm{CSP}(\mathcal{D}(\mathbb{A}))$ in the case that $\mathbb{A}$ itself is a digraph, so that $k = 2$. We are considering stage 3, so that $\mathbb{G}$ is a single connected digraph of height 4. The vertices at height 0 are $b_1, \dots, b_6$, and the vertices at height 4 are $e_1, \dots, e_4$. The shaded regions depict internal components: each is labelled by a subset of $\{1, 2\}$, depicting $\Gamma(C)$.



Let us examine how Stage 3A proceeds. We arrive at the first height 4 vertex e_1. For $i = 1$, discover no internal components with e_1 as the top, and with $1 \in \Gamma(C)$ (both have $\Gamma(C) = \varnothing$, so we are in case 1.1.2) and therefore return $\{x_1\}$ for the vertex set in the coordinate 1. For $i = 2$, we have the same outcome, so the generalized hyperedge that is actually written is $[\{x_1\}, \{x_2\}]_{e_1}$.

Then we proceed to the next height 4 vertex e_2. We encounter just one internal component C with $1 \in \Gamma(C)$, and its base vertices are $\{b_2\}$ (so this is in case 1.1.1). For $i = 2$ we also find just one internal component whose Γ value contains 2, and it has $\{b_3\}$ as the base vertices (also case 1.1.1). The generalized hyperedge $[\{b_2\}, \{b_3\}]_{e_3}$ is written.

For e_3 and $i = 1$ we encounter two internal components producing base vertices. We find b_2 as the only base vertex of the first, and b_4 for the second (case 1.1.1), so the first coordinate of the generalized hyperedge is $\{b_2, b_4\}$. For $i = 2$, no internal components yield a base vertex (case 1.1.2), so we output $\{x_3\}$. The actual generalized hyperedge written is $[\{b_2, b_4\}, \{x_3\}]_{e_2}$.

The vertex e_4 similarly results in the generalized hyperedge $[\{b_4, x_4\}, \{x_5\}]_{e_4}$.

This completes step 1.1 and we continue with step 1.2. We discover the height 0 vertex b_1 as the base of an internal component C with no top. We find $1 \notin \Gamma(C)$, so $\{x_6\}$ is written to the first coordinate of a generalized hyperedge (step 1.2.2 for $i = 1$). For $i = 2$ we find $2 \in \Gamma(C)$ so return $\{b_1\}$ for the second coordinate. The actual output written is $[\{x_6\}, \{b_1\}]$ (there are no subscripts to hyperedges from step 1.2). Level 0 vertices b_4 and b_5 also lead to the creation of generalized hyperedges. The overall output after the completion of steps 1.1 and 1.2 is

$$\begin{array}{ll}
[\{x_1\}, \{x_2\}]_{e_1} & \text{(from } e_1\text{, step 1.1)} \\
[\{b_2\}, \{b_3\}]_{e_2} & \text{(from } e_2\text{, step 1.1)} \\
[\{b_2, b_4\}, \{x_3\}]_{e_3} & \text{(from } e_3\text{, step 1.1)} \\
[\{b_4, x_4\}, \{x_5\}]_{e_4} & \text{(from } e_4\text{, step 1.1)} \\
[\{x_6\}, \{b_1\}] & \text{(from } b_1\text{, step 1.2)} \\
[\{x_7\}, \{x_8\}] & \text{(from } b_4\text{, step 1.2)} \\
[\{x_9\}, \{x_{10}\}] & \text{(from } b_5\text{, step 1.2)} \\
[\{b_5\}, \{x_{11}\}] & \text{(from } b_5\text{, step 1.2)} \\
[\{b_6\}, \{x_{12}\}] & \text{(from } b_6\text{, step 1.2)}
\end{array}$$

For step 2 of the algorithm, we output the following equalities

$$\begin{array}{ll}
e_1 = e_2, & \text{(from 2.1)} \\
b_4 = b_5, b_5 = b_6 & \text{(from 2.2)}
\end{array}$$

This completes Stage 3A: the list just given is $\mathbb{B}'$. We note that hyperedges such as $[\{x_7\}, \{x_8\}]$ will be no hinderance to satisfiability of $\mathbb{B}$ in $\mathbb{A}$, and we could word our algorithm to avoid writing these altogether.

Stage 3B then produces the digraph $\mathbb{B}$ with hyperedges

$$\begin{array}{l}
[\{b_2, b_4, b_5, b_6, x_1, x_4\}, \{b_3, x_2\}] \\
[\{b_2, b_4, b_5, b_6, x_1, x_4\}, \{b_3, x_2\}] \\
[\{b_2, b_4, b_5, b_6, x_1, x_4\}, \{x_3\}] \\
[\{b_2, b_4, b_5, b_6, x_1, x_4\}, \{x_5\}] \\
[\{x_6\}, \{b_1\}] \\
[\{x_7\}, \{x_8\}] \\
[\{x_9\}, \{x_{10}\}] \\
[\{b_2, b_4, b_5, b_6, x_1, x_4\}, \{x_{11}\}] \\
[\{b_2, b_4, b_5, b_6, x_1, x_4\}, \{x_{12}\}]
\end{array}$$

Which is a digraph with 12 vertices (namely, the 12 different sets of vertices appearing in hyperedges).

The algorithm itself is the composite of stage 3A and stage 3B.

7 Discussion

We conclude our paper with some applications and further research directions.

An example

Our construction allows us to create examples (and counterexamples) of digraph CSPs with certain desired properties, which were previously unknown or significantly harder to construct.

Example 7.1. *Let $\mathbb{A}$ be the structure on $\{0, 1\}$ with a single 4-ary relation*

$$R = \{(0, 0, 0, 1), (0, 1, 1, 1), (1, 0, 1, 1), (1, 1, 0, 1)\}.$$

Clearly $\mathbb{A}$ is a core. Using the fact that $R = \{(w, x, y, z) \in A^4 \mid w \oplus x = y \ \ \& \ \ z = 1\}$ (where $\oplus$ denotes addition modulo 2), it can be shown that the polymorphisms of $\mathbb{A}$ are the idempotent term functions of the two element group, and from this it follows that CSP($\mathbb{A}$) *is solvable by the few subpowers algorithm of [24], but is not bounded width. Then the CSP over the digraph $\mathcal{D}(\mathbb{A})$ is also solvable by few subpowers but is not bounded width (that is, is not solvable by local consistency check).*

Prior to the announcement of this example it had been temporarily conjectured by some researchers that solvability by the few subpowers algorithm implied solvability by local consistency check in the case of digraphs (this was the opening conjecture in Maróti's keynote presentation at the Second International Conference on Order, Algebra and Logics in Krakow 2011 for example). With 78 vertices and 80 edges, Example 7.1 also serves as a simpler alternative to the 368-vertex, 432-edge digraph whose CSP was shown by Atserias in [1, §4.2] to be tractable but not solvable by local consistency check.

In [25], Example 7.1, Corollary 5.2 and some fresh results on polymorphisms are used to construct digraph CSPs with every possible combination of the main polymorphism properties related to decision CSPs (allowing for Kazda's Maltsev implies majority result [28]).

Which properties are preserved?

Theorem 5.1 and Corollary 5.2 demonstrate that our reduction preserves almost all Maltsev conditions corresponding (or conjectured to be equivalent) to important algorithmic properties of decision CSPs. However, we were not able to extend our result to include all Maltsev conditions (in particular, nonbalanced identities in more than two variables). *Is it possible to characterize Maltsev conditions preserved by our construction? In particular, does it preserve all Maltsev conditions which hold in the zigzag?*

In [11] Bulatov established a dichtomy for counting CSPs (see also [17]). The algebraic condition separating tractable (FP) problems from #P-complete ones is called *congruence singularity.* It is not hard to see that the structure $\mathbb{A}$ from Example 7.1 satisfies this condition and thus the corresponding counting CSP is tractable. However, congruence singularity implies congruence permutability (i.e., having a Maltsev polymorphism) which fails in $\mathcal{D}(\mathbb{A})$. Therefore, counting CSP for $\mathcal{D}(\mathbb{A})$ is #P-complete. We conclude that our reduction does not preserve the complexity of counting. In fact, counting for $\mathcal{D}(\mathbb{A})$ is essentially always hard. *Is there a reduction of general CSPs to digraph CSPs which preserves complexity of counting?*

There are several other interesting variants or generalizations of CSPs in which algebraic conditions seem to play an important role as well. For example, infinite template CSPs (see below), valued CSPs [13, 40], or approximability of CSPs [14, 30]. *Can our construction be applied to obtain interesting results in these areas as well?*

Infinite template CSPs

CSPs over infinite templates are widely encountered in artificial intelligence; see [22, 29, 38] for example. Efforts to obtain a mathematical foundation for understanding the computational complexity of these problems have often involved assumptions of model theoretic properties on the template (such as ω-categoricity), as well as the presence of polymorphisms of certain kinds; see [6, 7, 8] for example. The results of the present article apply for such CSPs too: the proofs of Theorems 5.1 and 6.1 did not assume finiteness of A, only that $\mathbb{A}$ has only finitely many relations.

Remark. *Theorem 5.1 and Theorem 6.1 extend to infinite template CSPs consisting of only finitely many relations. Furthermore, since $\mathbb{A}$ and $\mathcal{D}(\mathbb{A})$ are first-order interdefinable, $\mathbb{A}$ is ω-categorical if and only if $\mathcal{D}(\mathbb{A})$ is ω-categorical.*

Special classes of CSPs

Hell and Rafiey [21] showed that all tractable list homomorphism problems over digraphs have the bounded width property, and from this it follows that there can be no translation from general CSPs to digraph CSPs preserving *conservative* polymorphisms (the polymorphisms related to list homomorphism problems). *Find a simple restricted class of list homomorphism problems for which there is a polymorphism-preserving translation from general list homomorphisms to the ones in this class.*

Another class of interest are the CSPs over generalized trees. *Is there a translation from generalized trees to oriented trees that preserves CSP tractability, or preserves polymorphism properties*?

Feder and Vardi's paper [18] also contains a polynomial reduction of general CSPs to CSPs over bipartite graphs. Payne and Willard announced preliminary results on a project similar to ours: to understand which Maltsev conditions are preserved by that reduction to bipartite graphs.

First order reductions

The logspace reduction in Lemma 3.4 *cannot* be replaced by first order reductions. Indeed, it is not hard to show that $\mathcal{D}(\mathbb{A})$ is never first order definable. More generally though, the only first order definable CSPs over balanced digraphs are the degenerate ones: over the single edge, or over a single vertex and no edges (see [26, Theorem C]), while deciding first order definability in general is NP-complete [32, Theorem 6.1]. Thus it seems unlikely that there is any other polynomial time computable construction to translate general CSPs to balanced digraph CSPs (as this would give P=NP). *Is there a different construction that translates general CSPs to* (*nonbalanced*!) *digraph CSPs with first order reductions in both directions*?

Various reductions of CSPs to digraphs

Feder and Vardi [18] and Atserias [1] provide polynomial time reductions of CSPs to digraph CSPs. We vigorously conjecture that their reductions preserve the properties of possessing a WNU polymorphism (and of being cores; but this is routinely verified). *Do these or other constructions preserve the precise arity of WNU polymorphisms*? *What other polymorphism properties are preserved*? *Do they preserve the bounded width property*?

Translations from general CSPs to digraph CSPs need not in general be as well behaved as the $\mathcal{D}$ construction of the present article. The third and fourth authors with Kowalski [25] have recently shown that a minor variation of the $\mathcal{D}$ construction preserves k-ary WNU polymorphisms (and thus the properties of being Taylor and having bounded width) but always fails to preserve many other polymorphism properties (such as those witnessing strict width, or the few subpowers property).

Acknowledgements

The authors would like to thank Libor Barto, Marcin Kozik, Miklós Maróti and Barnaby Martin for their thoughtful comments and discussions and Andrei Krokhin for carefully reading the paper and pointing out a number of technical issues. The first author was supported by the Grant Agency of the Czech Republic project no. 13-01832S, Charles University Grant Agency projects 558313 and SVV-2014-260107, Ministry of Education of the Czech Republic project 7AMB 13P10 13 and the Austrian Science Fund project P24285. The second author gratefully acknowledges support by the Natural Sciences and Engineering Research Council of Canada in the form of a Discovery Grant. The third and fourth authors were supported by ARC Discovery Project DP1094578, while the third author was also supported by Future Fellowship FT120100666.

References

[1] Atserias, A.: On digraph coloring problems and treewidth duality. European J. Combin. 29(4), 796–820 (2008)

[2] Barto, L., Kozik, M.: Absorbing subalgebras, cyclic terms, and the constraint satisfaction problem. Log. Methods Comput. Sci. 8(1), 1:07, 27 (2012)

[3] Barto, L., Kozik, M.: Constraint satisfaction problems solvable by local consistency methods. J. ACM 61(1), 3:1–3:19 (Jan 2014)

[4] Barto, L., Kozik, M., Niven, T.: The CSP dichotomy holds for digraphs with no sources and no sinks (a positive answer to a conjecture of Bang-Jensen and Hell). SIAM J. Comput. 38(5), 1782–1802 (2008/09)

[5] Berman, J., Idziak, P., Marković, P., McKenzie, R., Valeriote, M., Willard, R.: Varieties with few subalgebras of powers. Trans. Amer. Math. Soc. 362(3), 1445–1473 (2010)

[6] Bodirsky, M.: Constraint Satisfaction Problems with Infinite Templates. In: Creignou, N., Kolaitis, P.G., Vollmer, H. (eds.) Complexity of Constraints. Lecture Notes in Computer Science, vol. 5250, pp. 196–228. Springer (2008)

[7] Bodirsky, M., Chen, H.: Qualitative Temporal and Spatial Reasoning Revisited. J. Log. Comput. 19(6), 1359–1383 (2009)

[8] Bodirsky, M., Hils, M., Martin, B.: On the scope of the universal-algebraic approach to constraint satisfaction. Log. Methods Comput. Sci. 8(3), 3:13, 30 (2012)

[9] Bulatov, A., Jeavons, P., Krokhin, A.: Classifying the complexity of constraints using finite algebras. SIAM J. Comput. 34(3), 720–742 (2005)

[10] Bulatov, A.A.: A dichotomy theorem for constraint satisfaction problems on a 3-element set. J. ACM 53(1), 66–120 (2006)

[11] Bulatov, A.A.: The complexity of the counting constraint satisfaction problem. J. ACM 60(5), 34 (2013)

[12] Bulín, J., Delić, D., Jackson, M., Niven, T.: On the reduction of the CSP dichotomy conjecture to digraphs. In: Schulte, C. (ed.) CP. Lecture Notes in Computer Science, vol. 8124, pp. 184–199. Springer (2013)

[13] Cohen, D.A., Cooper, M.C., Creed, P., Jeavons, P.G., Živný, S.: An algebraic theory of complexity for discrete optimization. SIAM Journal on Computing 42(5), 1915–1939 (2013)

[14] Dalmau, V., Krokhin, A.: Robust satisfiability for CSPs: Hardness and algorithmic results. ACM Trans. Comput. Theory 5(4) (2013)

[15] Dalmau, V., Larose, B.: Maltsev + Datalog $\Rightarrow$ Symmetric Datalog. In: Proceedings of the 23rd IEEE Symposium on Logic in Computer Science. pp. 297–306. LICS '08 (2008)

[16] Dalmau, V., Pearson, J.: Closure Functions and Width 1 Problems. In: Jaffar, J. (ed.) CP. Lecture Notes in Computer Science, vol. 1713, pp. 159–173. Springer (1999)

[17] Dyer, M.E., Richerby, D.: An effective dichotomy for the counting constraint satisfaction problem. SIAM J. Comput. 42(3), 1245–1274 (2013)

[18] Feder, T., Vardi, M.Y.: The computational structure of monotone monadic SNP and constraint satisfaction: a study through Datalog and group theory. SIAM J. Comput. 28(1), 57–104 (electronic) (1999)

[19] Hell, P., Nešetřil, J.: On the complexity of H-coloring. J. Combin. Theory Ser. B 48(1), 92–110 (1990)

[20] Hell, P., Nešetřil, J.: Graphs and homomorphisms, Oxford Lecture Series in Mathematics and its Applications, vol. 28. Oxford University Press, Oxford (2004)

[21] Hell, P., Rafiey, A.: The dichotomy of list homomorphisms for digraphs. In: Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms. pp. 1703–1713. SIAM, Philadelphia, PA (2011)

[22] Hirsch, R.: Expressive Power and Complexity in Algebraic Logic. J. Log. Comput. 7(3), 309–351 (1997)

[23] Hobby, D., McKenzie, R.: The structure of finite algebras, Contemporary Mathematics, vol. 76. American Mathematical Society, Providence, RI (1988)

[24] Idziak, P., Marković, P., McKenzie, R., Valeriote, M., Willard, R.: Tractability and learnability arising from algebras with few subpowers. SIAM J. Comput. 39(7), 3023–3037 (2010)

[25] Jackson, M., Kowalski, T., Niven, T.: Digraph related constructions and the complexity of digraph homomorphism problems (2013), arXiv:1304.4986 [math.CO]

[26] Jackson, M., Trotta, B.: Constraint Satisfaction, Irredundant Axiomatisability and Continuous Colouring. Studia Logica 101(1), 65–94 (2013)

[27] Jeavons, P., Cohen, D., Gyssens, M.: Closure properties of constraints. J. ACM 44(4), 527–548 (1997)

[28] Kazda, A.: Maltsev digraphs have a majority polymorphism. European J. Combin. 32(3), 390–397 (2011)

[29] Krokhin, A., Jeavons, P., Jonsson, P.: Reasoning About Temporal Relations: The Tractable Subalgebras Of Allen's Interval Algebra. Journal of the ACM 50, 2003 (2001)

[30] Kun, G., O'Donnell, R., Tamaki, S., Yoshida, Y., Zhou, Y.: Linear programming, width-1 CSPs, and robust satisfaction. In: Goldwasser, S. (ed.) ITCS. pp. 484–495. ACM (2012)

[31] Ladner, R.E.: On the structure of polynomial time reducibility. J. Assoc. Comput. Mach. 22, 155–171 (1975)

[32] Larose, B., Loten, C., Tardif, C.: A characterisation of first-order constraint satisfaction problems. Log. Methods Comput. Sci. 3(4), 4:6, 22 (2007)

[33] Larose, B., Tesson, P.: Universal algebra and hardness results for constraint satisfaction problems. Theor. Comput. Sci. 410(18), 1629–1647 (2009)

[34] Maróti, M., McKenzie, R.: Existence theorems for weakly symmetric operations. Algebra Universalis 59(3-4), 463–489 (2008)

[35] Montanari, U.: Networks of constraints: fundamental properties and applications to picture processing. Information Sci. 7, 95–132 (1974)

[36] Reingold, O.: Undirected st-connectivity in log-space. In: Gabow, H.N., Fagin, R. (eds.) STOC. pp. 376–385. ACM (2005)

[37] Reingold, O.: Undirected connectivity in log-space. J. ACM 55(4) (2008)

[38] Renz, J.: Qualitative Spatial Reasoning with Topological Information, Lecture Notes in Computer Science, vol. 2293. Springer (2002)

[39] Schaefer, T.J.: The complexity of satisfiability problems. In: Conference Record of the Tenth Annual ACM Symposium on Theory of Computing (San Diego, Calif., 1978), pp. 216–226. ACM, New York (1978)

[40] Thapper, J., Živný, S.: The complexity of finite-valued CSPs. In: Proceedings of the Forty-fifth Annual ACM Symposium on Theory of Computing. pp. 695–704. STOC '13, ACM, New York, NY, USA (2013)

Part II

Decidability of absorption

Decidability of absorption in relational structures of bounded width

Jakub Bulín

Abstract

Absorption theory of Barto and Kozik has proven to be a very useful tool in the algebraic approach to the Constraint Satisfaction Problem and structure of finite algebras in general. We address the following problem: Given a finite relational structure $\mathbb{A}$ and a subset $B \subseteq A$, is it decidable whether B is an absorbing subuniverse? We provide an affirmative answer in the case when $\mathbb{A}$ has bounded width (i.e., the algebra of polymorphisms of $\mathbb{A}$ generates a congruence meet semidistributive variety). As a by-product we confirm that in this case the notion of Jónsson absorption coincides with the usual absorption. We also show that several open questions about absorption in relational structures can be reduced to digraphs.

Introduction

Universal algebra has recently found a fruitful application to theoretical computer science in the field of Constraint Satisfaction Problem (`CSP`). At the core of the so-called algebraic approach to the `CSP` lies the fact that every relational structure can be naturally associated with an algebra (the *algebra of polymorphisms*), which allows application of deep algebraic theories to the originally purely combinatorial problem (see [8]). This led to a rapid progress towards the `CSP` dichotomy conjecture of Feder and Vardi [10]. The interplay between constraint satisfaction and universal algebra has evolved into a mutually beneficial relationship. Studying `CSP`s has uncovered new (and sometimes surprising) properties of finite algebras and tools to study their structure. A prominent example is the theory of absorption of Barto and Kozik.

One of the milestones on the path towards the `CSP` dichotomy conjecture was the problem of characterizing `CSP`s solvable by *local consistency checking* or, equivalently, relational structures of *bounded width.* Larose and Zádori [11] conjectured that a finite core relational structure has bounded width, if and only if its algebra of polymorphisms generates a congruence meet semidistributive (SD($\wedge$)) variety. In the same paper they proved the "only if" part. Barto and Kozik solved this conjecture, first in the congruence distributive (CD) case [4] and later in the general case [5]. An essential idea of the proof of the so-called Bounded Width Theorem is that the instance of the `CSP` can be reduced to certain subsets of the relational structure, the *absorbing subuniverses* of its algebra of polymorphisms. This idea can be traced back to the CD case where a similar notion is used: *Jónsson ideals.*

The idea of absorption has proven to be very useful in a number of other problems related to the CSP and finite algebras. We refer the reader to [3] and [6] for some of the applications. Absorption theory is still in its beginning and there are many open problems that need to be understood. Our paper provides partial solutions to a few of them.

In [2] Barto used absorption to provide a new algorithm for solving conservative CSPs, significantly simplifying a result of Bulatov [7]. This algorithm uses the knowledge of absorbing subuniverses of the algebra of polymorphisms as a blackbox, which led Barto to formulate the following problem.

Problem 1 (Problem 24 in [14]). Given a finite relational structure $\mathbb{A}$ and a subset B, is it decidable whether B is an absorbing subuniverse of the algebra of polymorphisms of $\mathbb{A}$?

The main result of our paper is a partial solution to this problem. We prove that if the algebra of polymorphisms of $\mathbb{A}$ generates an SD($\wedge$) variety, then there is a co-NEXPTIME algorithm to test for absorption (EXPTIME if one assumes that the input subset B is always a subuniverse).

The above problem can be seen as a generalization of the following problem: Given a finite relational structure $\mathbb{A}$, is it decidable whether $\mathbb{A}$ admits a *near unanimity* polymorphism? A positive answer to this problem was given by Barto in [1], where he proves the Zádori conjecture: Every finite relational structure whose algebra of polymorphisms generates a CD variety admits a near unanimity polymorphism. Our proof mimics the proof of Barto. The idea is to encode the problem of existence of an absorbing term as an instance of the CSP over $\mathbb{A}$.

Our proof shows that every absorption in a finite SD($\wedge$) relational structure is realized by a term whose arity is bounded by the size of its universe and arities of its relations. This bound is doubly exponential and Problem 25 in [14] asks for a better bound. Also, as a by-product, we obtain a proof that in the SD($\wedge$) case Jónsson ideals (a generalization of the notion used in [4]) coincide with absorbing subuniverses.

In the second part of our paper we show that many questions about absorption in relational structures can be reduced to digraphs. For that, we use the construction from [9] and prove that it preserves the property of being an absorbing subuniverse as well as the arity of absorbing terms and also, in a sense, it preserves the absorption free subuniverses. We conclude by discussing the problem of characterizing finite algebras which generate a pseudovariety containing no absorption free members – another open problem in absorption theory.

1 Algebras and relational structures

In this section we briefly present notions and fix notation used throughout the paper.

A *k-ary operation* on a set A is a mapping $f\colon A^k \to A$. By an *algebra* we mean a pair $\mathbf{A} = (A; \mathcal{F})$, where A is a set and $\mathcal{F}$ is a set of operations on A. A subset $B \subseteq A$ is a *subuniverse* of $\mathbf{A}$ (denoted by $B \leq \mathbf{A}$) if it is closed under all operations from $\mathcal{F}$. An operation is *idempotent* if it satisfies $f(x, x, \dots, x) \approx x$. An algebra is idempotent if all of its operations are idempotent. (Equivalently, $\{a\} \leq \mathbf{A}$ for every $a \in A$.)

Term operations of $\mathbf{A}$ (denoted by $\operatorname{Clo}\mathbf{A}$) are operations that can be obtained from $\mathcal{F}$ and the projection operations using composition of operations. A *semilattice operation* is a binary idempotent, commutative and associative operation. A *near unanimity operation* is an operation (at least ternary) satisfying the identities $t(x,\dots,x,y) \approx t(x,\dots,x,y,x) \approx \dots \approx t(y,x,\dots,x) \approx x$.

The *variety* generated by $\mathbf{A}$, $\mathrm{HSP}(\mathbf{A})$, is the class of all algebras constructed from $\mathbf{A}$ by taking products (P), subalgebras (S) and homomorphic images (H). A variety is *congruence meet semidistributive* ($\mathrm{SD}(\wedge)$) if the congruence lattices of all of its members satisfy the meet semidistributive law, and similarly for *congruence distributivity* (CD); $\mathbf{A}$ is said to be CD or $\mathrm{SD}(\wedge)$ if the variety generated by $\mathbf{A}$ is. For $\mathbf{A}$ finite, the *pseudovariety* generated by $\mathbf{A}$, $\mathrm{HSP}_{\mathrm{fin}}(\mathbf{A})$, is the class of all finite members of $\mathrm{HSP}(\mathbf{A})$.

An *n-ary relation* on A is a subset $R \subseteq A^n$. A *relational structure* is a pair $\mathbb{A} = (A;\mathcal{R})$, where A is a set and $\mathcal{R} = \{R_1,\dots,R_m\}$ is a finite set of relations on A. We say that a relation R is *pp-definable* from a relational structure $\mathbb{A}$ if there exists a *primitive positive* formula φ (i.e., an existentially quantified conjunction of atomic formulæ) in the language of $\mathbb{A}$ such that $(a_1,\dots,a_n) \in R$ if and only if $\varphi(a_1,\dots a_n)$ holds in $\mathbb{A}$.

A *polymorphism* of $\mathbb{A}$ is an operation on A which *preserves* all relations from $\mathcal{R}$. That is, if $(a_1^1,\dots,a_n^1),(a_1^2,\dots,a_n^2),\dots,(a_1^k,\dots,a_n^k) \in R$ for some n-ary relation $R \in \mathcal{R}$, then $(f(a_1^1,\dots,a_1^k),\dots,f(a_n^1,\dots,a_n^k)) \in R$. The set of all polymorphisms of $\mathbb{A}$ is denoted by $\operatorname{Pol}\mathbb{A}$.

To every relational structure $\mathbb{A}$ we can associate in a natural way an algebra: the *algebra of polymorphisms of* $\mathbb{A}$ is the algebra $\mathbf{alg}\,\mathbb{A} = (A;\operatorname{Pol}\mathbb{A})$. Relations pp-definable from $\mathbb{A}$ are precisely subuniverses of finite powers of the algebra $\mathbf{alg}\,\mathbb{A}$ (see [8] for details). An algebra is *finitely related* if it has the same term operations as the algebra of polymorphisms of some relational structure.

A relational structure is *binary* if it has binary relations only. A *digraph* is a binary relational structure with just one relation, the *edge* relation.

2 Absorption

The definitions, examples and observations in this section are standard in the theory of absorption of Barto and Kozik (see [3], [6]). We begin with the definition of an *absorbing subuniverse*.

Definition 2.1. Let $\mathbf{A}$ be an algebra and $B \leq \mathbf{A}$. We say that B is an *absorbing subuniverse* of $\mathbf{A}$ (and write $B \trianglelefteq \mathbf{A}$), if there exists an idempotent $t \in \operatorname{Clo}\mathbf{A}$ such that

$$\begin{gathered} t(A,B,B,\dots,B,B) \subseteq B, \\ t(B,A,B,\dots,B,B) \subseteq B, \\ \vdots \\ t(B,B,B,\dots,B,A) \subseteq B. \end{gathered}$$

The notion of absorbing subuniverse was motivated by algebras with a near unanimity operation, namely the following characterization.

Lemma 2.2. *A finite idempotent algebra* $\mathbf{A}$ *has a near unanimity term operation, if and only if* $\{a\} \trianglelefteq \mathbf{A}$ *for every* $a \in A$.

Proof. The proof is easy using the same trick as in [12], Lemma 3.10. If $\{b\} \trianglelefteq \mathbf{A}$ via a k-ary term t_b and $\{c\} \trianglelefteq \mathbf{A}$ via an n-ary t_c, then both of these absorptions are witnessed by the kn-ary term operation

$$t_{bc}(x_1, \ldots, x_{kn}) = t_b(t_c(x_1, \ldots, x_n), \ldots, t_c(x_{(k-1)n+1}, \ldots, x_{kn})).$$

Hence we can construct a term operation t witnessing $\{a\} \trianglelefteq \mathbf{A}$ for every $a \in A$; this is equivalent to t being a near unanimity operation. ☐

In applications of absorption theory an important role is played by algebras with no proper absorbing subuniverses, the *absorption free* algebras.

Definition 2.3. An algebra $\mathbf{A}$ is *absorption free*, if $|A| > 1$ and $B \trianglelefteq \mathbf{A}$ implies that $B = A$ or $B = \emptyset$.

An example of an absorption free algebra is the two element group.

The origins of absorption theory can be traced back to the study of the `CSP` for congruence distributive relational structures (see [4, Definition 6.5]). Let $\mathbf{A}$ be a CD algebra with a fixed Jónsson chain $p_1, \ldots, p_n$. A *Jónsson ideal* of $\mathbf{A}$ is a subuniverse $B \leq \mathbf{A}$ satisfying $p_i(B, A, B) \subseteq B$ for all i. Studying Jónsson ideals inspired the definition of absorbing subuniverse in [5].

Kozik recently discovered a new Maltsev condition, the so-called *directed Jónsson terms*, which characterizes finite CD algebras.

Theorem 2.4 (Kozik, M., personal communication, 2011)**.** *A finite algebra* $\mathbf{A}$ *is* CD*, if and only if there exists a sequence of ternary idempotent terms* $p_1, \ldots, p_n \in \operatorname{Clo} \mathbf{A}$ *satisfying the following identities.*

$$\begin{aligned} &p_1(x,x,y) \approx x,\ p_n(x,y,y) \approx y, \\ &p_i(x,y,y) \approx p_{i+1}(x,x,y) \text{ for all } i < n, \\ &p_i(x,y,x) \approx x \text{ for all } i \leq n. \end{aligned}$$

The notion of *$\mathcal{J}$-absorbing subuniverse* is a generalization of a "directed" Jónsson ideal.

Definition 2.5. Let $\mathbf{A}$ be an algebra and $B \leq \mathbf{A}$. We call B a *$\mathcal{J}$-absorbing subuniverse* of $\mathbf{A}$ ($B \trianglelefteq_{\mathcal{J}} \mathbf{A}$), if there exists a sequence of ternary idempotent terms $p_1, \ldots, p_n \in \operatorname{Clo} \mathbf{A}$ such that

$$\begin{aligned} &p_1(x,x,y) \approx x,\ p_n(x,y,y) \approx y, \\ &p_i(x,y,y) \approx p_{i+1}(x,x,y) \text{ for all } i < n \end{aligned}$$

and

$$p_i(B, A, B) \subseteq B \text{ for all } i \leq n.$$

$\mathcal{J}$-absorbing subuniverses are sometimes also called Jónsson ideals and the corresponding sequence of terms a Jónsson chain (of length n).

Similarly as with absorption, the following holds.

Lemma 2.6. *A finite idempotent algebra* $\mathbf{A}$ *is* CD*, if and only if* $\{a\} \trianglelefteq_{\mathcal{J}} \mathbf{A}$ *for every* $a \in A$*.*

Proof. Let $\{b\} \trianglelefteq_{\mathcal{J}} \mathbf{A}$ via a Jónsson chain $p_1, \dots, p_k$ and $\{c\} \trianglelefteq_{\mathcal{J}} \mathbf{A}$ via $q_1, \dots, q_n$. If we define $r_{ij}(x, y, z) = p_i(x, q_j(x, y, z), z)$ for $i \le k$ and $j \le n$, then it can be routinely verified that $r_{11}, \dots, r_{1n}, r_{21}, \dots, r_{kn}$ is a Jónsson chain witnessing both $\{b\} \trianglelefteq_{\mathcal{J}} \mathbf{A}$ and $\{c\} \trianglelefteq_{\mathcal{J}} \mathbf{A}$. Consequently, we can construct a single Jónsson chain witnessing $\{a\} \trianglelefteq_{\mathcal{J}} \mathbf{A}$ for every $a \in A$, which is equivalent to the directed Jónsson terms from Theorem 2.4. ☐

The next lemma shows that an absorbing subuniverse is always $\mathcal{J}$-absorbing. (The idea is the same as in the syntactical proof that an algebra with a near-unanimity term operation is CD.)

Lemma 2.7. *Let* $\mathbf{A}$ *be an algebra and* $B \le \mathbf{A}$*. If* $B \trianglelefteq \mathbf{A}$ *via an* n*-ary term* t*, then* $B \trianglelefteq_{\mathcal{J}} \mathbf{A}$ *via a Jónsson chain of length* n*.*

Proof. We can define the Jónsson chain in the following way: $p_1(x, y, z) = t(y, x, \dots, x)$, $p_n(x, y, z) = t(z, \dots, z, y)$, and $p_i(x, y, z) = t(z, \dots, z, y, x \dots, x)$ (where y is in the ith place) for $1 < i < n$. ☐

However, the converse is not true in general as we can see in the following example.

Example (The implication algebra)**.** Let $\mathbf{A} = (\{0, 1\}; \to)$, where $\to$ denotes the logical implication (as a binary operation). It is well known that $\mathbf{A}$ generates a congruence distributive variety, but does not have a near unanimity term operation. Hence both $\{0\}$ and $\{1\}$ are $\mathcal{J}$-absorbing, but not absorbing.

In Theorem 3.1 we prove that the two notions coincide in finite, finitely related congruence meet semidistributive algebras. The precise relation between absorption and $\mathcal{J}$-absorption is yet to be understood. We conclude this section with a notational remark.

Definition 2.8. Let $\mathbb{A}$ be a finite relational structure and $B \subseteq \mathbb{A}$. For brevity, we write $B \le \mathbb{A}$, $B \trianglelefteq \mathbb{A}$ and $B \trianglelefteq_{\mathcal{J}} \mathbb{A}$ instead of $B \le \mathbf{alg}\,\mathbb{A}$, $B \trianglelefteq \mathbf{alg}\,\mathbb{A}$ and $B \trianglelefteq_{\mathcal{J}} \mathbf{alg}\,\mathbb{A}$ (respectively). Moreover, we say that $\mathbb{A}$ is SD($\wedge$) meaning that its algebra of polymorphisms generates a congruence meet semidistributive variety.

Throughout the rest of our paper, all results about finite relational structures can be equivalently stated for finite, finitely related algebras.

3 Absorption in SD($\wedge$) structures

The following theorem and its corollary state the main results of our paper.

Theorem 3.1. *Let* $\mathbb{A}$ *be a finite* SD($\wedge$) *relational structure and* $B \subseteq A$*. The following holds:*

(i) If $B \trianglelefteq_{\mathcal{J}} \mathbb{A}$*, then* $B \trianglelefteq \mathbb{A}$*.*

(ii) If $B \trianglelefteq \mathbb{A}$*, then this absorption is realized via some term of arity at most* $4^{8^{|A|^k}} + 1$*, where* k *is the maximum arity of a relation of* $\mathbb{A}$*.*

We will prove Theorem 3.1 in Section 5. This theorem provides an algorithm to test absorption:

Corollary 3.2. *Given a finite* SD($\wedge$) *relational structure* $\mathbb{A}$ *and* $B \subseteq A$*, there exists a* `co-NEXPTIME` *algorithm that checks whether* $B \trianglelefteq \mathbb{A}$*.*

Proof. By the previous theorem, $B \trianglelefteq \mathbb{A}$ if and only if $B \trianglelefteq_{\mathcal{J}} \mathbb{A}$. Testing whether B is a subuniverse is in `co-NEXPTIME` (and might even be complete for this class, see [13]). Once this is established, we can check in `EXPTIME` whether there exists a Jónsson chain witnessing $B \trianglelefteq_{\mathcal{J}} \mathbb{A}$ in the same manner as in [1, Corollary 7.1]. (Basically, generate all ternary polymorphisms and check for a Jónsson chain there.) $\square$

Let us note that our proof uses the Bounded width theorem [5], and hence this approach cannot be generalized beyond the realm of congruence meet semidistributivity. (Still, the origins of absorption theory as well as a large portion of its applications do come from this context.) See the next section for details.

The algorithm that we provide may not be optimal. Some evidence suggests that there could be a faster algorithm (see [1, Subsection 7.1] for a related discussion). The following problem remains open.

Problem 2. Given an SD($\wedge$) relational structure and $B \subseteq \mathbb{A}$, determine the complexity of deciding whether $B \trianglelefteq \mathbb{A}$.

Our proof gives a double exponential bound on the minimal arity of an absorbing polymorphism, which may not be optimal.

Problem 3 (Problem 25 in [14])**.** Given an SD($\wedge$) relational structure and $B \trianglelefteq \mathbb{A}$, provide a better bound on the minimal arity of a polymorphism witnessing this absorption.

4 Instances of the `CSP`

In this section we briefly introduce a few definitions concerning instances of the `CSP` and the key result from [1] needed for the proof of Theorem 3.1. A more detailed treatment with proofs can be found in [1, Sections 5 and 6].

An *instance* of the `CSP` is a triple $\mathcal{P} = (V, A, \mathcal{C})$, where A and V are finite sets (the set of *variables* and the *domain*) and $\mathcal{C}$ is a finite set of *constraints*. Each constraint $C \in \mathcal{C}$ is a tuple $C = (\bar{x}, R)$, where $\bar{x} \in V^m$ and R is an m-ary relation on A. A *solution* to $\mathcal{P}$ is a mapping $\varphi\colon V \to A$ such that $\varphi(\bar{x}) \in R$ for every $(\bar{x}, R) \in \mathcal{C}$.

Let $\mathbf{A}$ be a finite idempotent algebra. An instance $\mathcal{P} = (V, A, \mathcal{C})$ is an instance of the `CSP` *over* $\mathbf{A}$ if all the relations in the constraints from $\mathcal{C}$ are subalgebras of finite powers of $\mathbf{A}$.

Definition 4.1. Let $\mathcal{P} = (V, A, \mathcal{C})$ be an instance of the `CSP`.

- $\mathcal{P}$ is a *simple binary instance*, if $\mathcal{C} = \{((x, y), R_{x,y}) \mid x, y \in V\}$, where $R_{x,y} = R_{y,x}^{-1}$ and $R_{x,x} \subseteq \{(a, a) \mid a \in A\}$ for all $x, y \in V$.

- $\mathcal{P}$ is a $(1,2)$-*system* with unary projections $\{R_x \mid x \in V\}$, if it is a simple binary instance, $\emptyset \neq R_x \subseteq A$ and for all $x, y \in V$, the projection of $R_{x,y}$ to the 1st coordinate is R_x.
- $\mathcal{P}$ is a $(2,3)$-*system*, if it is a $(1,2)$-system and for every $x, y, z \in V$ and $(a,b) \in R_{x,y}$ there exists $c \in A$ such that $(a,c) \in R_{x,z}$ and $(b,c) \in R_{y,z}$.

Let $\mathcal{P}$ be a simple binary instance and $\mathcal{S} = \{S_x \mid x \in V\}$ a family of subsets of A. The *restriction of* $\mathcal{P}$ *to* $\mathcal{S}$ is the simple binary instance $\mathcal{P}|_{\mathcal{S}}$ with $R^{\mathcal{P}|_{\mathcal{S}}}_{x,y} = R^{\mathcal{P}}_{x,y} \cap (S_x \times S_y)$.

When considering simple binary instances it is natural to talk about "realizing" trees labeled by variables.

Definition 4.2. Let $\mathcal{P} = (V, A, \mathcal{C})$ be a simple binary instance of the `CSP`.

- A $\mathcal{P}$-*tree* T is a tree (i.e., an undirected connected graph without cycles) together with a labeling $\mathrm{lbl}\colon \mathrm{V}(T) \to V$.
- A *realization of* T is a mapping $\mathrm{r}\colon \mathrm{V}(T) \to A$ such that $(\mathrm{r}(u), \mathrm{r}(v)) \in R_{\mathrm{lbl}(u),\mathrm{lbl}(v)}$ whenever $(u,v) \in \mathrm{E}(T)$.

The following theorem is the key to the proof of the Zádori conjecture as well as Theorem 3.1.

Theorem 4.3. *Let $\mathcal{P} = (V, A, \mathcal{C})$ be a $(2,3)$-system with unary projections $\{R_x \mid x \in V\}$ over an $\mathrm{SD}(\wedge)$ algebra $\mathbf{A}$ and let $\mathcal{S} = \{S_x \mid x \in V\}$ be a family of nonempty subuniverses of A such that $S_x \trianglelefteq_{\mathcal{J}} R_x$. If all $\mathcal{P}$-trees with at most $4^{8^{|A|}}$ vertices are realizable in $\mathcal{P}|_{\mathcal{S}}$, then $\mathcal{P}|_{\mathcal{S}}$ has a solution.*

This theorem is a minor refinement of results from [1]. The proof can be extracted from there, namely from Theorem 5.6, Proposition 5.3, Remark 5.4 and Theorem 5.7. We omit the proof, as it is quite long and technical. We are making only two refinements:

First, we relax the condition on $\mathbf{A}$ from being congruence distributive to meet semidistributive. The only time this assumption is used is when invoking Theorem 4.5 from [5] which says that every $(2,3)$-system over an $\mathrm{SD}(\wedge)$ algebra has a solution (see [1, Remark 5.8]). In fact, this property characterizes finite $\mathrm{SD}(\wedge)$ algebras. Note that this is the bottleneck of our method: one cannot use this approach for relational structures without bounded width.

Second, we are working with $\mathcal{J}$-absorbing subuniverses instead of Jónsson ideals, which slightly simplifies the proof. The only place in the proof where this difference matters is in [1, Lemma 6.1], namely in the "minimal counterexample" situation. The proof can be easily adapted; we present the new version here.

Lemma 4.4. *Let $\mathbf{B}$ be a finite algebra, $U \subseteq B$, $a \in U$, $b \in B \backslash U$ and $E \leq F \leq \mathbf{B}^2$ such that $E \trianglelefteq_{\mathcal{J}} F$. Assume that $(a,a) \in E$, $(b,b) \in E$ and $(a,b) \in F$. Then there exists $c \in U$ and $d \in B \setminus U$ such that $(c,d) \in E$.*

Proof. Let $p_1, \dots, p_n$ be a Jónsson chain witnessing $E \trianglelefteq_{\mathcal{J}} F$. We define a sequence of elements $c_0, c_1, \dots, c_n$ as follows:

$$\begin{aligned} c_0 &= p_1(a,a,b) = a, \\ c_i &= p_i(a,b,b) = p_{i+1}(a,a,b) \text{ for } 0 < i < n, \\ c_n &= p_n(a,b,b) = b. \end{aligned}$$

Since $(a,a), (b,b) \in E$ and $(a,b) \in F$, it follows that $(p_i(a,a,b), p_i(a,b,b)) \in E$, and thus $c_0, c_1, \dots, c_n$ is a directed path from a to b in the digraph $(B; E)$. As $a \in U$ and $b \notin U$, there exists i such that $c_{i-1} \in U$ and $c_i \in B \setminus U$. □

5 Proof of Theorem 3.1

In [1], the proof of the Zádori conjecture is significantly simplified by first reducing to binary relational structures. We use the same construction.

Lemma 5.1. *Let $\mathbb{A}$ be a relational structure and let k be the maximum arity of a relation of $\mathbb{A}$. There exists a binary relational structure $\bar{\mathbb{A}}$ with universe A^k satisfying the following:*

(i) $\mathbf{alg}\,\bar{\mathbb{A}} = (\mathbf{alg}\,\mathbb{A})^k$ *(and hence they generate the same variety).*

(ii) For any $B \subseteq A$, $B \trianglelefteq \mathbb{A}$ via an n-ary term iff $B^k \trianglelefteq \bar{\mathbb{A}}$ via an n-ary term. Similarly, $B \trianglelefteq_{\mathcal{J}} \mathbb{A}$ iff $B^k \trianglelefteq_{\mathcal{J}} \bar{\mathbb{A}}$.

Proof. For details of the construction we refer the reader to [1, Proposition 3.1]. Statement (i) is proved there and (ii) follows in the same fashion. □

Using Lemma 5.1, it's enough to prove that given a finite binary SD($\wedge$) relational structure $\mathbb{A}$ and $B \trianglelefteq_{\mathcal{J}} \mathbb{A}$, there exists $s \in \operatorname{Pol}\mathbb{A}$ of arity $4^{8^{|A|}} + 1$ such that $B \trianglelefteq \mathbb{A}$ via s. We can assume that B is nonempty, otherwise the claim is trivial.

Similarly as in [1], the idea is to encode n-ary polymorphisms witnessing absorption as solutions to an instance of the CSP and then use Theorem 4.3 to prove that for n "big enough" there is a solution to this instance. Let $n = 4^{8^{|A|}} + 1$. We define a simple binary instance $\mathcal{P} = (A^n, A, \{((x,y), R_{x,y}) \mid x, y \in A^n\})$, where

$$R_{(a_1,\dots,a_n),(b_1,\dots,b_n)} = \{(t(a_1,\dots,a_n), t(b_1,\dots,b_n)) \mid t \in \operatorname{Pol}\mathbb{A}, t \text{ is } n\text{-ary}\}.$$

We summarize the properties of $\mathcal{P}$ in a lemma.

Lemma 5.2. *The following holds:*

(i) $\mathcal{P}$ is an instance of the CSP over $\mathbf{alg}\,\mathbb{A}$.

(ii) $\mathcal{P}$ is a $(2,3)$-system with unary projections

$$R_{(a_1,\dots,a_n)} = \operatorname{Sg}_{\mathbf{alg}\,\mathbb{A}}(\{a_1,\dots,a_n\}).$$

(iii) Solutions to $\mathcal{P}$ are precisely n-ary polymorphisms of $\mathbb{A}$.

Proof. (i) and (ii) can be verified easily, for (iii) see [1, Proposition 4.5]. □

We want to isolate those solutions of $\mathcal{P}$, which witness absorption $B \trianglelefteq \mathbb{A}$. In order to do that, we restrict $\mathcal{P}$ to a family $\mathcal{S} = \{S_x \mid x \in A^n\}$ of subsets of A defined as follows:

$$S_{(a_1,\dots,a_n)} = \begin{cases} B \cap R_{(a_1,\dots,a_n)}, & \text{if } |\{i \mid a_i \notin B\}| \leq 1, \\ R_{(a_1,\dots,a_n)}, & \text{else.} \end{cases}$$

Lemma 5.3. *The following holds:*

(i) For every $x \in A^n$, $S_x \neq \emptyset$ and $S_x \trianglelefteq_{\mathcal{J}} R_x$.

(ii) Solutions to $\mathcal{P}|_{\mathcal{S}}$ are precisely n-ary polymorphisms of $\mathbb{A}$ witnessing $B \trianglelefteq \mathbb{A}$.

(iii) Every $\mathcal{P}$-tree with at most $n-1$ vertices is realizable in $\mathcal{P}|_{\mathcal{S}}$.

Proof. (i) is easily seen and (ii) follows from the construction of $\mathcal{S}$. To prove (iii), fix a $\mathcal{P}$-tree T. Note that for every $(a_1,\dots,a_n)$ there exists at most one i such that $a_i \notin S_{(a_1,\dots,a_n)}$. As T has less than n vertices, there exists i_0 such that $a_{i_0} \in S_{(a_1,\dots,a_n)}$ whenever $(a_1,\dots,a_n)$ is the label of a vertex of T. The projection to the i_0th coordinate (i.e., the mapping $\pi_{i_0}\colon A^n \to A$ defined by $\pi_{i_0}(a_1,\dots,a_n) = a_{i_0}$, which is a polymorphism of $\mathbb{A}$) provides a realization of T in $\mathcal{P}|_{\mathcal{S}}$. $\square$

Finally, we are ready to apply Theorem 4.3 and conclude that the instance $\mathcal{P}|_{\mathcal{S}}$ has a solution s, which is a $(4^{8^{|A|}}+1)$-ary polymorphism witnessing $B \trianglelefteq \mathbb{A}$.

6 Reduction to digraphs

This section presents a way to reduce a variety of questions about absorption in finite relational structures to digraphs. In [9] the authors introduced a construction devised to show that every `CSP` over a relational structure $\mathbb{A}$ is `LOGSPACE` equivalent to the `CSP` over a digraph $\mathcal{D}(\mathbb{A})$ and, moreover, the algebra of polymorphisms of the digraph $\mathcal{D}(\mathbb{A})$ shares many interesting equational properties (i.e., Maltsev conditions) with the algebra of polymorphisms of $\mathbb{A}$.

We observe that, interestingly, this construction behaves nicely with respect to absorption theory as well. Namely, it preserves the property of being an ($\mathcal{J}$-)absorbing subuniverse as well as the arity of a polymorphism (length of a Jónsson chain) witnessing the ($\mathcal{J}$-)absorption. Moreover, in a sense, it does not create any new absorption-free subalgebras. Below we present the construction in a concise form. For a more thorough description we refer the reader to [9, Section 4].

6.1 Preliminaries

An oriented path is a digraph obtained from an undirected path by giving each edge an orientation. We require that oriented paths have direction and hence an *initial* and a *terminal* vertex. By a *zigzag* and a *single edge* we mean the oriented paths $\bullet \to \bullet \leftarrow \bullet \to \bullet$ and $\bullet \to \bullet$, respectively. Let us denote by $\dotplus$ the concatenation of oriented paths (identifying the terminal vertex of the first one with the initial vertex of the second one).

Let $\mathbb{G}$ be a digraph and $a, b \in G$. We say that a and b are *connected*, if there exists an oriented path $\mathbb{P}$ and a homomorphism $\varphi : \mathbb{P} \to \mathbb{G}$ mapping the initial and terminal vertex of $\mathbb{P}$ to a and b, respectively. Connectedness is an equivalence relation on G; its classes are the *connected components* of $\mathbb{G}$ and $\mathbb{G}$ is *connected* if it consists of a single connected component.

A connected digraph is *balanced* if it admits a *level function* L, where $\mathrm{L}(b) = \mathrm{L}(a) + 1$ whenever (a, b) is an edge and the minimum level is 0. The maximum level is called the *height* of the digraph.

Given a digraph $\mathbb{G} = (G; E)$ and $n > 0$, the *nth direct power of* $\mathbb{G}$ is the digraph $\mathbb{G}^n = (G^n; E^n)$, i.e., $(\bar{a}, \bar{b})$ is an edge in $\mathbb{G}^n$ iff $a_i \to b_i$ for all $i \leq n$. (The *direct product* of possibly distinct digraphs is defined similarly.)

6.2 The construction

Let $\mathbb{A} = (A; R_1, \dots, R_n)$ be a finite relational structure. Denote by k the sum of arities of the relations R_i and by R the k-ary relation $R_1 \times \cdots \times R_n$. (As $\mathbf{alg}\,\mathbb{A} = \mathbf{alg}(A; R)$, we could without loss of generality assume that $\mathbb{A}$ has just one relation.) We make a technical assumption that $R \neq \emptyset$.

For every $e \in A \times R$, we define an oriented path $\mathbb{P}_e$ (of height $k+2$). Say that $e = (a, (a_1, \dots, a_k))$, then

$$\mathbb{P}_e = \bullet \to \bullet \dotplus \mathbb{P}_{e,1} \dotplus \mathbb{P}_{e,2} \dotplus \dots \dotplus \mathbb{P}_{e,k} \dotplus \bullet \to \bullet$$

where $\mathbb{P}_{e,i}$ is a single edge if $a = a_i$, and a zigzag else. Finally, let $\mathcal{D}(\mathbb{A})$ be the digraph obtained from $(A \cup R; A \times R)$ by replacing every $e = (a, r) \in A \times R$ by the oriented path $\mathbb{P}_e$ (identifying the initial and terminal vertices of $\mathbb{P}_e$ with a and r, respectively).

The digraph $\mathcal{D}(\mathbb{A})$ is balanced of height $2k + 1$. For computational questions it is useful to note that the size of $\mathcal{D}(\mathbb{A})$ is roughly $k|A|^k$. Precise bounds can be found in [9, Remark 1].

6.3 Results

It is straightforward to prove that $\mathbb{A}$ is pp-definable from $\mathcal{D}(\mathbb{A})$ (see [9, Lemma 3]). It follows that A and R are subuniverses of $\mathcal{D}(\mathbb{A})$ and for any $f \in \mathrm{Pol}\,\mathcal{D}(\mathbb{A})$, the restriction $f|_A$ is a polymorphism of $\mathbb{A}$. Consequently, for any idempotent Maltsev condition Σ, $\mathbf{alg}\,\mathcal{D}(\mathbb{A}) \models \Sigma$ implies that $\mathbf{alg}\,\mathbb{A} \models \Sigma$.

The following result, which will appear in [9], provides a partial converse to this implication.

Theorem 6.1. *Let $\mathbb{A}$ be a finite relational structure. Let Σ be a linear idempotent Maltsev condition such that the algebra of polymorphisms of the zigzag satisfies Σ and each identity in Σ is either balanced or contains at most two variables. Then* $\mathbf{alg}\,\mathbb{A} \models \Sigma$ *if and only if* $\mathbf{alg}\,\mathcal{D}(\mathbb{A}) \models \Sigma$.

The above condition on Σ includes most of the Maltsev conditions commonly encountered in the algebraic approach to the `CSP` (some of them are listed in [9]) with the one important exception being having a Maltsev term.

The following lemma demonstrates that the reduction to digraphs indeed works well with the notion of absorption.

Lemma 6.2. *Let $\mathbb{A}$ be a finite relational structure and $B \subseteq \mathbb{A}$. Then the following holds.*

(i) $B \trianglelefteq \mathbb{A}$ via an m-ary polymorphism iff $B \trianglelefteq \mathcal{D}(\mathbb{A})$ via an m-ary polymorphism.

(ii) $B \trianglelefteq_{\mathcal{J}} \mathbb{A}$ via a chain of length m iff $B \trianglelefteq_{\mathcal{J}} \mathcal{D}(\mathbb{A})$ via a chain of length m.

Proof. To prove (i), assume that $B \trianglelefteq \mathbb{A}$ via an m-ary $t \in \operatorname{Pol}\mathbb{A}$. As t is an operation on A, it also acts coordinatewise on R. For every $\bar{a} \in A^m$ and $\bar{r} \in R^m$ there exists a homomorphism from the connected component of $\prod_{i=1}^m \mathbb{P}_{(a_i,r_i)}$ containing $\bar{a}$ to $\mathbb{P}_{(t(\bar{a}),t(\bar{r}))}$ (see [9] for the technical details). Let us fix such a homomorphism and call it $\Phi_{\bar{a},\bar{r}}$.

We define an m-ary operation t' on $\mathcal{D}(\mathbb{A})$. Let $\bar{x} \in \mathcal{D}(\mathbb{A})^m$ be arbitrary.

1. If $\{x_1, \ldots, x_m\} \subseteq A$ or $\{x_1, \ldots, x_m\} \subseteq R$, we put $t'(\bar{x}) = t(\bar{x})$. (Note that t acts coordinatewise on R.)

2. If $\mathrm{L}(x_1) = \ldots \mathrm{L}(x_m) = l$ for some $l \notin \{0, k+2\}$, then let a_i, r_i be such that $x_i \in \mathbb{P}_{(a_i,r_i)}$. We define $t'(\bar{x}) = \Phi_{\bar{a},\bar{r}}(\bar{x})$ if $\bar{x}$ and $\bar{a}$ are connected in $\mathcal{D}(\mathbb{A})^m$, and $t'(\bar{x}) = x_1$ else.

3. If there exists $i \in [m]$ and $l > l'$ such that $\mathrm{L}(x_i) = l$ and $\mathrm{L}(x_j) = l'$ for all $j \neq i$, then $t'(\bar{x}) = x_2$ if $i = 1$ and $t'(\bar{x}) = x_1$ else.

4. In all other cases we define $t'(\bar{x}) = x_1$.

The absorption condition is obviously satisfied. To verify that $t' \in \operatorname{Pol}\mathcal{D}(\mathbb{A})$, note that if $x_i \to y_i$ for $i \in [m]$, then $\bar{x}$ and $\bar{y}$ fall under the same case of the definition of t'. The rest is easy. We omit the proof of (ii), as it is similar yet more technical. One can adapt the proof of Theorem 6.1 to obtain both (i) and (ii). $\square$

The next lemma describes absorption free subalgebras of $\mathcal{D}(\mathbb{A})$.

Lemma 6.3. *Let $C \leq \mathcal{D}(\mathbb{A})$ be absorption free. Then there exists an absorption free algebra $B \in \mathrm{SP}_{\mathrm{fin}}(\mathbf{alg}\,\mathbb{A})$ and a congruence $\alpha \in \operatorname{Con} C$ such that $B \simeq C/\alpha$. Moreover, classes of α are at most two-element and possess a semilattice operation.*

Proof. Let us define the following binary operation $\star$ on $\mathcal{D}(\mathbb{A})$.

1. If $\mathrm{L}(x) > \mathrm{L}(y)$, then we define $x \star y = y$.

2. If $\mathrm{L}(x) = \mathrm{L}(y)$, $x, y \in \mathbb{P}_e$ for some $e \in A \times R$ and y is closer to the initial vertex of $\mathbb{P}_e$ than x, then we define $x \star y = y$.

3. In all other cases we put $x \star y = x$.

It is not hard to verify that $\star$ is a polymorphism of $\mathcal{D}(\mathbb{A})$ and that $A \trianglelefteq \mathcal{D}(\mathbb{A})$ via $\star$; and so $C \neq \mathcal{D}(\mathbb{A})$.

We will prove that all the vertices from C have the same level. Let D denote the set of elements of C with minimum level, i.e., $D = \{d \in C \mid \mathrm{L}(d) \leq$

$\mathrm{L}(c)$ for all $c \in C\}$. Below we show that D is pp-definable from $\mathcal{D}(\mathbb{A})$, and thus $D \leq C$. Since $D \trianglelefteq C$ via $\star$ and C is absorption free, it then follows that $D = C$.

Let $D = \{d_1, \ldots, d_m\}$ be an enumeration of all vertices from D. The tuple $\bar{d} = (d_1, \ldots, d_m)$ cannot form a singleton connected component of $\mathcal{D}(\mathbb{A})^m$. (Otherwise we could define for every $v \in \mathcal{D}(\mathbb{A})$ an m-ary polymorphism t_v by setting $t_v(\bar{d}) = v$ and $t_v(\bar{x}) = x_1$ for all other connected components; and so the subuniverse generated by D would be the whole $\mathcal{D}(\mathbb{A})$.) It follows that there exist $\bar{a} \in A^m$ and an oriented path $\mathbb{P}$ connecting $\bar{a}$ to $\bar{d}$ in $\mathcal{D}(\mathbb{A})^m$ (see [9], Lemma 6). Let us denote the initial and terminal vertex of $\mathbb{P}$ by p and q, respectively. The set D can be expressed in the following way:

$$D = C \cap \{\varphi(q) \mid \varphi : \mathbb{P} \to \mathcal{D}(\mathbb{A}) \text{ is a homomorphism and } \varphi(p) \in A\}.$$

The right hand side is definable by a primitive positive formula in $\mathcal{D}(\mathbb{A})$, which concludes the proof that D is a subuniverse.

We may assume that neither $C \leq A$ nor $C \leq R$, otherwise the proof is trivial. For every $c \in C$ let $a_c \in A$ and $r_c \in R$ be such that $c \in \mathbb{P}_{(a_c, r_c)}$. Define $A_C = \{a_c \mid c \in C\}$, $R_C = \{r_c \mid c \in C\}$, $B = A_C \times R_C$ and $\alpha = \{(c, c') \in C^2 \mid (a_c, r_c) = (a_{c'}, r_{c'})\}$.

Using a similar argument as before, it is not hard to verify that $A_C \leq A$ and $R_C \leq R$; therefore $B \in \mathrm{SP}_{\mathrm{fin}}(\mathbf{alg}\,\mathbb{A})$. The equivalence relation α is pp-definable (the defining formula is either $(x, y \in C)\&(\exists z)(x \to z \leftarrow y)$ or $(x, y \in C)\&(\exists z)(x \leftarrow z \to y)$); and so $\alpha \in \operatorname{Con} C$. Every α-class K is a subset of $C \cap \mathbb{P}_e$ for some $e \in A \times R$, so K is at most two element, and $\star|_K$ is a semilattice operation.

The mapping $\varphi \colon C \to B$ defined by $\varphi(c) = (a_c, r_c)$ is a surjective homomorphism whose kernel is α. (This is an easy exercise.) Thus, by the First Isomorphism Theorem, $C/\alpha \simeq B$. To conclude the proof, note that since C is absorption free, then so is $C/\alpha \simeq B$, that is, unless C/α is one element. However, in that case $\star$ is a semilattice operation on C and thus C cannot be absorption free; a contradiction. $\square$

7 Always absorbing algebras

We conclude the paper by presenting an open problem in absorption theory that we find particularly interesting: the problem of *always absorbing algebras.*

Definition 7.1. Let $\mathbf{A}$ be a finite idempotent algebra. We say that $\mathbf{A}$ is *always absorbing (AA)* if for every $B \leq \mathbf{A}$ there exists $b \in B$ such that $\{b\} \trianglelefteq B$.

The property of being always absorbing is inherited by all finite algebras in the variety generated by $\mathbf{A}$. The following lemma gives an equivalent definition of always absorbing algebras. (The proof is elementary; all the facts about absorption that are needed can be found in [6].)

Lemma 7.2. *A finite idempotent algebra* $\mathbf{A}$ *is always absorbing, if and only if there exist no absorption free algebras in* $\mathrm{HSP}_{\mathrm{fin}}(\mathbf{A})$.

Barto and Kozik recently discovered a new characterization of finite idempotent $\mathrm{SD}(\wedge)$ algebras. An algebra $\mathbf{A}$ has a *pointed term* if there exists $t \in \operatorname{Clo} \mathbf{A}$

(say n-ary), $(a_1, \ldots, a_n) \in A^n$ and $c \in A$ such that

$$t(b_1, \ldots, b_n) = c \text{ whenever } |\{i \mid a_i \neq b_i\}| \leq 1.$$

Theorem 7.3 (to appear in [6])**.** *A finite idempotent algebra* $\mathbf{A}$ *is* SD($\wedge$)*, if and only if every* $B \leq \mathbf{A}$ *has a pointed term.*

It is easily seen that if an algebra has a singleton absorbing subuniverse, then the term witnessing this absorption is a pointed term, which implies the following:

Corollary 7.4. *Let* $\mathbf{A}$ *be a finite idempotent algebra. If* $\mathbf{A}$ *is AA, then it is* SD($\wedge$).

Not every finite idempotent SD($\wedge$) algebra is AA (for example, the "rock, paper, scissors" 2-semilattice on $\{0, 1, 2\}$ given by $0 < 1$, $1 < 2$ and $2 < 0$ is not). However, the class of AA algebras includes two important classes of SD($\wedge$) algebras: algebras with a near unanimity term and algebras with a semilattice term. Moreover, several theorems, proofs and algorithms around the algebraic approach to the `CSP` (a prototypical example being the Bounded width theorem [5]) get significantly simpler when restricted to AA algebras. Hence the following problem:

Problem 4. Characterize AA algebras, at least in the finitely related case.

Lemma 6.3 shows that the class of finitely related AA algebras is determined by digraphs.

Acknowledgements

The author would like to thank Libor Barto and Marcin Kozik for their thoughtful comments and helpful discussions and the anonymous reviewer for pointing out an error in the manuscript. This research was supported by the Grant Agency of the Czech Republic project no. 201/09/H012, Charles University Grant Agency projects no. GA UK 67410, SVV-2012-265317 and the Fields Institute for Research in Mathematical Sciences.

References

[1] Barto, L.: Finitely related algebras in congruence distributive varieties have near unanimity terms. Canadian Journal of Mathematics **65**/1, 3–21 (2013)

[2] Barto, L.: The dichotomy for conservative constraint satisfaction problems revisited. In: Proceedings of the 26th IEEE Symposium on Logic in Computer Science, LICS'11, 301–310 (2011)

[3] Barto, L., Kozik, M.: Absorbing subalgebras, cyclic terms and the constraint satisfaction problem. Logical Methods in Computer Science **8**/1:07, 1–26 (2012)

[4] Barto, L., Kozik, M.: Congruence distributivity implies bounded width. SIAM Journal on Computing **39**/4, 1531–1542 (2009)

[5] Barto, L., Kozik, M.: Constraint satisfaction problems of bounded width. In: Proceedings of the 50th Annual IEEE Symposium on Foundations of Computer Science, FOCS'09, 595–603 (2009)

[6] Barto, L., Kozik, M., Stanovský, D.: Mal'tsev conditions, lack of absorption, and solvability (2013, preprint) `http://www.karlin.mff.cuni.cz/~stanovsk/math/maltsev_absorption_solvability.pdf`

[7] Bulatov, A.: Tractable conservative constraint satisfaction problems. In: Proceedings of the 18th IEEE Symposium on Logic in Computer Science, LICS'03, 321–330 (2003)

[8] Bulatov, A., Jeavons, P., Krokhin, A.: Classifying the complexity of constraints using finite algebras. SIAM Journal on Computing **34**, 720–742 (2005)

[9] Bulín, J., Delić, D., Jackson, M., Niven, T.: On the reduction of the CSP dichotomy conjecture to digraphs. In: Proceedings of the 19th International Conference on Principles and Practice of Constraint Programming (CP 2013). Lecture Notes in Computer Science **8124**, 184–199 (2013)

[10] Feder, T., Vardi, M.Y.: The computational structure of monotone monadic SNP and constraint satisfaction: a study through Datalog and group theory. SIAM Journal on Computing **28**, 57–104 (1999)

[11] Larose, B., Zádori, L.: Bounded width problems and algebras. Algebra Universalis **56**(3–4), 439–466 (2007)

[12] Maróti, M.: On the (un)decidability of a near-unanimity term. Algebra Universalis **57**(2), 215-–237 (2007)

[13] Willard, R.: Testing expressibility is hard. In: Proceedings of the 16th International Conference on Principles and Practice of Constraint Programming (CP 2010). Lecture Notes in Computer Science **6308**, 9–23 (2010)

[14] Open Problems from the Workshop on Algebra and CSPs. Fields Institute for Research in Mathematical Sciences, Toronto (2011) `http://www.karlin.mff.cuni.cz/~barto/Articles/fields-2011-problems.pdf`

Part III

Oriented trees

On the complexity of $\mathbb{H}$-coloring for special oriented trees

Jakub Bulín

Abstract

For a fixed digraph $\mathbb{H}$, the $\mathbb{H}$-coloring problem is the problem of deciding whether a given input digraph $\mathbb{G}$ admits a homomorphism to $\mathbb{H}$. The CSP dichotomy conjecture of Feder and Vardi is equivalent to proving that, for any $\mathbb{H}$, the $\mathbb{H}$-coloring problem is in in **P** or **NP**-complete. We confirm this dichotomy for a certain class of oriented trees, which we call special trees (generalizing earlier results on special triads and polyads). Moreover, we prove that every tractable special oriented tree has bounded width, i.e., the corresponding $\mathbb{H}$-coloring problem is solvable by local consistency checking. Our proof relies on recent algebraic tools, namely characterization of congruence meet-semidistributivity via pointing operations and absorption theory.

Introduction

The Constraint Satisfaction Problem (CSP) provides a common framework for various problems from theoretical computer science as well as for many real-life applications (e.g. in graph theory, database theory, artificial intelligence, scheduling). Its history dates back to 1970s and it has been central to the development of theoretical computer science in the past few decades.

For a fixed (finite) relational structure $\mathbb{A}$, the Constraint Satisfaction Problem with template $\mathbb{A}$, or CSP($\mathbb{A}$) for short, is the following decision problem:

INPUT: A relational structure $\mathbb{X}$ (of the same type as $\mathbb{A}$).
QUESTION: Is there a homomorphism from $\mathbb{X}$ to $\mathbb{A}$?

For a (directed) graph $\mathbb{H}$, CSP($\mathbb{H}$) is also commonly referred to as the $\mathbb{H}$-coloring problem.

A lot of interest in this class of problems was sparked by a seminal work of Feder and Vardi [20], in which the authors established a connection to computational complexity theory: they conjectured a large natural class of **NP** decision problems avoiding the complexity classes strictly between **P** and **NP**-complete (assuming that **P**$\neq$**NP**). Many natural decision problems, such as k-SAT, graph k-colorability or solving systems of linear equations over finite fields belong to this class. They also proved that each problem from this class can be reduced in polynomial time to CSP($\mathbb{A}$), for some relational structure $\mathbb{A}$. Hence their conjecture can be formulated as follows.

Conjecture 1 (The CSP dichotomy conjecture)**.** *For every (finite) relational structure* $\mathbb{A}$*,* CSP($\mathbb{A}$) *is in* $\boldsymbol{P}$ *or* $\boldsymbol{NP}$*-complete.*

At that time this conjecture was supported by two major cases: Schaefer's dichotomy result for two-element domains [33] and the dichotomy theorem for undirected graphs by Hell and Nešetřil [26]. A major breakthrough followed the work of Jeavons, Cohen and Gyssens [29], later refined by Bulatov, Jeavons and Krokhin [15], which uncovered an intimate connection between the constraint satisfaction problem and universal algebra. This connection brought a better understanding of the known results as well as a number of new results which seemed out of reach for pre-algebraic methods. The most important results include dichotomy for three-element domains [14] and for conservative structures (i.e., containing all subsets as unary relations) [13] by Bulatov (see also [2]), a characterization of solvability by the *few subpowers* algorithm (a generalization of Gaussian elimination) by Berman et al [11, 28] and solvability by local consistency checking (so-called *bounded width*) by Barto and Kozik [6] (conjectured in [31]). Larose and Tesson [30] successfully applied the theory to study finer complexity classes of CSPs.

The connection between CSPs and algebras turned out to be fruitful in both directions; it has lead to a discovery of important structural properties of finite algebras. Of particular importance to us is the theory of absorption by Barto and Kozik [4, 9] and a characterization of congruence meet-semiditributivity via pointing operations by Barto, Kozik and Stanovský [5, 9].

In the paper [20], Feder and Vardi also constructed, for every structure $\mathbb{A}$, a directed graph $\mathcal{D}(\mathbb{A})$ such that CSP($\mathbb{A}$) and CSP($\mathcal{D}(\mathbb{A})$) are polynomial-time equivalent. Hence the CSP dichotomy conjecture is equivalent to its restriction to digraphs. A variant of this reduction (which is, in fact, logspace) is studied by the author, Delić, Jackson and Niven in [17, 18], where we prove that most properties relevant to the CSP carry over from $\mathbb{A}$ to $\mathcal{D}(\mathbb{A})$. As a consequence, the algebraic conjectures characterizing CSPs solvable in **P** [15], **NL** and **L** [30] are equivalent to their restrictions to digraphs. The digraphs $\mathcal{D}(\mathbb{A})$ are, in fact, *special balanced digraphs* in the terminology of this paper, a generalization of special triads, special polyads and special trees discussed below.

Using the algebraic approach, Barto, Kozik and Niven confirmed the conjecture of Bang-Jensen and Hell and proved dichotomy for *smooth digraphs* (i.e., digraphs with no sources and no sinks) [8]. The dichotomy was also established for a number of other classes of digraphs, e.g. oriented paths (which are all tractable) [22] or oriented cycles [19].

This paper is concerned with $\mathbb{H}$-coloring for oriented trees. In the class of all digraphs, oriented trees are in some sense very far from smooth digraphs, and the algebraic tools seem to be not yet developed enough to deal with them. Hence oriented trees serve as a good field-test for new methods.

Except the oriented paths, the simplest class of oriented trees are *triads* (i.e., oriented trees with one vertex of degree 3 and all other vertices of degree 2 or 1); the CSP dichotomy remains open even for triads. Among the triads, Hell, Nešetřil and Zhu [23, 24] identified a (fairly restricted) subclass, for which they coined the term *special triads* and which allowed them to handle at least some examples. For instance, they constructed a special triad with **NP**-complete $\mathbb{H}$-coloring.

In [7], Barto et al used algebraic methods to prove that every special triad has **NP**-complete $\mathbb{H}$-coloring, or a compatible majority operation (so-called *strict width 2*) or compatible totally symmetric idempotent operations of all arities (so-called *width 1*). In [3], the author and Barto established the CSP dichotomy conjecture for *special polyads*, a generalization of special triads where the one vertex of degree > 2 is allowed to have an arbitrary degree. In particular, every tractable core special polyad has bounded width. However, there are special polyads which have bounded width, but neither bounded strict width nor width 1.

In this paper we study *special trees*, a broad generalization of special triads and special polyads. Special trees have an underlying structure of a height 1 oriented tree (see the definition in Section 1) and while for special triads it has only 7 vertices and for special polyads it has radius 2, for general special trees it can be arbitrary.

We confirm the CSP dichotomy conjecture for special trees and, moreover, prove that every tractable core special tree has bounded width. The proof uses modern tools from the algebraic approach to the CSP (in particular, absorption and pointing operations [9]) and is somewhat simpler and more natural than the proofs in [7] and [3]. Therefore we believe that there is hope for further generalization. In particular, we conjecture that tractability implies bounded width for all oriented trees.

1 Special trees & the main result

In this section we define special trees and state the main result of this paper. The notions used here will be defined later, in Sections 2 and 3.

Definition 1.1. An oriented path $\mathbb{P}$ with initial vertex a and terminal vertex b is *minimal* if

- $\mathrm{lvl}(a) = 0$,
- $\mathrm{lvl}(b) = \mathrm{hgt}(\mathbb{P})$, and
- $0 < \mathrm{lvl}(v) < \mathrm{hgt}(\mathbb{P})$ for every $v \in P \setminus \{a, b\}$.

Minimal paths have the property that their *net length* (the number of forward edges minus the number of backward edges) is strictly greater than the net length of any of their subpaths; hence the name. An example of a minimal path is depicted in Figure 1 below.

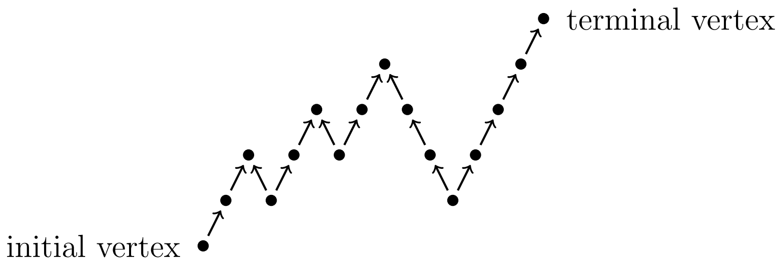


Figure 1: A minimal path

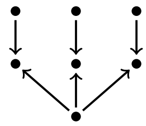
We will need the following well known fact. The proof can be found in [27].

Lemma 1.2. *Let $\mathbb{P}_1, \mathbb{P}_2, \dots \mathbb{P}_k$ be minimal paths of the same height h. There exists a minimal path $\mathbb{Q}$ of height h such that for every $i \in [k]$ there exists an onto homomorphism $\mathbb{Q} \to \mathbb{P}_i$.*

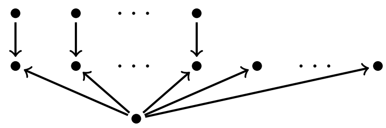
We are now ready to define *special trees.*

Definition 1.3. Let $\mathbb{T} = (T; E)$ be an oriented tree of height 1. A $\mathbb{T}$-*special tree* of height h is an oriented tree obtained from $\mathbb{T}$ by replacing every edge $(a, b) \in E$ with some minimal path $\mathbb{P}_{(a,b)}$ of height h, preserving orientation. (That is, identifying the initial vertex of $\mathbb{P}_{(a,b)}$ with a and the terminal vertex with b. We require the vertex sets of the minimal paths to be pairwise disjoint and also disjoint with T.)

- A *special triad* (as defined in [7]) is a $\mathbb{T}$-special tree with

  $\mathbb{T} =$ 

- A *special polyad* (as defined in [3]) is a $\mathbb{T}$-special tree with

  $\mathbb{T} =$ 

- A *special tree* is simply a $\mathbb{T}$-special tree for some height 1 oriented tree $\mathbb{T}$.

As an example, in Figure 2 below we present a special triad constructed in [7], which has **NP**-complete $\mathbb{H}$-coloring (and is conjectured to be the smallest oriented tree with this property). The vertices from the bottom and top level are marked by $\blacksquare$ and $\square$, respectively.

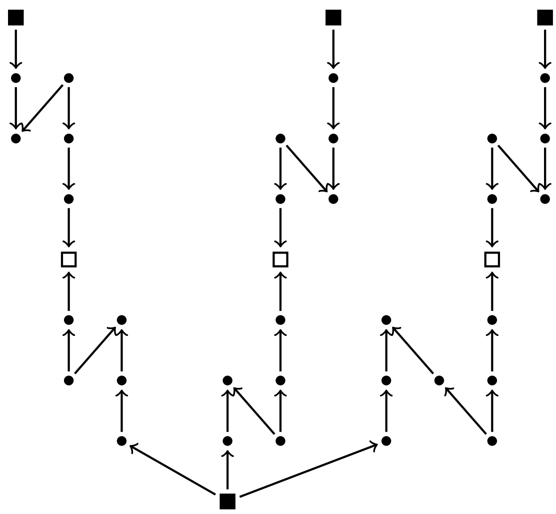


Figure 2: A special triad; the smallest known oriented tree with **NP**-complete $\mathbb{H}$-coloring problem (39 vertices).

The following theorem is the main algebraic result of our paper.

Theorem 1.4. *Let $\mathbb{H}$ be a special tree. If the algebra of idempotent polymorphisms of $\mathbb{H}$ is Taylor, then it is congruence meet-semidistributive.*

As a consequence, we confirm the dichotomy of $\mathbb{H}$-coloring for special trees.

Corollary 1.5. *The CSP dichotomy conjecture holds for special trees. For any core special tree $\mathbb{H}$, $\mathrm{CSP}(\mathbb{H})$ is **NP**-complete or $\mathbb{H}$ has bounded width.*

We will prove Theorem 1.4 and Corollary 1.5 in Section 4.

2 Preliminaries

In this section we introduce basic notions and fix notation used throughout the paper. We assume the reader possesses some knowledge of graph theory and basic universal algebra.

We recommend [25] for a detailed exposition of digraphs, relational structures (under the name "general relational systems") and their homomorphisms as well as an introduction to graph coloring and constraint satisfaction. For an introduction to the notions from universal algebra that are not explained in detail in this paper we invite the reader to consult [10]. Primary source for the algebraic approach to the CSP is the paper [15].

Our aim is to make the paper accessible to a wider audience outside of universal algebra. Thus we refrain from using specialist terminology wherever possible, or move it to explanatory remarks which the reader may skip.

2.1 Notation

For a positive integer n we denote the set $\{1, 2, \dots, n\}$ by $[n]$; we set $[0] = \emptyset$. We write tuples using boldface notation, e.g., $\mathbf{a} = (a_1, a_2, \dots, a_k) \in A^k$. When ranging over tuples we use superscripts, e.g. $(\mathbf{a}^1, \mathbf{a}^2, \dots, \mathbf{a}^n) \in (A^k)^n$, where $\mathbf{a}^i = (a^i_1, a^i_2, \dots, a^i_k)$, for $i \in [n]$. We sometimes write $\langle a_1 a_2 \dots \rangle$ to denote a sequence of elements.

2.2 Relational structures

An *n-ary relation* on a set A is a subset $R \subseteq A^n$. A (finite) *relational structure* $\mathbb{A}$ is a finite, nonempty set A equipped with finitely many relations $R_1 \dots R_m$ on A; we write $\mathbb{A} = (A; R_1, \dots, R_m)$.

Let $\mathbb{B} = (B; S_1, \dots, S_m)$ be a relational structure of the same *type* as $\mathbb{A}$ (i.e., same number of relations and corresponding relations have the same arity). A mapping $\varphi : A \to B$ is a *homomorphism* from $\mathbb{A}$ to $\mathbb{B}$, if for each $i \in [m]$ and $\mathbf{a} \in R_i$ (say k-ary) we have $(\varphi(a_1), \dots, \varphi(a_k)) \in S_i$. We write $\varphi : \mathbb{A} \to \mathbb{B}$ to mean that φ is a homomorphism from $\mathbb{A}$ to $\mathbb{B}$, and $\mathbb{A} \to \mathbb{B}$ to mean that there exists a homomorphism from $\mathbb{A}$ to $\mathbb{B}$.

For every $\mathbb{A}$ there exists a relational structure $\mathbb{A}'$ such that $\mathbb{A} \to \mathbb{A}'$ and $\mathbb{A}' \to \mathbb{A}$ and $\mathbb{A}'$ is of minimal size with respect to these properties; that structure $\mathbb{A}'$ is called the *core of* $\mathbb{A}$ (it is unique up to isomorphism); $\mathbb{A}$ is a *core* if it is the core of itself.

We will be almost exclusively interested in a special type of relational structures: *directed graphs.*

2.3 Digraphs

A *digraph* (short for "*directed graph*") is a relational structure $\mathbb{G} = (G; \to)$ with a single binary relation $\to \subseteq G^2$. We call $u \in G$ and $(u, v) \in \to$ (usually written as $u \to v$) *vertices* and *edges* of $\mathbb{G}$, respectively. A digraph $\mathbb{G}' = (G'; \to')$ is a *subgraph* of $\mathbb{G}$, if $G' \subseteq G$ and $\to' \subseteq \to$. It is an *induced subgraph* if $\to' = \to \cap (G')^2$.

An *oriented path* is a digraph $\mathbb{P}$ which consists of a non-repeating sequence of vertices $\langle v_0 v_1 \dots v_k \rangle$ (allowing for the degenerate case $k = 0$) such that precisely one of $(v_{i-1}, v_i), (v_i, v_{i-1})$ is an edge, for each $i \in [k]$. We require oriented paths to have a fixed direction, and thus an *initial* and a *terminal* vertex.

For $a, b \in G$ we say that a is *connected* to b in $\mathbb{G}$ via an oriented path $\mathbb{P}$, if $\mathbb{P}$ is a subgraph of $\mathbb{G}$ and a and b are the initial and terminal vertex of $\mathbb{P}$, respectively. The *distance* of a and b in $\mathbb{G}$ is then the number of edges in the shortest oriented path $\mathbb{P}'$ connecting a to b in $\mathbb{G}$. Connectivity is an equivalence relation, its classes are *components of connectivity* of $\mathbb{G}$ and $\mathbb{G}$ is *connected* if it consists of a single component of connectivity.

For $n > 0$, the *nth direct power of* $\mathbb{G}$ is the digraph $\mathbb{G}^n = (G^n, \to^n)$, i.e., its vertices are n-tuples of vertices of $\mathbb{G}$ and the edge relation is

$$\{(\mathbf{u}, \mathbf{v}) \in (G^n)^2 \mid u_i \to v_i \text{ for all } i \in [n]\}.$$

Connectivity in direct powers of digraphs will play an important role.

An *oriented tree* is a connected digraph containing no oriented cycles. Equivalently, it is a digraph in which every two vertices are connected via a unique oriented path. Oriented paths and trees are natural examples of balanced digraphs: a connected digraph is *balanced* if it admits a *level function* $\mathrm{lvl} : G \to \mathbb{N} \cup \{0\}$, where $\mathrm{lvl}(b) = \mathrm{lvl}(a) + 1$ whenever (a, b) is an edge, and the minimum level is 0. The maximum level is called *height* and denoted by $\mathrm{hgt}(\mathbb{G})$.

2.4 Algebras

A *k-ary operation* on a set A is a mapping $f \colon A^k \to A$. By an *algebra* we mean a pair $\mathbf{A} = (A; \mathcal{F})$, where A is a nonempty set and $\mathcal{F}$ is a set of operations on A (so-called *basic operations* of $\mathbf{A}$). We denote by $\mathrm{Clo}(\mathbf{A})$ the set of all *term operations* of $\mathbf{A}$ (i.e., operations obtained from $\mathcal{F}$ together with the *projection* operations by composition).

A subset $B \subseteq A$ is a *subuniverse* of $\mathbf{A}$ (denoted by $B \leq \mathbf{A}$) if it is closed under all (basic, or equivalently term) operations of $\mathbf{A}$. A nonempty subuniverse B is an algebra in its own right, equipped with operations of $\mathbf{A}$ restricted to B, i.e., $(B; \{f|_B \mid f \in \mathcal{F}\})$. We will frequently use the fact that an intersection of subuniverses is again a subuniverse.

An operation is *idempotent* if $f(x, x, \dots, x) = x$ for all $x \in A$. An algebra is idempotent if all of its (basic, or equivalently term) operations are idempotent. Note that an algebra $\mathbf{A}$ is idempotent, if and only if $\{a\} \leq \mathbf{A}$ for every $a \in A$.

For $n > 0$, the *nth power* of $\mathbf{A}$ is the algebra $\mathbf{A}^n = (A^n; \{f \times \dots \times f \mid f \in \mathcal{F}\})$ where $f \times \dots \times f$ means that f is applied to n-tuples of elements coordinatewise.

We write $C \leq B \leq \mathbf{A}$ to mean that both B and C are subuniverses of $\mathbf{A}$ and $C \subseteq B$. In particular, if B and C are subuniverses of $\mathbf{A}$, then $E \leq B \times C$ means that E is a subuniverse of $\mathbf{A}^2$ contained in $B \times C$ (which is a subuniverse of $\mathbf{A}^2$ as well).

All algebras we will work with will be subuniverses of a certain finite idempotent algebra (or rarely of its 2nd power): the *algebra of idempotent polymorphisms* of some fixed relational structure.

2.5 Algebra of idempotent polymorphisms

Note that a digraph homomorphism is simply an edge-preserving mapping. The notion of digraph *polymorphism* is a natural generalization to higher arity operations:

Let $\mathbb{G} = (G; \to)$ be a digraph. A k-ary ($k > 0$) operation φ on G is a *polymorphism* of $\mathbb{G}$, if it is a homomorphism from $\mathbb{G}^k$ to $\mathbb{G}$. This means that φ preserves edges in the following sense: if $a_i \to b_i$ for $i \in [k]$, then $\varphi(\mathbf{a}) \to \varphi(\mathbf{b})$. The notions of kth direct power, preserving a relation, and polymorphism generalize naturally to relational structures.

Let $\mathbb{A}$ be a relational structure. The *algebra of idempotent polymorphisms* of $\mathbb{A}$ is the algebra $\mathbf{alg}\,\mathbb{A} = (A; \mathrm{IdPol}(\mathbb{A}))$, where $\mathrm{IdPol}(\mathbb{A})$ denotes the set of all idempotent polymorphisms of $\mathbb{A}$; we write $\mathrm{IdPol}_k(\mathbb{A})$ to denote its k-ary part.

A relation $S \subseteq A^n$ is *primitive positive definable* from $\mathbb{A}$ *with constants*, if it is definable by an existentially quantified conjunction of atomic formulæ of the form $x_i = a$ or $R(x_{i_1}, \ldots, x_{i_j})$, where $a \in A$ and R is one of the relations of $\mathbb{A}$. The following fact, based on the Galois correspondence between clones and relational clones [12, 21] is central to the algebraic approach to the CSP.

Lemma 2.1 (see [15, Proposition 2.21])**.** *A relation $S \subseteq A^n$ is primitive positive definable from $\mathbb{A}$ with constants, if and only if S is a subuniverse of* $(\mathbf{alg}\,\mathbb{A})^n$.

The connection between universal algebra and constraint satisfaction is discussed in detail in [15, 16].

3 Algebraic tools

In this section we introduce the universal algebraic tools we will use in our proof. Recall that for a fixed relational structure $\mathbb{A}$, the *Constraint satisfaction problem* for $\mathbb{A}$ is membership problem for the set $\mathrm{CSP}(\mathbb{A}) = \{\mathbb{X} \mid \mathbb{X} \to \mathbb{A}\}$. Note that if $\mathbb{A}'$ is the core of $\mathbb{A}$, then $\mathrm{CSP}(\mathbb{A}) = \mathrm{CSP}(\mathbb{A}')$.

Of particular importance to the CSP are the following two well known classes of finite algebras: *Taylor* algebras (called "active" in [10]) and *congruence meet-semidistributive* (*$SD(\wedge)$*) algebras[1]. Instead of providing direct definitions, we present the following characterization from [32].

Definition 3.1. A *weak near-unanimity* (*WNU*) on a set A is an n-ary ($n \geq 2$) idempotent operation ω such that for all $x, y \in A$,

$$\omega(x, \ldots, x, y) = \omega(x, \ldots, x, y, x) = \cdots = \omega(y, x, \ldots, x).$$

Theorem 3.2 ([32])**.** *Let* $\mathbf{A}$ *be a finite algebra.*

[1] Taylor and SD($\wedge$) algebras are also commonly referred to as "omitting type 1" and "omitting types 1, 2"; this terminology comes from Tame Congruence Theory (see [10, Chapter 8]).

- $\mathbf{A}$ *is Taylor, if and only if there exists a WNU operation* $\omega \in \mathrm{Clo}(\mathbf{A})$.
- $\mathbf{A}$ *is* SD($\wedge$), *if and only if there exists* n_0 *such that for all* $n \geq n_0$ *there exists an* n*-ary WNU operation* $\omega_n \in \mathrm{Clo}(\mathbf{A})$.

The Algebraic CSP dichotomy conjecture ([15], see also [16, Conjecture 1]) asserts that being Taylor is what distinguishes (algebras of idempotent polymorphisms of) tractable core relational structures from the **NP**-complete ones; the hardness part is known.

Theorem 3.3 ([15]). *Let* $\mathbb{A}$ *be a core relational structure. If* $\mathbf{alg}\,\mathbb{A}$ *is not Taylor, then* $\mathrm{CSP}(\mathbb{A})$ *is* ***NP****-complete.*

A relational structure $\mathbb{A}$ is said to have *bounded width* [20], if $\mathrm{CSP}(\mathbb{A})$ is solvable by "local consistency checking" algorithm (or rather algorithmic principle). We refer the reader to [6] for a detailed exposition. This property is characterized (for cores) by congruence meet-semidistributivity; the characterization was conjectured, and the "only if" part proved, in [31].

Theorem 3.4 ([6],"Bounded Width Theorem"). *A core relational structure* $\mathbb{A}$ *has bounded width (implying that* $\mathrm{CSP}(\mathbb{A})$ *is in* ***P****), if and only if* $\mathbf{alg}\,\mathbb{A}$ *is* SD($\wedge$).

The proof of the Bounded Width Theorem uncovered a new characterization of SD($\wedge$) algebras via so-called *pointing operations* as well as the concept of *absorbing subuniverse*, which turned out to be quite useful even outside of the realm of congruence meet-semidistributivity (see [4, 9]).

3.1 Pointing operations

Pointing operations were first used in [5]. More details as well as a proof of the characterization theorem we need are in the manuscript [9].

Definition 3.5. Let f be an n-ary idempotent operation on a set A and X, Y nonempty subsets of A. We say that f *weakly points* X to Y, if there exist $\mathbf{a^1}, \ldots, \mathbf{a^n} \in A^n$ such that for every $i \in [n]$ and $x \in X$ we have

$$f(a^i_1, \ldots, a^i_{i-1}, x, a^i_{i+1}, \ldots, a^i_n) \in Y$$

(where x is in the ith place). We refer to $\mathbf{a^1}, \ldots, \mathbf{a^n}$ as *witnessing tuples.*

The word "weakly" means that we can have different witnessing tuples for different coordinates, as opposed to (strongly) pointing operations from [9]. For $f : A^k \to A$ and $g : A^n \to A$, we denote by $g \lessdot f$ the kn-ary operation on A defined by

$$(g \lessdot f)(x_1, \ldots, x_{kn}) = g(f(x_1, \ldots, x_k), f(x_{k+1}, \ldots, x_{2k}), \ldots, f(x_{(n-1)k+1}, \ldots, x_{nk})).$$

We will need the following easy observation.

Observation 3.6. *If* $f : A^k \to A$ *weakly points* X *to* Y *and* $g : A^n \to A$ *weakly points* Y *to* Z*, then* $g \lessdot f$ *weakly points* X *to* Z.

Proof. Let the witnessing tuples for f weakly pointing X to Y and g weakly pointing Y to Z be $\mathbf{a^1}, \ldots, \mathbf{a^k}$ and $\mathbf{b^1}, \ldots, \mathbf{b^n}$, respectively. For $i \in [n]$ and $j \in [k]$ define $\mathbf{c^{i,j}} \in A^{nk}$ to be the following tuple:

$$\begin{aligned}\mathbf{c^{i,j}} = (&b_1^i, b_1^i, \ldots, b_1^i, b_2^i, b_2^i, \ldots, b_2^i, \ldots, b_{i-1}^i, b_{i-1}^i, \ldots, b_{i-1}^i,\\ &a_1^j, a_2^j, \ldots, a_k^j, b_{i+1}^i, b_{i+1}^i, \ldots, b_{i+1}^i, \ldots, b_n^i, b_n^i, \ldots, b_n^i),\end{aligned}$$

where b_l^i appears k-times for every $l \in [n] \setminus \{i\}$. It is straightforward to verify (using idempotency of f) that $g \lessdot f$ weakly points X to Z with witnessing tuples $\mathbf{c^{1,1}}, \mathbf{c^{1,2}}, \ldots, \mathbf{c^{1,k}}, \mathbf{c^{2,1}}, \ldots, \mathbf{c^{n,k}}$. □

Of particular interest are term operations weakly pointing the whole algebra (or a subuniverse) to a singleton, due to the following characterization of congruence meet-semidistributivity.

Definition 3.7. Let $\mathbf{A}$ be a finite idempotent algebra. We say that $\mathbf{A}$ *has a weakly pointing operation*, if there exists $\tau \in \operatorname{Clo} \mathbf{A}$ and $a \in A$ such that τ weakly points A to $\{a\}$.

Theorem 3.8 ([9, Theorem 1.3])**.** *A finite idempotent algebra* $\mathbf{A}$ *is* $\mathrm{SD}(\wedge)$*, if and only if every nonempty subuniverse* $B \leq \mathbf{A}$ *has a weakly pointing operation.*

Remark. Using this characterization it is easy to prove that given a finite idempotent algebra $\mathbf{A}$, the class of all $\mathrm{SD}(\wedge)$ members of the pseudovariety generated by $\mathbf{A}$ (that is, quotients of subuniverses of finite powers of $\mathbf{A}$) is closed under taking products, subalgebras and quotients. In particular, we will need the following fact.

Lemma 3.9 ([9, Proposition 2.1(6)])**.** *Let* $\mathbf{A}$ *be a finite idempotent algebra and* B, C *its nonempty subuniverses. If* B *and* C *are* $\mathrm{SD}(\wedge)$*, then* $B \times C$ *(considered as a subuniverse of* $\mathbf{A}^2$*) is* $\mathrm{SD}(\wedge)$ *as well.*

3.2 Absorbing subuniverses

We briefly introduce basic notions and facts from the theory of absorption of Barto and Kozik. For more details see [4, 9].

Definition 3.10. Let $\mathbf{A}$ be an algebra and $B \leq \mathbf{A}$ a nonempty subuniverse. We say that B is an *absorbing subuniverse* of $\mathbf{A}$, and write $B \trianglelefteq \mathbf{A}$, if there exists an idempotent $\tau \in \operatorname{Clo} \mathbf{A}$ such that

$$\begin{aligned}\tau(A, B, B, \ldots, B, B) &\subseteq B,\\ \tau(B, A, B, \ldots, B, B) &\subseteq B,\\ &\vdots\\ \tau(B, B, B, \ldots, B, A) &\subseteq B.\end{aligned}$$

We also say that B absorbs $\mathbf{A}$ *via* τ and call τ an *absorbing operation.*

Note that B absorbs $\mathbf{A}$ *via* τ (say n-ary), if and only if τ (strongly) points A to B and any tuple $\mathbf{b} \in B^n$ can serve as a witnessing tuple for that. Hence absorption is somewhat stronger than pointing operations.

In applications of absorption theory an important role is played by algebras with no proper absorbing subuniverses, the *absorption-free* algebras.

Definition 3.11. An algebra $\mathbf{A}$ is *absorption-free*, if $|A| > 1$, and $B \trianglelefteq \mathbf{A}$ implies that $B = A$.

The following corollary, which is an easy consequence of Theorem 3.8, will be applied several times in our proof.

Corollary 3.12 (see [9, Corollary 2.13])**.** *A finite idempotent algebra* $\mathbf{A}$ *is* $\mathrm{SD}(\wedge)$*, if and only if every absorption-free subuniverse* $B \leq \mathbf{A}$ *has a weakly pointing operation.*

We will use without further notice the following easy facts about absorption:

Lemma 3.13 ([4, Proposition 2.4])**.** *Let* $\mathbf{A}$ *be a finite idempotent algebra.*

- *If* $B \trianglelefteq \mathbf{A}$ *and* $C \trianglelefteq B$*, then* $C \trianglelefteq \mathbf{A}$*.*
- *If* $B \trianglelefteq \mathbf{A}$ *(via* τ*) and* $C \leq A$ *and* $B \cap C \neq \emptyset$*, then* $B \cap C \trianglelefteq C$ *(via* $\tau|_C$*).*

4 The proof

Let us start by introducing notation used throughout the proof. Let $\mathbb{T} = (T; E)$ be an oriented tree of height 1, with $T = A \dot\cup B$ and $E \subseteq A \times B$. We will sometimes write $a \dashrightarrow b$ to mean $(a, b) \in E$. Let $\mathbb{H} = (H; \to)$ be a $\mathbb{T}$-special tree of height h such that $\mathbf{alg}\,\mathbb{H}$ is Taylor. Our aim is to prove that $\mathbf{alg}\,\mathbb{H}$ is $\mathrm{SD}(\wedge)$. We divide the proof into several steps organized into subsections.

4.1 Reduction to the top and bottom levels

Our first step is to show that we can focus only on the top and bottom level of $\mathbb{H}$, i.e., the sets (indeed, subuniverses) A and B. This is the property that justifies the definition of special trees. The reduction was already described in detail in [3] (although the construction there is different).

Lemma 4.1. *Both* A *and* B *are subuniverses of* $\mathbf{alg}\,\mathbb{H}$*. Moreover,* $E \leq A \times B$ *(*$\leq (\mathbf{alg}\,\mathbb{H})^2$*).*

Proof. By Lemma 2.1, it is enough to show that A, B and E are primitive positive definable from $\mathbb{H}$ with constants (although in fact, we will not need the constants). Let $\mathbb{Q}$ be a minimal oriented path of height h which maps homomorphically onto $\mathbb{P}_e$ for all $e \in E$, given by Lemma 1.2. Let us denote by u and v the initial and terminal vertex of $\mathbb{Q}$, respectively. The binary relation E is equal to the set

$$\{(\varphi(u), \varphi(v)) \mid \varphi : \mathbb{Q} \to \mathbb{H} \text{ is a homomorphism}\},$$

which can be expressed by a primitive positive formula. Consequently, $(\exists y)(x \dashrightarrow y)$ and $(\exists y)(y \dashrightarrow x)$ provides us with primitive positive definitions of A and B, respectively. $\square$

It is useful to observe that an n-ary polymorphism can be defined on different components of connectivity of $\mathbb{H}^n$ independently; to verify that it preserves the edges one has to be concerned with inputs from one component at a time only. Among the components a prominent one is the component containing the diagonal: For $n > 0$ we denote by Δ_n the component of connectivity of the digraph $\mathbb{H}^n$ containing the diagonal (i.e., the set $\{(v, \dots, v) : v \in H\}$).

Lemma 4.2. *For any $n > 0$, $(A^n \cup B^n) \subseteq \Delta_n$.*

Proof. It is easily seen that the set $(A^n \cup B^n)$ is connected in the digraph $\mathbb{T}^n$. Let $(\mathbf{a}, \mathbf{b})$ be an edge in $\mathbb{T}^n$ (i.e., $a_i \dashrightarrow b_i$ for $i \in [n]$). Let $\mathbb{Q}$ be a minimal oriented path of height h which maps homomorphically onto all the paths $\{\mathbb{P}_{(a_i,b_i)} \mid i \in [n]\}$, whose existence is provided by Lemma 1.2. For every $i \in [n]$ let $\varphi_i : \mathbb{Q} \to \mathbb{P}_{(a_i,b_i)}$ be a homomorphism. Then the mapping $\varphi : \mathbb{Q} \to \mathbb{H}^n$ given by $\varphi(\mathbf{x}) = (\varphi_1(x_1), \ldots, \varphi_n(x_n))$ is also a homomorphism and it maps the initial and terminal vertex of $\mathbb{Q}$ to $\mathbf{a}$ and $\mathbf{b}$, respectively. This shows that $\mathbf{a}$ and $\mathbf{b}$ are connected in $\mathbb{H}^n$ (via $\varphi(\mathbb{Q})$). Consequently, the whole set $(A^n \cup B^n)$ is connected in $\mathbb{H}^n$. As it intersects the diagonal, it follows that $(A^n \cup B^n) \subseteq \Delta_n$. $\square$

In the next lemma we prove that every polymorphism which is a WNU on the top and bottom levels can be modified to obtain a polymorphism satisfying the WNU property everywhere. In Corollary 4.4 below we combine this fact with Theorem 3.2 to obtain the desired result. The assumption that $n > 2$ is there only to avoid a technical nuisance; in fact, the claim is true for $n = 2$ as well (see [3]).

Lemma 4.3. *Let $n \geq 3$ and let $\tau \in \mathrm{IdPol}_n(\mathbb{H})$ be such that $\tau|_A$ and $\tau|_B$ are WNU operations on A and B, respectively. Then there exists $\tau' \in \mathrm{IdPol}_n(\mathbb{H})$ which is a WNU on H.*

Proof. Let us fix an arbitrary linear order $\leq_E$ of the set E. We define the following linear order $\sqsubseteq$ on the set $H \setminus (A \cup B)$: for $x \in \mathbb{P}_{(a,b)}$ and $y \in \mathbb{P}_{(a',b')}$ we put $x \sqsubset y$ if

- $(a, b) <_E (a', b')$, or
- $(a, b) = (a', b')$ and x is closer to a than y (in $\mathbb{H}$).

We split the definition of τ' into several cases. Fix $\mathbf{x} \in H^n$.

1. If $\mathbf{x} \in A^n \cup B^n$, then we set $\tau'(\mathbf{x}) = \tau(\mathbf{x})$.
2. If $\mathbf{x} \in \Delta_n \setminus (A^n \cup B^n)$, then
   (a) if $\{x_1, \ldots, x_n\} \subseteq \mathbb{P}_{(a,b)}$ for some $(a, b) \in E$, then we define $\tau'(\mathbf{x})$ to be the $\sqsubseteq$-minimal element from $\{x_1, \ldots, x_n\}$,
   (b) if there exists $i \in [n]$ and $e \neq e' \in E$ such that $x_i \in \mathbb{P}_e$ and $x_j \in \mathbb{P}_{e'}$ for all $j \neq i$, then we define
   $$\tau'(\mathbf{x}) = \tau(x_i, x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n),$$
   (c) in all other cases we set $\tau'(\mathbf{x}) = \tau(\mathbf{x})$.
3. If $\mathbf{x} \notin \Delta_n$, then
   (a) if $\mathrm{lvl}(x_1) = \mathrm{lvl}(x_2) = \cdots = \mathrm{lvl}(x_n)$, then we define $\tau'(\mathbf{x})$ to be the $\sqsubseteq$-minimal element from $\{x_1, \ldots, x_n\}$,
   (b) if there exists $i \in [n]$ and $k \neq l$ such that $\mathrm{lvl}(x_i) = k$ and $\mathrm{lvl}(x_j) = l$ for all $j \neq i$, then we define $\tau'(\mathbf{x}) = x_i$,

(c) in all other cases we define $\tau'(\mathbf{x}) = x_1$.

Let us first comment on subcase (2b) of the construction. Since τ is a polymorphism, for any $(a_i, b_i) \in E$, $i \in [n]$, it induces a homomorphism from $\Delta_n \cap \prod_{i=1}^n \mathbb{P}_{(a_i,b_i)}$ (as an induced subgraph of $\mathbb{H}^n$) to $\mathbb{P}_{(\tau(\mathbf{a}),\tau(\mathbf{b}))}$. However, typically there are many such homomorphisms. Even if $\tau(\mathbf{a}) = \tau(\mathbf{a}')$, $\tau(\mathbf{b}) = \tau(\mathbf{b}')$ and $\mathbf{a}', \mathbf{b}'$ are just permutations of $\mathbf{a}, \mathbf{b}$, the two corresponding homomorphisms induced by τ can be different. That is why we cannot simply define $\tau'(\mathbf{x}) = \tau(\mathbf{x})$ in subcase (2b); the WNU property might not hold.

We divide the proof into two separate claims.

Claim. τ' is a polymorphism of $\mathbb{H}$.

Let $(\mathbf{x}, \mathbf{y})$ be an edge in $\mathbb{H}^n$. For every $i \in [n]$ let $e_i = (a_i, b_i) \in E$ be such that $x_i, y_i \in \mathbb{P}_{e_i}$. If $\mathbf{x}$ falls under case (1) of the construction, then $\tau'(\mathbf{x}) = \tau(\mathbf{a})$, and $\mathbf{y}$ falls under one of the subcases of (2). If it is (2a), then $e_1 = \cdots = e_n = e$ for some $e = (a, b) \in E$ and $y_1 = \cdots = y_n = \tau'(\mathbf{y}) = y$, where y is the unique vertex from $\mathbb{P}_e$ such that $a \to y$. Hence $\tau'(\mathbf{x}) \to \tau'(\mathbf{y})$ holds. If it is subcase (2b), then $\mathbf{x} = \mathbf{a} = (a, \dots, a, a', a, \dots, a)$ for some $a, a' \in A$ (where a' is in the ith coordinate) and $\mathbf{y} = (y, \dots, y, y', y, \dots, y)$. Using both that τ is a WNU and a polymorphism we get that $\tau'(\mathbf{x}) = \tau(\mathbf{a}) = \tau(a', a, \dots, a) \to \tau(y', y, \dots, y) = \tau'(\mathbf{y})$. If $\mathbf{y}$ falls under subcase (2c), then for every $i \in [n]$, y_i is the unique vertex from $\mathbb{P}_{e_i}$ such that $a_i \to y_i$ and since τ is a polymorphism we get that $\tau'(\mathbf{x}) = \tau(\mathbf{a}) \to \tau(\mathbf{y}) = \tau'(\mathbf{y})$.

The argument is similar when $\mathbf{y}$ falls under case (1) (and so $\mathbf{x}$ under (2)). In all other situations both $\mathbf{x}$ and $\mathbf{y}$ fall under the same subcase of the construction. Note that since $x_i \to y_i$ and $x_i \notin A$, $y_i \notin B$ (for all $i \in [n]$), it follows that there is an edge between the $\sqsubseteq$-minimal element of $\{x_1, \dots, x_n\}$ and of $\{y_1, \dots, y_n\}$. This implies $\tau'(\mathbf{x}) \to \tau'(\mathbf{y})$ for cases (2a) and (3a).

In cases (2c) and (3c) the polymorphism condition follows immediately from the fact that τ is a polymorphism (in (2c)) and that $x_1 \to y_1$ (in (3c)). For the remaining cases, (2b) and (3b), we have to add the observation that the distinguished coordinate $i \in [n]$ is the same for both $\mathbf{x}$ and $\mathbf{y}$.

Claim. τ' is a WNU on H.

Let $x, y \in H$ be arbitrary. Note that all of the tuples

$$(y, x, \dots, x), (x, y, x, \dots, x), \dots, (x, \dots, x, y)$$

fall under the same case (and subcase) of the construction, and that it can be neither (2c) nor (3c). In case (1) the WNU property follows from the fact that τ is a WNU on A and B while in cases (2a) and (3a) from the fact that the construction in these cases is independent of order and repetition of elements. In case (2b) the result is $\tau(y, x, \dots, x)$ for all the tuples in question while in case (3b) the result is always y. $\square$

Corollary 4.4. *If both A and B are* SD($\wedge$)*, then* $\mathbf{alg}\,\mathbb{H}$ *is* SD($\wedge$).

Proof. By Lemma 3.9, $A \times B$ ($\leq (\mathbf{alg}\,\mathbb{H})^2$) is SD($\wedge$) as well. Hence, by Theorem 3.2, there exists n_0 such that for every $n \geq n_0$ there exists $\tau_n \in \mathrm{IdPol}_n(\mathbb{H})$ such that $(\tau_n \times \tau_n)|_{A \times B}$ is a WNU on $A \times B$. This implies that the restrictions of τ_n to A and B are WNUs. Using Lemma 4.3 we obtain, for every $n \geq \max(n_0, 3)$, a WNU $\tau'_n \in \mathrm{IdPol}_n(\mathbb{H})$. The proof concludes by another application of Theorem 3.2. $\square$

4.2 Singleton absorbing subuniverse

Our next step is to prove that either A or B has a singleton absorbing subuniverse. This is the one and only place where we use the assumption that $\mathbf{alg}\,\mathbb{H}$ is Taylor.

Since $\mathbf{alg}\,\mathbb{H}$ is Taylor, by Theorem 3.2 there exists a WNU operation $\omega \in \mathrm{IdPol}(\mathbb{H})$. Let $\circ : H^2 \to H$ be the *binary polymer* of the WNU ω, that is,

$$x \circ y = \omega(x, x, \dots, y) = \cdots = \omega(y, x, \dots, x)$$

for $x, y \in H$. Note that $\circ \in \mathrm{IdPol}_2(\mathbb{H})$.

We can and will assume that ω is *special* in the sense of [1, Definition 6.2], that is, satisfies $x \circ (x \circ y) = x \circ y$. (Here the word *special* is unrelated to our definition of *special* trees.) This property can be enforced by an iterated composition of ω with itself (i.e., $\omega \lessdot \omega \lessdot \dots \lessdot \omega$, $|H|!$-times, see [1, Lemma 6.4]).

For $x, y \in A \cup B$ we denote by $\mathrm{dist}_E(a, b)$ the *distance* of x and y in $\mathbb{T}$. For a subset $C \subseteq A$ we define the *E-neighbourhood* of C, denoted by $E_+(C)$, to be the set $\{b \in B \mid c \dashrightarrow b \text{ for some } c \in C\}$. Similarly, the E-neighbourhood of $D \subseteq B$ is the set $E_-(D) = \{a \in A \mid a \dashrightarrow d \text{ for some } d \in D\}$. For brevity we write $E_+(c)$, $E_-(d)$ instead of $E_+(\{c\})$, $E_-(\{d\})$. Moreover, for every $k \geq 0$, $C \subseteq A$ and $D \subseteq B$ we inductively define the sets $E_k(C)$ and $E_k(D)$ as follows:

- $E_0(C) = C$ and $E_0(D) = D$,
- $E_1(C) = E_+(C)$ and $E_1(D) = E_-(D)$, and
- $E_k(C) = E_1(E_{k-1}(C))$ and $E_k(D) = E_1(E_{k-1}(D))$ for $k > 1$.

Note that the above definition can be reformulated as follows:

$$E_k(C) = \{x \in A \cup B \mid (\exists\, c \in C)\ \mathrm{dist}_E(x, c) \leq k \ \&\ \mathrm{dist}_E(x, c) \equiv k \pmod 2\},$$

and similarly for $E_k(D)$. We will frequently use the following easy facts (as well as the obvious "dual" versions for $D \leq D' \leq B$), which are all consequences of the fact that $E \leq (\mathbf{alg}\,\mathbb{H})^2$. We leave the proof to the reader.

Observation 4.5. *If $C \leq C' \leq A$, then the following holds:*

- *$E_+(C) \leq E_+(C') \leq B$,*
- *$E_k(C) \leq A$ for k even and $E_k(C) \leq B$ for k odd,*
- *if $k \leq l$ and $l - k$ is even, then $E_k(C) \leq E_l(C)$,*
- *if $C \neq 0$, then there exists k such that $E_k(C) = A$ and $E_{k+1}(C) = B$, and*
- *if $C \trianglelefteq C'$, then for every $k \geq 0$, $E_k(C) \trianglelefteq E_k(C')$ as well and, moreover, the absorption is via the same $\tau \in \mathrm{IdPol}(\mathbb{H})$.* [2]

We are now ready to prove that either A or B has a singleton absorbing subuniverse and, moreover, that this absorption is realized via the WNU operation ω.

Lemma 4.6. *There exists $o \in A \cup B$ such that $\{o\} \trianglelefteq E_2(o)$ via ω.*

[2] Technically, the absorbing operation is $\tau|_{C'}$ in the first case while it is $\tau|_{E_k(C')}$ in the second case, but we will neglect this formality.

Proof. Suppose for contradiction that no such element exists. It follows that for every $u \in A \cup B$ there exists $w \in E_2(u)$ such that $u \circ w = v \neq u$. Since the WNU ω is special, we have that $u \circ v = u \circ (u \circ w) = u \circ w = v$. Consider the binary relation $\gg$ on $A \cup B$ defined by setting $u \gg v$ if and only if $v \in E_2(u) \setminus \{u\}$ and $u \circ v = v$. We have proved that for every $u \in A \cup B$ there exists v such that $u \gg v$.

Let k be maximal such that there exists a sequence $\langle u_0 u_1 \dots u_k \rangle$ of elements of $A \cup B$ with the following properties:

(1) $\mathrm{dist}_E(u_0, u_i) = i$ for all $i \in [k]$, and

(2) $u_i \gg u_{i+2}$ for all $0 \leq i \leq k-2$.

Note that (1) ensures that the sequence is non-repeating and thus, by finiteness of $A \cup B$, such a maximal k exists. The previous paragraph shows that $k \geq 2$: just take $\langle a, b, a' \rangle$ for any $a, a' \in E_-(b)$ such that $a \gg a'$.

Let us assume that $u_k \in A$; the proof for $u_k \in B$ is analogous. Let $u'_k \in A$ and $u'_{k+1} \in B$ be such that $u_{k-1} \gg u'_{k+1}$ and $u_{k-1}, u'_{k+1} \in E_+(u'_k)$ (see the figure below). We will prove that the sequence $\langle u_0 u_1 \dots u_{k-1} u'_k u'_{k+1} \rangle$ also satisfies properties (1) and (2); a contradiction with maximality of k.

$$\begin{tikzcd}
 & & & u_k \arrow[dl, dashed] & \\
\cdots & u_{k-2} \arrow[l, dashed] \arrow[r, dashed] & u_{k-1} & & \\
 & & & u'_k \arrow[ul, dashed] \arrow[r, dashed] & u'_{k+1}
\end{tikzcd}$$

First we prove (1). From $u_{k-1} \gg u'_{k+1}$ we get that $\mathrm{dist}_E(u_{k-1}, u'_k) = 1$ and $\mathrm{dist}_E(u_{k-1}, u'_{k+1}) = 2$. Since $\mathbb{T}$ is a tree, it suffices to rule out the possibility that $u'_k = u_{k-2}$. In that case $u_{k-2} \dashrightarrow u'_{k+1}$, $u_{k-2} \dashrightarrow u_{k-1}$ and $u_k \dashrightarrow u_{k-1}$ would give

$$\omega(u_{k-2}, u_{k-2}, \dots, u_{k-2}, u_k) \dashrightarrow \omega(u'_{k+1}, u_{k-1}, \dots, u_{k-1}, u_{k-1}).$$

The left hand side is $u_{k-2} \circ u_k = u_k$ while the right hand side is $u_{k-1} \circ u'_{k+1} = u'_{k+1}$; and so we get $u_k \dashrightarrow u'_{k+1}$. But $u_k \in E_-(u_{k-1}) \cap E_-(u'_{k+1})$ would imply that $u_k = u'_k = u_{k-2}$ which contradicts $u_{k-2} \gg u_k$.

To prove (2) we only need to establish $u_{k-2} \gg u'_k$. From $u_{k-2} \dashrightarrow u_{k-1}$, $u'_k \dashrightarrow u'_{k+1}$ and the fact that $\circ$ preserves E we get

$$u_{k-2} \circ u'_k \dashrightarrow u_{k-1} \circ u'_{k+1} = u'_{k+1}.$$

On the other hand, $\{u_{k-2}, u'_k\} \subseteq E_-(u_{k-1})$, which is a subuniverse, and thus $u_{k-2} \circ u'_k \dashrightarrow u_{k-1}$. It follows that $u_{k-2} \circ u'_k = u'_k$; and $u_{k-2} \neq u'_k$ is proved above. □

Fix $o \in A \cup B$ given by the previous lemma. To simplify the exposition we choose that $o \in A$. The proofs are essentially the same in the other case (moreover, note that reversing edges of $\mathbb{H}$ does not change $\mathbf{alg}\,\mathbb{H}$).

Since $\mathbb{H}$ is an oriented tree, it follows that for every $v \in H$ there exists a unique oriented path $\mathbb{Q}_{o,v}$ connecting o to v in $\mathbb{H}$. We define a partial order $\preceq$ on H by setting $u \preceq v$ if and only if $u \in \mathbb{Q}_{o,v}$. Note that o is the minimum element in this order. Furthermore, for $u, v \in A \cup B$, $u \preceq v$ implies $\mathrm{dist}_E(o, u) \leq \mathrm{dist}_E(o, v)$.

Lemma 4.7. *If $a, a' \in A$ and $a \preceq a'$, then $a \circ a' = a$ (and similarly for $b, b' \in B$). In particular, $\{o\} \trianglelefteq A$ via ω.*

Proof. If $a = a'$, then $a \circ a' = a$ follows trivially from idempotency of ω. Else, there exists $k \geq 0$ such that $a \in E_k(o)$ and $a' \in E_{k+2}(o) \setminus E_k(o)$. From Lemma 4.6 and the last item of Observation 4.5 it follows that $E_k(o) \trianglelefteq E_{k+2}(o)$ via ω and so $a \circ a' \in E_k(o)$. In particular, $a \circ a' \neq a'$.

Note that $l = \mathrm{dist}_E(a, a')$ is even and that there exists a unique vertex $u \in \mathbb{Q}_{o,a'} \cap (A \cup B)$ such that $\mathrm{dist}_E(a, u) = \mathrm{dist}_E(u, a') = l/2$. Since $a, a' \in E_{l/2}(u)$, which is a subuniverse, we have $a \circ a' \in E_{l/2}(u)$ while a is the $\preceq$-minimal element of $E_{l/2}(u)$. It follows that $a \preceq a \circ a'$.

Suppose for contradiction that $a \neq a \circ a'$. Then repeating the arguments from the first paragraph with $a \circ a'$ in the role of a' yields $a \circ (a \circ a') \neq a \circ a'$ which contradicts the fact that ω is a special WNU.

Hence we have proved that $a \circ a' = a$. The proof for $b \preceq b'$ is essentially the same. The fact that $\{o\} \trianglelefteq A$ via ω now follows immediately from the definition of absorption and the fact that o is the $\preceq$-minimum element of A. $\square$

Remark. Incidentally, the Absorption Theorem of Barto and Kozik [4, Theorem 2.3] applied to A, B and E immediately yields that either A or B has a singleton absorbing subuniverse. We need a slightly stronger fact for our proof (namely that the absorbing operation is a WNU); it is however likely that the claim of the Absorption theorem can be strengthened to replace the above ad hoc argument. Our argument can be viewed as a proof of a special case of the Absorption Theorem, where the relation E is acyclic.

Existence of the singleton absorbing subuniverse $\{o\}$ already significantly restricts living space for possible absorption-free subuniverses in A and B, as we can see in the next lemma. (Of course, the dual version for $D \leq B$ is also true.)

Lemma 4.8. *If $C \leq A$ is absorption-free, then there exists $k > 0$ such that $\mathrm{dist}_E(o, c) = k$ for all $c \in C$.*

Proof. Let k be the minimum from the set $\{\mathrm{dist}_E(o, c) \mid c \in C\}$. Since $\{o\} \trianglelefteq A$, by Observation 4.5 we have $E_k(o) \trianglelefteq E_k(A) = A$, and thus also $C \cap E_k(o) \trianglelefteq C \cap A = C$. Since C is absorption-free, it follows that $C \cap E_k(o) = C$.

We have proved that $k \leq \mathrm{dist}_E(o, c) \leq k$ for all $c \in C$. Note that $k > 0$, since otherwise $C = \{o\}$ which is not absorption-free by definition. $\square$

4.3 E-neighbourhoods of singletons are $\mathrm{SD}(\wedge)$

In this subsection we prove that E-neighbourhoods of elements from $A \cup B$ are $\mathrm{SD}(\wedge)$. Our strategy is to show that whenever they have an absorption-free subuniverse, it must have a weakly pointing operation (and then apply Corollary 3.12). For the rest of this subsection we fix $b \in B$ and an absorption-free subuniverse $C \leq E_-(b)$. (The proof for $D \leq E_+(a)$ is analogous.)

From Lemma 4.8 and the fact that $|C| > 1$ (and that $\mathbb{T}$ is a tree) we see that $b \prec c$ for all $c \in C$. In the first step we prove that elements from B which are $\preceq$-above C are "absorbed by b" via a certain binary operation $\star$. (Note that such elements do not need to form a subuniverse, and so it is not absorption in the sense we defined.) Later we will use this operation to construct various binary polymorphisms and then build up a weakly pointing operation for C from them.

Let us denote by $\star$ the binary idempotent polymorphism of $\mathbb{H}$ given by

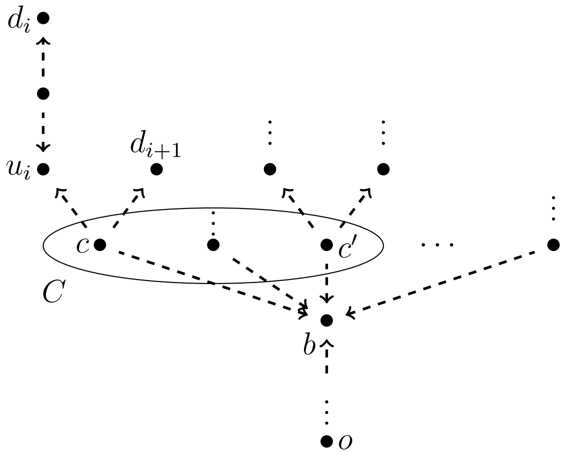
$$x \star y = (\dots(((x \underbrace{\circ y) \circ y) \circ \dots \circ}_{|H|\times} y),$$

where the operation $\circ$ appears $|H|$-times (just for good measure).

Lemma 4.9. *If $d \in B$ is such that $c \prec d$ for some $c \in C$, then $b \star d = d \star b = b$.*

Proof. From Lemma 4.7 we get that $b \circ d = b$ and thus also $b \star d = b$. To prove the other equality, fix $d \in B$ and $c \in C$ with $b \prec c \prec d$ and consider the sequence $\langle d_0 d_1 \dots d_{|H|} \rangle$ of elements of B defined inductively by setting $d_0 = d$ and $d_i = d_{i-1} \circ b$ for $i \in [|H|]$. Observe that $d_{|H|} = d \star b$ which we want to equate to b.

Let k_i denote the distance $\mathrm{dist}_E(d_i, b)$. We will prove that for every $0 \le i \le |H|$, $b \preceq d_i$ and $k_i \le k_{i-1}$ (we set $k_{-1} = k_0$). The proof uses induction on i; the case $i = 0$ is trivial. Assume that the claim holds for some $i < |H|$. Following the same argument as in the proof of Lemma 4.7, there exists $u_i \in A \cup B$ such that $\mathrm{dist}_E(u_i, b) = \mathrm{dist}_E(u_i, d_i) = k_i/2$. Since $b \preceq d_i$, it follows that b is the $\preceq$-minimal (and d_i a $\preceq$-maximal) element of $E_{k_i/2}(u_i)$. Consequently, $d_{i+1} = d_i \circ b \in E_{k_i/2}(u_i)$ implies that $b \preceq d_{i+1}$ and $k_{i+1} \le k_i$ (see the figure below).



Note that $k_0 < |H|$, and so there must exist $i < |H|$ such that $k_i = k_{i+1}$. Denote this distance by k and suppose for contradiction that $k \neq 0$ (and so $k \ge 2$, since k is even). Pick any $c' \in C$. Since $c \in E_{k-1}(d_i)$, $c' \in E_-(b) \le E_{k-1}(b)$, $d_i \circ b = d_{i+1}$ and $\circ$ preserves E, it follows that $c \circ c' \in E_{k-1}(d_{i+1})$. But we also have $c \circ c' \in C$ and $E_{k-1}(d_{i+1}) \cap C = \{c\}$. Thus we have proved that $c \circ c' = c$ for all $c' \in C$, which means that $\{c\} \trianglelefteq C$ via ω, a contradiction with C being absorption-free. Therefore it must be the case that $k = 0$, which means $d_i = b$ and thus by idempotency of $\circ$ also $d_{|B|} = d \star b = b$. $\square$

Let us denote by $\mathcal{F}$ the smallest set of binary operations on H satisfying

- $x \star y \in \mathcal{F}$, $y \star x \in \mathcal{F}$,
- if $\varphi(x, y) \in \mathcal{F}$, then $\{x \star \varphi(x, y), y \star \varphi(x, y), \varphi(x, y) \star x, \varphi(x, y) \star y\} \subseteq \mathcal{F}$,

- if $\varphi(x,y), \varphi'(x,y) \in \mathcal{F}$, then $(\varphi(x,y) \star \varphi'(x,y)) \in \mathcal{F}$.

From Lemma 4.9 and the construction of $\mathcal{F}$ we immediately obtain the following:

Corollary 4.10. *If $d \in B$ is such that $c \prec d$ for some $c \in C$, then $\varphi(b,d) = \varphi(d,b) = b$ for every $\varphi \in \mathcal{F}$.*

For every $c, c' \in C$ let $S_{c,c'}$ be the set $\{\varphi(c,c') \mid \varphi(x,y) \in \mathcal{F}\} \subseteq C$. We will use the following easy facts:

- $S_{c,c} = \{c\}$,
- $S_{c,c'} = S_{c',c}$,
- both $S_{c,c'}$ and $S_{c,c'} \cup \{c, c'\}$ are closed under the operation $\star$,
- in particular, if $x, y \in S_{c,c'}$, then $S_{x,y} \subseteq S_{c,c'}$.

Remark. Alternatively, using terminology from universal algebra, we could have defined $\mathcal{F}$ to be the set of all binary terms in the binary operation symbol $\star$ which contain both the variables x and y. Then $S_{c,c'}$ would be the image of $\mathcal{F}$ under the homomorphism from the absolutely free two-generated algebra to $(C; \{\star\})$ given by $x \mapsto c$ and $y \mapsto c'$.

Note that $\mathcal{F} \subseteq \mathrm{IdPol}_2(\mathbb{H})$. In the next lemma we prove that, in fact, $\mathbb{H}$ has many more binary idempotent polymorphisms.

Lemma 4.11. *Let $\gamma : C^2 \to C$ be any binary operation such that $\gamma(c,c') \in S_{c,c'}$ for all $c, c' \in C$. Then there exists $\tau \in \mathrm{IdPol}_2(\mathbb{H})$ extending γ (i.e., $\tau|_C = \gamma$).*

Proof. For every $c, c' \in C$ we fix some $\varphi_{c,c'}(x,y) \in \mathcal{F}$ witnessing that $\gamma(c,c') \in S_{c,c'}$. For $x, y \in H$ we define $\tau(x,y)$ in the following way:

1. If there exist $c, c' \in C$ such that
   - $b \prec x \prec c$ or $c \preceq x$,
   - $b \prec y \prec c'$ or $c' \preceq y$, and
   - $\mathrm{lvl}(x) = \mathrm{lvl}(y)$,

   then we set $\tau(x,y) = \varphi_{c,c'}(x,y)$.
2. Else, we define $\tau(x,y) = x \star y$.

It follows immediately from the construction that τ is idempotent and $\tau|_C = \gamma$. To prove that $\tau \in \mathrm{IdPol}(\mathbb{H})$, let $x \to u$, $y \to v$ be arbitrary edges of $\mathbb{H}$. Note that since $\star$ and $\varphi_{c,c'}$ (for any $c, c' \in C$) are polymorphisms of $\mathbb{H}$, $\tau(x,y) \to \tau(u,v)$ follows immediately if both $\{x,y\}$ and $\{u,v\}$ fall under the same case of the construction. If they do not, then it must be the case that $\{x,y\}$ falls under case (1) while $\{u,v\}$ under case (2) (it cannot be the opposite, since if $b \prec u$ and $x \to u$, then $b \prec x$ as well, and similarly for the other conditions). Thus $\tau(x,y) = \varphi_{c,c'}(x,y)$ for some $c, c' \in C$ and $\tau(u,v) = u \star v$. Moreover it must be the case that $b \preceq u$, $b \preceq v$ and $b \in \{u,v\}$. As $\mathrm{lvl}(x) = \mathrm{lvl}(y)$ implies that $\mathrm{lvl}(u) = \mathrm{lvl}(v)$, we get $u, v \in B$. It follows that $\tau(u,v) = u \star v = b = \varphi_{c,c'}(u,v)$, either by Corollary 4.10 or by idempotency (in case that $u = v = b$). We conclude that $\tau(x,y) \to \tau(u,v)$ in this case as well. □

As an easy consequence of this lemma, we can prove that C has a binary idempotent commutative operation (i.e., a binary WNU).

Corollary 4.12. *There exists $\varphi \in \mathrm{IdPol}_2(\mathbb{H})$ such that $\varphi|_C$ is commutative.*

Proof. For every $c, c' \in C$ define $\gamma(c, c') = \gamma(c', c)$ to be an arbitrary element from $S_{c,c'}$ thus making γ commutative, and then apply Lemma 4.11. $\square$

The above corollary implies that $|C| > 2$, since a binary WNU on a 2-element set is a semilattice operation which would violate absorption-freeness. Unfortunately, a binary WNU is not enough to construct a weakly pointing operation for C; we need a slightly more involved argument.

Lemma 4.13. *C has a weakly pointing operation.*

Proof. We start by showing that every two-element set is weakly pointed to a singleton by some operation, with an additional "symmetry" property.

Claim. For every $x, y \in C$ there exist $\varphi \in \mathrm{IdPol}(\mathbb{H})$ (say it is n-ary), $z \in C$, $\mathbf{c^1}, \dots, \mathbf{c^n} \in C^n$ and $\alpha : C \to C$ such that the following hold:

1. $\varphi|_C$ weakly points $\{x, y\}$ to $\{z\}$ with witnessing tuples $\mathbf{c^1}, \dots, \mathbf{c^n}$.
2. For every $i \in [n]$ and $u \in C$, $\varphi(c^i_1, c^i_2, \dots, c^i_{i-1}, u, c^i_{i+1}, \dots, c^i_n) = \alpha(u)$.

We will prove the claim by induction on $|S_{x,y} \cup \{x, y\}|$. Assume first that $S_{x,y} \cap \{x, y\} \neq \emptyset$, say $x \in S_{x,y}$ (the argument for $y \in S_{x,y}$ is analogous). In that case we can apply Lemma 4.11 to construct $\varphi \in \mathrm{IdPol}_2(\mathbb{H})$ such that $\varphi(x, y) = \varphi(y, x) = \varphi(x, x) = x$ and $\varphi|_C$ is commutative (see the proof of Corollary 4.12). The claim follows since $\varphi|_C$ weakly points $\{x, y\}$ to $\{x\}$, the witnessing tuple is (x, x) for both coordinates and $\alpha(u) = \varphi(u, x)$ for all $u \in C$. This also covers the base step of our induction (i.e., $S_{x,y} \subseteq \{x, y\}$).

We can now assume that $S_{x,y} \cap \{x, y\} = \emptyset$. Let us define $c = x \star y$, $x' = x \star c$ and $y' = y \star c$. Using Lemma 4.11 we can construct $\varphi \in \mathrm{IdPol}_2(\mathbb{H})$ such that $\varphi(x, c) = \varphi(c, x) = x'$ and $\varphi(y, c) = \varphi(c, y) = y'$ and $\varphi|_C$ is commutative. In particular, $\varphi|_C$ points $\{x, y\}$ to $\{x', y'\}$, the witnessing tuple is (c, c) for both coordinates.

Since $x', y' \in S_{x,y}$, it follows that $S_{x',y'} \cup \{x', y'\} \subseteq S_{x,y} \subsetneq S_{x,y} \cup \{x, y\}$. Hence, by induction assumption, the claim holds for x', y'. Let it be witnessed by $\psi \in \mathrm{IdPol}(\mathbb{H})$ weakly pointing $\{x', y'\}$ to $\{z\}$ and let $\alpha' : C \to C$ be the corresponding mapping from (2).

Using Observation 3.6 we get that $(\psi \lessdot \varphi)|_C$ weakly points $\{x, y\}$ to $\{z\}$ and it is not hard to see from its proof that (2) holds as well, with $\alpha : C \to C$ given by $\alpha(u) = \alpha'(\varphi(u, c))$, for $u \in C$. We leave the verification to the reader.

We will now compose the operations from this claim to construct a weakly pointing operation for C; we use another induction argument.

Claim. For every nonempty $X \subseteq C$ there exists $c \in C$ and $\varphi \in \mathrm{IdPol}(\mathbb{H})$ such that $\varphi|_C$ weakly points X to $\{c\}$.

We prove the claim by induction on $|X|$. If $X = \{x\}$, then the claim is trivial: take any $\varphi \in \mathrm{IdPol}(\mathbb{H})$, $z = x$ and witnessing tuple $(x, x, \dots, x)$ for all coordinates. Let $|X| = k > 1$ and assume that the claim holds for all at most $(k-1)$-element subsets of C. Pick any $x, y \in X$, $x \neq y$ and let $\varphi \in \mathrm{IdPol}(\mathbb{H})$ (say

n-ary), $z \in C$ and $\alpha : C \to C$ be the objects given by the previous claim applied to x and y. It is easy to see that $\varphi|_C$ weakly points X to $Y = \{\alpha(x) \mid x \in X\}$ (this is why we need the "symmetry" property from the previous claim). Since $\alpha(x) = z = \alpha(y)$, it follows that $|Y| < |X|$. By induction assumption, there exists $\psi \in \mathrm{IdPol}(\mathbb{H})$ and $c \in C$ such that $\psi|_C$ weakly points Y to $\{c\}$. Using Observation 3.6 we get that $(\psi \lessdot \varphi)|_C$ weakly points X to $\{c\}$ which concludes the proof.

□

We have achieved the goal of this subsection, i.e., the following corollary.

Corollary 4.14. *For every $b \in B$, $E_-(b)$ is* $\mathrm{SD}(\wedge)$. *Similarly for $a \in A$ and $E_+(a)$.*

Proof. By Lemma 4.13, every absorption-free subuniverse $C \leq E_-(b)$ has a weakly pointing operation and so we can apply Corollary 3.12. The proof for $a \in A$ is analogous. □

4.4 Absorption-free subuniverses are $\mathrm{SD}(\wedge)$

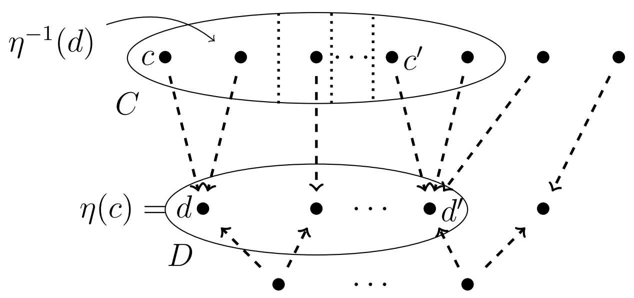
The last step of our proof is to show that every absorption-free subuniverse C of A or B has a weakly pointing operation. Theorem 1.4 will then follow from Corollary 3.12 and Corollary 4.4.

Lemma 4.15. *Every absorption-free subuniverse C of A or B has a weakly pointing operation.*

Proof. Recall that by Lemma 4.8, for every absorption-free subuniverse C of A or B there exists $k > 0$ such that $\mathrm{dist}_E(c, o) = k$ for all $c \in C$. We will proceed by induction on this distance k. The base step, $k = 1$, follows from Lemma 4.13 from the previous subsection, since in that case $C \leq E_+(o)$.

Let $k > 1$ and assume that $C \leq A$ (the proof for $C \leq B$ is analogous). Let us denote by D the subuniverse $D = E_+(C) \cap E_{k-1}(o) \leq B$. If $D = \{d\}$ for some $d \in B$, then $C \leq E_-(d)$ and C has a weakly pointing operation by Lemma 4.13. Thus we can assume that $|D| > 1$.

The binary relation $E \cap (C \times D)$ induces an onto mapping $\eta : C \to D$ defined by $\eta(c) = d$, where $d \in D$ is unique such that $(c, d) \in E$ (this is because $\mathbb{T}$ is a tree; see the figure below).



The relation $E \cap (C \times D)$ is preserved by every $\varphi \in \mathrm{IdPol}(\mathbb{H})$ (see Lemma 4.1). The following are easy consequences of this fact:

- for every $D' \leq D$ the set $\eta^{-1}(D')$ is a subuniverse of C,
- if $D' \trianglelefteq D$, then $\eta^{-1}(D') \trianglelefteq C$ (the absorbing polymorphism is the same),
- for $D' \leq D$, $D' = D$ if and only if $\eta^{-1}(D') = C$ (since η is onto).

Combining these facts together with the fact that C is absorption-free yields that D is absorption-free. Hence by induction assumption D has a weakly pointing operation.

Let $\varphi \in \mathrm{IdPol}(\mathbb{H})$ (say n-ary) be such that $\varphi|_D$ weakly points D to $\{d\}$ with witnessing tuples $\mathbf{d^1}, \ldots, \mathbf{d^n}$. It is easy to verify that $\varphi|_C$ weakly points C to $\eta^{-1}(d)$; any $\mathbf{c^1}, \ldots, \mathbf{c^n} \in C^n$ such that $\eta(c^i_j) = d^i_j$ (for $i, j \in [n]$) can serve as witnessing tuples.

Since $\eta^{-1}(d) \leq E_-(d)$, it follows from Corollary 4.14 and Theorem 3.8 that $\eta^{-1}(d)$ has a weakly pointing operation. Let $\psi \in \mathrm{IdPol}(\mathbb{H})$ and $c \in \eta^{-1}(d)$ be such that $\psi|_{\eta^{-1}(d)}$ weakly points $\eta^{-1}(d)$ to $\{c\}$. In particular, $\psi|_C$ weakly points $\eta^{-1}(d)$ to $\{c\}$ and thus by Observation 3.6, $(\psi \lessdot \varphi)|_C$ weakly points C to c. □

Remark. In the language of universal algebra, the relation $E \cap (C \times D)$ is the graph of an onto homomorphism $\eta : C \to D$ and thus, by the First Isomorphism Theorem, D is isomorphic to the quotient of C over the kernel of η. The induction step in the previous lemma follows easily from this observation.

Proof of Theorem 1.4 and Corollary 1.5. Let $\mathbb{H}$ be a special tree such that $\mathbf{alg}\,\mathbb{H}$ is Taylor. In Lemma 4.15 we proved that every absorption-free subuniverse of A or B has a weakly pointing operation. By Corollary 3.12, both A and B are $\mathrm{SD}(\wedge)$ and thus it follows from Corollary 4.4 that $\mathbf{alg}\,\mathbb{H}$ is $\mathrm{SD}(\wedge)$.

It is easy to see that the core of a special tree is again a special tree. If $\mathbb{H}$ is a core, then either $\mathbf{alg}\,\mathbb{H}$ is not Taylor, in which case $\mathrm{CSP}(\mathbb{H})$ is **NP**-complete by Theorem 3.3, or $\mathbf{alg}\,\mathbb{H}$ is $\mathrm{SD}(\wedge)$ and $\mathbb{H}$ has bounded width by Theorem 3.4. □

5 Discussion

We believe that given the evidence, it is reasonable to conjecture that our result generalizes to all oriented trees. Moreover, we hope that the techniques developed in this paper will be useful in pursuit of the proof.

Conjecture 2. *For every oriented tree $\mathbb{H}$, either* $\mathbf{alg}\,\mathbb{H}$ *is not Taylor or it is* $\mathrm{SD}(\wedge)$*. In particular, if $\mathbb{H}$ is a core, then $\mathbb{H}$ has bounded width or* $\mathrm{CSP}(\mathbb{H})$ *is* ***NP****-complete.*

The reader may wonder why we need two different characterizations of $\mathrm{SD}(\wedge)$ algebras, i.e., why we use WNU operations for the proof of Corollary 4.4. The reason is that our techniques used later in the proof are not well suited to deal with non-diagonal components of connectivity of powers of $\mathbb{H}$. This is one obstacle to generalizing the result to all oriented trees.

Another shortcoming is that we cannot get a good handle of polymorphisms of higher arities than binary. For example, it follows from Corollary 4.12 that neither A nor B can have a two-element absorption-free subuniverse and in fact, we can prove that $\mathbf{alg}\,\mathbb{H}$ (if it is Taylor) cannot have a two-element absorption-free subuniverse at all (we will not present the argument here, but it is similar in spirit to the proof of Lemma 4.3). We do not know if this result can be extended to more than two elements. Hence the following open problem.

Problem. Let $\mathbb{H}$ be a (special, or any oriented) tree such that $\mathbf{alg}\,\mathbb{H}$ is Taylor. Is $\mathbf{alg}\,\mathbb{H}$ *always absorbing*?

A finite idempotent algebra $\mathbf{A}$ is *always absorbing*, if for every nonempty $B \leq \mathbf{A}$ there exists $b \in B$ such that $\{b\} \trianglelefteq B$. (Equivalently, there are no absorption-free algebras in the pseudovariety generated by $\mathbf{A}$, see [9, Proposition 2.1].) By Corollary 3.12, always absorbing algebras are SD($\wedge$). A positive answer to this problem would significantly simplify our proof.

Special balanced digraphs, a natural relaxation of the definition of special trees to balanced digraphs, appear naturally in the reduction of constraint satisfaction problems to digraph $\mathbb{H}$-coloring [17, 18, 20]. The reader may notice similarities with some of the proofs in [18]. Can our techniques be adapted to obtain interesting results about special balanced digraphs?

Acknowledgements

The author would like to thank Libor Barto for thoughtful comments and discussions and valuable inputs which improved this paper in several places. This research was supported by the Grant Agency of the Czech Republic projects no.201/09/H012 and 13-01832S, Charles University Grant Agency projects 67410, 558313 and SVV-2014-260107 and Ministry of Education of the Czech Republic project 7AMB 13P10 13.

References

[1] Barto, L., Kozik, M.: Constraint satisfaction problems of bounded width. In: 50th Annual IEEE Symposium on Foundations of Computer Science, 2009. FOCS ’09. pp. 595–603 (2009)

[2] Barto, L.: The dichotomy for conservative constraint satisfaction problems revisited. In: In Proceedings of the 26th Annual IEEE Symposium on Logic in Computer Science, LICS 2011. pp. 21–24. IEEE Computer Society (2011)

[3] Barto, L., Bulín, J.: CSP dichotomy for special polyads. International Journal of Algebra and Computation 23(5), 1151–1174 (2013)

[4] Barto, L., Kozik, M.: Absorbing subalgebras, cyclic terms, and the constraint satisfaction problem. Logical Methods in Computer Science 8(1), 1:07, 27 (2012)

[5] Barto, L., Kozik, M.: Robust satisfiability of constraint satisfaction problems. In: Proceedings of the Forty-fourth Annual ACM Symposium on Theory of Computing. pp. 931–940. STOC ’12, ACM, New York, NY, USA (2012)

[6] Barto, L., Kozik, M.: Constraint satisfaction problems solvable by local consistency methods. Journal of the ACM 61(1), Art. 3, 19 (2014)

[7] Barto, L., Kozik, M., Maróti, M., Niven, T.: CSP dichotomy for special triads. Proceedings of the American Mathematical Society 137(9), 2921–2934 (2009)

[8] Barto, L., Kozik, M., Niven, T.: The CSP dichotomy holds for digraphs with no sources and no sinks (a positive answer to a conjecture of Bang-Jensen and Hell). SIAM Journal on Computing 38(5), 1782–1802 (2008)

[9] Barto, L., Kozik, M., Stanovský, D.: Mal'tsev conditions, lack of absorption, and solvability. Manuscript (2013), `http://www.karlin.mff.cuni.cz/~stanovsk/math/maltsev_absorption_solvability.pdf`

[10] Bergman, C.: Universal Algebra: Fundamentals and Selected Topics. Chapman and Hall/CRC, Boca Raton, 1st edn. (2011)

[11] Berman, J., Idziak, P., Marković, P., McKenzie, R., Valeriote, M., Willard, R.: Varieties with few subalgebras of powers. Transactions of the American Mathematical Society 362(3), 1445–1473 (2010)

[12] Bodnarčuk, V.G., Kalužnin, L.A., Kotov, V.N., Romov, B.A.: Galois theory for post algebras. i, II. Otdelenie Matematiki, Mekhaniki i Kibernetiki Akademii Nauk Ukrainskoi SSR. Kibernetika no. 5(3), 1–9 (1969)

[13] Bulatov, A.: Tractable conservative constraint satisfaction problems. In: Proceedings of the 18th Annual IEEE Symposium on Logic in Computer Science. p. 321. LICS '03, IEEE Computer Society, Washington, DC, USA (2003)

[14] Bulatov, A.: A dichotomy theorem for constraint satisfaction problems on a 3-element set. J. ACM 53(1), 66–120 (2006)

[15] Bulatov, A., Jeavons, P., Krokhin, A.: Classifying the complexity of constraints using finite algebras. SIAM Journal on Computing 34(3), 720–742 (2005)

[16] Bulatov, A., Valeriote, M.: Recent results on the algebraic approach to the CSP. In: Complexity of Constraints. pp. 68–92 (2008)

[17] Bulín, J., Delić, D., Jackson, M., Niven, T.: On the reduction of the CSP dichotomy conjecture to digraphs. In: Schulte, C. (ed.) Principles and Practice of Constraint Programming, pp. 184–199. No. 8124 in Lecture Notes in Computer Science, Springer Berlin Heidelberg (2013)

[18] Bulín, J., Delić, D., Jackson, M., Niven, T.: A finer reduction of constraint problems to digraphs. arXiv:1406.6413 [cs, math] (2014), `http://arxiv.org/abs/1406.6413`

[19] Feder, T.: Classification of homomorphisms to oriented cycles and of k-partite satisfiability. SIAM J. Discret. Math. 14(4), 471–480 (2001)

[20] Feder, T., Vardi, M.Y.: The computational structure of monotone monadic SNP and constraint satisfaction: a study through datalog and group theory. SIAM Journal on Computing 28(1), 57–104 (electronic) (1999)

[21] Geiger, D.: Closed systems of functions and predicates. Pacific Journal of Mathematics 27, 95–100 (1968)

[22] Gutjahr, W., Welzl, E., Woeginger, G.: Polynomial graph-colorings. Discrete Applied Mathematics. The Journal of Combinatorial Algorithms, Informatics and Computational Sciences 35(1), 29–45 (1992)

[23] Hell, P., Nešetřil, J., Zhu, X.: Complexity of tree homomorphisms. Discrete Applied Mathematics 70(1), 23–36 (1996)

[24] Hell, P., Nešetřil, J., Zhu, X.: Duality and polynomial testing of tree homomorphisms. Transactions of the American Mathematical Society 348(4), 1281–1297 (1996)

[25] Hell, P., Nešetřil, J.: Graphs and Homomorphisms. Oxford University Press, Oxford; New York (2004)

[26] Hell, P., Nešetřil, J.: On the complexity of H-coloring. J. Comb. Theory Ser. B 48(1), 92–110 (1990)

[27] Häggkvist, R., Hell, P., Miller, D.J., Neumann Lara, V.: On multiplicative graphs and the product conjecture. Combinatorica. An International Journal of the János Bolyai Mathematical Society 8(1), 63–74 (1988)

[28] Idziak, P., Mckenzie, R., Willard, R.: Tractability and learnability arising from algebras with few subpowers. In: In LICS'07. pp. 213–224. IEEE Computer Society (2007)

[29] Jeavons, P., Cohen, D., Gyssens, M.: Closure properties of constraints. J. ACM 44(4), 527–548 (1997)

[30] Larose, B., Tesson, P.: Universal algebra and hardness results for constraint satisfaction problems. Theoretical Computer Science 410(18), 1629–1647 (2009), automata, Languages and Programming (ICALP 2007)

[31] Larose, B., Zádori, L.: Bounded width problems and algebras. Algebra universalis 56(3-4), 439–466 (2007)

[32] Maróti, M., McKenzie, R.: Existence theorems for weakly symmetric operations. Algebra Universalis 59(3-4), 463–489 (2008)

[33] Schaefer, T.J.: The complexity of satisfiability problems. In: Proceedings of the Tenth Annual ACM Symposium on Theory of Computing. pp. 216–226. STOC '78, ACM, New York, NY, USA (1978)