Constraint satisfaction problems for reducts of homogeneous graphs

Manuel Bodirsky[*1], Barnaby Martin[†2], Michael Pinsker[‡3], and András Pongrácz[§4]

1 **Institut für Algebra, TU Dresden, Germany**
2 **School of Science and Technology, Middlesex University, London**
3 **Department of Algebra, Charles University, Czech Republic and Institute für Computersprachen, TU Wien, Austria.**
4 **Department of Algebra and Number Theory, University of Debrecen, Hungary**

Abstract

For $n \geq 3$, let (H_n, E) denote the n-th Henson graph, i.e., the unique countable homogeneous graph with exactly those finite graphs as induced subgraphs that do not embed the complete graph on n vertices. We show that for all structures Γ with domain H_n whose relations are first-order definable in (H_n, E) the constraint satisfaction problem for Γ is either in P or is NP-complete.

We moreover show a similar complexity dichotomy for all structures whose relations are first-order definable in a homogeneous graph whose reflexive closure is an equivalence relation.

Together with earlier results, in particular for the random graph, this completes the complexity classification of constraint satisfaction problems of structures first-order definable in countably infinite homogeneous graphs: all such problems are either in P or NP-complete.

1998 ACM Subject Classification F.2.2 Nonnumerical Algorithms and Problems

Keywords and phrases Constraint Satisfaction, Homogeneous Graphs, Computational Complexity, Universal Algebra, Ramsey Theory

Digital Object Identifier 10.4230/LIPIcs.CVIT.2016.23

1 Introduction

1.1 Constraint satisfaction problems

A *constraint satisfaction problem* (CSP) is a computational problem in which the input consists of a finite set of variables and a finite set of *constraints*, and where the question is whether there exists a mapping from the variables to some fixed domain such that all the constraints are satisfied. We can thus see the possible constraints as relations on the domain, and in an instance of the CSP, we are asked to assign domain values to the variables such that certain specified tuples of variables become elements of certain specified relations.

* The first and fourth authors have received funding from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013 Grant Agreement no. 257039). Manuel Bodirsky has also been supported by the DFG-funded project 'Topo-Klon' (Project number 622397).
† The second and fourth authors were supported by EPSRC grant EP/L005654/1.
‡ The third author has received funding from project P27600 of the Austrian Science Fund (FWF).
§ The fourth author was also supported by the Hungarian Scientific Research Fund (OTKA) grant no. K109185..

When the domain is finite, and arbitrary constraints are permitted, then the CSP is NP-complete. However, when only constraints from a restricted set of relations on the domain are allowed in the input, there might be a polynomial-time algorithm for the CSP. The set of relations that is allowed to formulate the constraints in the input is often called the *constraint language*. The question which constraint languages give rise to polynomial-time solvable CSPs has been the topic of intensive research over the past years. It has been conjectured by Feder and Vardi [21] that CSPs for constraint languages over finite domains have a complexity dichotomy: they are either in P or NP-complete. This conjecture remains unsettled, although dichotomy is now known on substantial classes (for example when the domain has at most three elements [30, 19] or when the constraint language contains a single binary relation without sources and sinks [24, 1]). Various methods, combinatorial (graph-theoretic), logical, and universal-algebraic have been brought to bear on this classification project, with many remarkable consequences. A conjectured delineation for the dichotomy was given in the algebraic language in [20].

When the domain is infinite, the complexity of the CSP can be outside NP, and even undecidable [11]. But for natural classes of such CSPs there is often the potential for structured classifications, and this has proved to be the case for structures first-order definable over the order $(\mathbb{Q}, <)$ of the rationals [8] or over the integers with successor [9]. Another classification of this type has been obtained for CSPs where the constraint language is first-order definable over the random (Rado) graph [14], making use of structural Ramsey theory. This paper was titled 'Schaefer's theorem for graphs' and it can be seen as lifting the famous classification of Schaefer [30] from Boolean logic to logic over finite graphs, since the random graph is universal for the class of finite graphs.

1.2 Homogeneous graphs and their reducts

The notion of *homogeneity* from model theory plays an important role when applying techniques from finite-domain constraint satisfaction to constraint satisfaction over infinite domains. A relational structure is *homogeneous* if every isomorphism between finite induced substructures can be extended to an automorphism of the entire structure. Homogeneous structures are uniquely (up to isomorphism) given by the class of finite structures that embed into them. The structure $(\mathbb{Q}, <)$ and the random graph are among the most prominent examples of homogeneous structures. The class of structures that are definable over a homogeneous structure with finite relational signature is a very large generalisation of the class of all finite structures, and CSPs for those structures have been studied independently in many different areas of theoretical computer science, e.g. in temporal and spatial reasoning, phylogenetic analysis, computational linguistics, scheduling, graph homomorphisms, and many more; see [4] for references.

While homogeneous relational structures are abundant, there are remarkably few countably infinite homogeneous (undirected, irreflexive) *graphs*; they have been classified by Lachlan and Woodrow [26]. Besides the random graph mentioned earlier, an example of such a graph is the countable homogeneous *universal triangle-free* graph, one of the fundamental structures that appears in most textbooks in model theory. This graph is the up to isomorphism unique countable triangle-free graph (H_3, E) with the property that for every finite independent set $X \subseteq H_3$ and for every finite set $Y \subseteq H_3$ there exists a vertex $x \in H_3 \setminus (X \cup Y)$ such that x is adjacent to every vertex in X and to no vertex in Y.

Further examples of homogeneous graphs are the graphs (H_3, E), (H_4, E), (H_5, E), ..., called the *Henson graphs*, and their complements. Here, (H_n, E) for $n > 3$ is the generalisation of the graph (H_3, E) above from triangles to cliques of size n. Finally, the list of Lachlan

and Woodrow contains only one more family of infinite graphs, namely the graphs (C_n^s, E) whose reflexive closure Eq is an equivalence relation with n classes of equal size s, where $1 \leq n, s \leq \omega$ and either n or s equals ω, as well as their complements. We remark that (C_n^s, Eq) is itself homogeneous and first-order interdefinable with (C_n^s, E), and so we shall sometimes refer to the *homogeneous equivalence relations*.

All countable homogeneous graphs, and even all structures which are first-order definable over homogeneous graphs, are *ω-categorical*, that is, all countable models of their first-order theory are isomorphic. Moreover, all countably infinite homogeneous graphs Γ are *finitely bounded* in the sense that the *age* of Γ, i.e., the class of finite structures that embed into Γ, can be described by finitely many forbidden substructures. Finitely bounded homogeneous structures also share with finite structures the property of having a finite description: up to isomorphism, they are uniquely given by the finite list of forbidden structures that describes their age. Recent work indicates the importance of finite boundedness for complexity classification [2, 10], and it has been conjectured that all structures with a first-order definition in a finitely bounded homogeneous structure enjoy a complexity dichotomy, i.e., their CSP is either in P or NP-complete (cf. [16, 2]). The structures first-order definable in homogeneous graphs therefore provide the most natural class on which to test further the methods developed in [14] specifically for the random graph.

In this article we obtain a complete classification of the computational complexity of CSPs where all constraints have a first-order definition in one of the Henson graphs. We moreover obtain such a classification for CSPs where all constraints have a first-order definition in a countably infinite homogeneous graph whose reflexive closure is an equivalence relation, expanding earlier results for the special cases of one single equivalence class (so-called *equality constraints* [7]) and infinitely many infinite classes [18]. Together with the above-mentioned result on the random graph, this completes the classification of CSPs for constraints with a first-order definition in any countably infinite homogeneous graph, by Lachlan and Woodrow's classification.

Following an established convention [33, 12], we call a structure with a first-order definition in another structure Δ a *reduct* of Δ. That is, for us a reduct of Δ is as the classical definition of a reduct with the difference that we first allow a first-order expansion of Δ. With this terminology, the present article provides a complexity classification of the CSPs for all reducts of countably infinite homogeneous graphs. In other words, for every such reduct we determine the complexity of deciding its *primitive positive theory*, which consists of all sentences which are existentially quantified conjunctions of atomic formulas and which hold in the reduct. We remark that all reducts of such graphs can be defined by quantifier-free first-order formulas, by homogeneity and ω-categoricity.

For reducts of (H_n, E), the CSPs express computational problems where the task is to decide whether there exists a finite graph without any clique of size n that meets certain constraints. An example of a reduct whose CSP can be solved in polynomial time is $(H_n, \neq, \{(x, y, u, v) : E(x, y) \Rightarrow E(u, v)\})$, where $n \geq 3$ is arbitrary. As it turns out, for every CSP of a reduct of a Henson graph which is solvable in polynomial time, the corresponding reduct over the Rado graph, i.e., the reduct whose relations are defined by the same quantifier-free formulas, is also polynomial-time solvable. On the other hand, the CSP of the reduct $(H_n, \{(x, y, u, v) : E(x, y) \vee E(u, v)\})$ is NP-complete for all $n \geq 3$, but the corresponding reduct over the random graph can be decided in polynomial time.

Similarly, for reducts of the graph (C_n^s, E) whose reflexive closure is an equivalence relation with n classes of size s, where $1 \leq n, s \leq \omega$, the computational problem is to decide whether there exists an equivalence relation with n classes of size s that meets certain constraints.

1.3 Results

Our first result is the complexity classification of the CSPs of all reducts of Henson graphs, showing in particular that a uniform approach to infinitely many "base structures" (namely, the n-th Henson graph for each $n \geq 3$) is possible.

▶ **Theorem 1.** *Let $n \geq 3$, and let Γ be a finite signature reduct of the n-th Henson graph (H_n, E). Then* CSP(Γ) *is either in* P *or* NP*-complete.*

We then obtain a similar complexity dichotomy for reducts of homogeneous equivalence relations, expanding earlier results for special cases [18, 7].

▶ **Theorem 2.** *Let (C_n^s, E) be an infinite graph whose reflexive closure Eq is an equivalence relation with n classes of size s, where $1 \leq n, s \leq \omega$. Then for any finite signature reduct Γ of (C_n^s, E), the problem* CSP(Γ) *is either in* P *or* NP*-complete.*

Together with the classification of countable homogeneous graphs, and the fact that the complexity of the CSPs of the reducts of the Rado graph have been classified [14], this completes the CSP classification of reducts of all countably infinite homogeneous graphs, confirming further instances of the open conjecture that CSPs of reducts of finitely bounded homogeneous structures are either in P or NP-complete [16, 2].

▶ **Corollary 3.** *Let Γ be a finite signature reduct of a countably infinite homogeneous graph. Then* CSP(Γ) *is either in* P *or* NP*-complete.*

1.4 The strategy

The method we employ follows to a large extent the method invented in [14] for the corresponding classification problem where the 'base structure' is the random graph. The key component of this method is the usage of Ramsey theory (in our case, a result of Nešetřil and Rödl [27]) and the concept of *canonical functions* introduced in [13]. There are, however, some interesting differences and novelties that appear in the present proof, as we now shortly outline.

1.4.1 Henson graphs

When studying the proofs in [14], one might get the impression that the complexity of the method grows with the model-theoretic complexity of the base structure, and that for the random graph we have really reached the limits of bearableness for applying the Ramsey method.

However, quite surprisingly, when we step from the random graph to the graphs (H_n, E), which are in a sense more complicated structures from a model-theoretic point of view[1], the classification and its proof become easier again. It is one of the contributions of the present article to explain the reasons behind this effect. Essentially, certain *behaviours* of canonical functions existing on the random graph can not be realised in (H_n, E). For example the canonical polymorphisms of behaviour "max" (cf. preliminaries) play no role for the present classification, but account over the random graph for the tractability of, inter alia, the 4-ary relation defined by the formula $E(x, y) \vee E(u, v)$.

[1] For example, the random graph has a *simple* theory [32], whereas the Henson graphs are the most basic examples of structures whose theory is *not* simple.

Interestingly, we are able to reuse results about canonical functions over the random graph, since the calculus for composing behaviours of canonical functions is the same for any other structure with the same type space, and in particular the Henson graphs. Via this meta-argument we can, on numerous occasions, make statements about canonical functions over the Henson graphs which were proven earlier for the Rado graph, ignoring completely the actual underlying structure; even more comfortably, we can *a posteriori* rule out some possibilities in those statements because of the K_n-freeness of the Henson graphs. Examples of this phenomenon appear in Lemma 14.

On the other hand, along with these simplifications, there are also new additional difficulties that appear when investigating reducts of (H_n, E) and that were not present in the classification of reducts of the random graph, which basically stem from the lower degree of symmetry of (H_n, E) compared to the Rado graph. For example, in expansions of Henson graphs by finitely many constants, not all orbits induce copies of Henson graphs; the fact that the analogous statement does hold for the Rado graph was used extensively in [14].

1.4.2 Equivalence relations

Similarly to the situation for the equivalence relation with infinitely many infinite classes studied in [18], there are two interesting sources of NP-hardness for the reducts Γ of other homogeneous equivalence relations: namely, if the equivalence relation is invariant under the polymorphisms of Γ, then the structure obtained from Γ by factoring by the equivalence relation might have a NP-hard CSP, implying NP-hardness for the CSP of Γ itself; or, roughly, for a fixed equivalence class the restriction of Γ to that class might have a NP-hard CSP, again implying NP-hardness of the CSP of Γ (assuming that Γ is a *model-complete core*, see Sections 3 and 6). But whereas for the equivalence relation with infinitely many infinite classes both the factor structure and the restriction to a class are again infinite structures, for the other homogeneous equivalence relations one of the two is a finite structure, obliging us to combine results about CSPs of finite structures with those of infinite structures. As it turns out, the two-element case is, not surprisingly, different from the other finite cases and, quite surprisingly, significantly more involved than the other cases.

2 Preliminaries

Let E denote the irreflexive edge relation of our respective graphs and N its irreflexive complement. The following lemma has been first stated in [25] for finite domain structures Γ only, but the proof there also works for arbitrary infinite structures.

▶ **Lemma 4.** *Let $\Gamma = (D, R_1, \ldots, R_\ell)$ be a relational structure, and let R be a relation that has a primitive positive (pp) definition in Γ. Then* $\mathrm{CSP}(\Gamma)$ *and* $\mathrm{CSP}(D, R, R_1, \ldots, R_\ell)$ *are polynomial-time equivalent.*

▶ **Theorem 5** (from [11])**.** *Let Γ be a countable ω-categorical structure. Then the relations preserved by the polymorphisms* $\mathrm{Pol}(\Gamma)$ *are precisely those having a primitive positive definition in Γ.*

These facts make it possible to apply a *universal algebraic approach*, and classify the complexity of reducts of an ω-categorical structure through understanding the polymorphism clones of these reducts. In fact, we can state our results in terms of the polymorphism clones, see Theorems 21 and 37. Roughly speaking, we will conclude that if Γ is a reduct of a homogeneous graph with a finite relational language, then $\mathrm{CSP}(\Gamma)$ is NP-complete iff for

some finite tuple c in Γ, the clone $\mathrm{Pol}(\Gamma, c)$ maps to the clone of projections via a continuous homomorphism.

If such a homomorphism does not exist, it implies the existence of functions in $\mathrm{Pol}(\Gamma)$, which satisfy an equation that prevents them from being mapped to projections with a homomorphism. The idea then is to find patterns in the behaviour of these functions, and show that they must generate one out of a given finite number of well-behaved functions; those appearing in Theorems 21 and 37, and whose incidence in $\mathrm{Pol}(\Gamma)$ automatically make $\mathrm{CSP}(\Gamma)$ solvable in polynomial time. These well-behaved functions in $\mathrm{Pol}(\Gamma)$, called *canonical functions*, are obtained by the method mentioned in Section 1.4 using Ramsey theory.

▶ **Definition 6.** Let Δ be a structure. The *type* $\mathrm{tp}(a)$ of an n-tuple a in Δ is the set of first-order formulas with free variables $x_1, \ldots, x_n$ that hold for a in Δ. For structures $\Delta_1, \ldots, \Delta_k$ and tuples $a^1, \ldots, a^n \in \Delta_1 \times \cdots \times \Delta_k$, the type $\mathrm{tp}(a^1, \ldots, a^n)$ of $(a^1, \ldots, a^n)$ in $\Delta_1 \times \cdots \times \Delta_k$ is the k-tuple containing the types of $(a_i^1, \ldots, a_i^n)$ in Δ_i for each $1 \leq i \leq k$.

It is well-known that in homogeneous structures such as (H_n, E) and (C_n^k, E), two n-tuples have the same type if and only if they are in the same orbit of the automorphism group.

▶ **Definition 7.** Let $\Delta_1, \ldots, \Delta_k$ and Λ be structures. A *behaviour* B between $\Delta_1, \ldots, \Delta_k$ and Λ is a partial function from the types over $\Delta_1, \ldots, \Delta_k$ to the types over Λ. Pairs (s, t) with $B(s) = t$ are also called *type conditions*. We say that a function $f\colon \Delta_1 \times \cdots \times \Delta_k \to \Lambda$ *satisfies the behaviour* B if whenever $B(s) = t$ and $(a^1, \ldots, a^n)$ has type s in $\Delta_1, \ldots, \Delta_k$, then the n-tuple $(f(a_1^1, \ldots, a_k^1), \ldots, f(a_1^n, \ldots, a_k^n))$ has type t in Λ. A function $f\colon \Delta_1 \times \cdots \times \Delta_k \to \Lambda$ is *canonical* if it satisfies a behaviour which is a total function from the types over $\Delta_1, \ldots, \Delta_k$ to the types over Λ.

To provide immediate examples for these notions, we now define some behaviours that will appear in our proof as well as in the precise CSP classification. For m-ary relations $R_1, \ldots, R_k$ over a set D, we will in the following write $R_1 \cdots R_k$ for the m-ary relation on D^k that holds between k-tuples $x^1, \ldots, x^m \in D^k$ iff $R_i(x_i^1, \ldots, x_i^m)$ holds for all $1 \leq i \leq k$.

▶ **Definition 8.** Given a homogeneous graph (G, E) we say that a binary injective operation $f\colon G^2 \to G$ is

- *balanced in the first argument* if for all $u, v \in G^2$ we have that $\mathrm{E}{=}(u, v)$ implies $E(f(u), f(v))$ and $\mathrm{N}{=}(u, v)$ implies $N(f(u), f(v))$;
- *E-dominated (N-dominated) in the first argument* if for all $u, v \in G^2$ with $\neq{=}(u, v)$ we have that $E(f(u), f(v))$ $(N(f(u), f(v)))$;
- *balanced/E-dominated/N-dominated in the second argument* if $(x, y) \mapsto f(y, x)$ is balanced/E-dominated/N-dominated in the first argument;
- *balanced/E-dominated/N-dominated* if f is balanced/E-dominated/N-dominated in both arguments, and *unbalanced* if f is not balanced;
- *of behaviour p_1* if for all $u, v \in G^2$ with $\neq{\neq}(u, v)$ we have $E(f(u), f(v))$ iff $E(u_1, v_1)$;
- *of behaviour p_2* if $(x, y) \mapsto f(y, x)$ is of behaviour p_1;
- *of behaviour projection* if it is of behaviour p_1 or p_2;
- *of behaviour* min if for all $u, v \in G^2$ with $\neq{\neq}(u, v)$ we have $E(f(u), f(v))$ iff $\mathrm{EE}(u, v)$.

A ternary canonical injection $f\colon G^3 \to G$ is

- *hyperplanely of behaviour projection* iff the functions $(u, v) \mapsto f(c, u, v)$, $(u, v) \mapsto f(u, c, v)$, and $(u, v) \mapsto f(u, v, c)$ are of behaviour projection for all $c \in G$. Similarly other hyperplane behaviours, such as hyperplanely E-dominated, are defined.
- of behaviour *minority* if for all $u, v \in G^3$ with $\neq{\neq}{\neq}(u, v)$ we have $E(f(u), f(v))$ if and only if $\mathrm{EEE}(u, v)$, $\mathrm{NNE}(u, v)$, $\mathrm{NEN}(u, v)$, or $\mathrm{ENN}(u, v)$.

2.1 Overview

This article is organised as follows. Basic notions and definitions, as well as the fundamental facts of the method we are going to use, are deferred for reasons of space to the appendix. Our notation and definitions may also be found in [14].

Sections 3 to 5 deal with the Henson graphs: Section 3 is complexity-free and investigates the structure of reducts of Henson graphs via polymorphisms and Ramsey theory. In Section 4, we provide hardness results for different classes of reducts. In Section 5 we rephrase Theorem 1, and we discuss the complexity classification in more detail, formulating in particular a tractability criterion for CSPs of reducts of Henson graphs.

Section 6 investigates the structure of reducts of homogeneous equivalence relations via polymorphisms and Ramsey theory and describes the polynomial-time cases.

3 Polymorphisms over Henson graphs

We investigate polymorphisms of reducts of (H_n, E). We start with unary polymorphisms in Section 3.1, obtaining that we can assume that the relations E and N are pp-definable in our reducts. We then turn to binary polymorphisms in Section 3.2, obtaining Proposition 15 telling us that we may further assume the existence of a binary injective polymorphism. Building on the results of those sections, we show in Section 3.3 via an analysis of ternary polymorphisms that for any reduct which pp-defines the relations E and N, either the polymorphisms preserve a certain relation H, or there is a canonical polymorphism of behaviour min (Proposition 17).

3.1 The unary case: model-complete cores

A countable ω-categorical structure Δ is called a *model-complete core* if $\mathrm{Aut}(\Delta)$ is dense in $\mathrm{End}(\Delta)$, or equivalently, every endomorphism of Δ is an elementary self-embedding, i.e., preserves all first-order formulas over Δ. Every countable ω-categorical structure Γ is *homomorphically equivalent* to an up to isomorphism unique ω-categorical model-complete core Δ, that is, there exists homomorphisms from Γ into Δ and vice-versa [3]. Since the CSPs of homomorphically equivalent structures are equal, it has proven fruitful in classification projects to always work with model-complete cores. The following proposition essentially calculates the model-complete cores of the reducts of Henson graphs.

▶ Proposition 9. Let Γ be a reduct of (H_n, E). Then either $\mathrm{End}(\Gamma)$ contains a function whose image induces an independent set, or $\mathrm{End}(\Gamma) = \overline{\mathrm{Aut}(\Gamma)} = \overline{\mathrm{Aut}(H_n, E)}$.

In the first case of Proposition 9, the model-complete core of the reduct is in fact a reduct of equality. Since the CSPs of reducts of equality have been classified [7], we do not have to consider any further reducts with an endomorphism whose image induces an independent set.

▶ **Lemma 10.** *Let Γ be a reduct of (H_n, E), and assume that $\mathrm{End}(\Gamma)$ contains a function whose image is an independent set. Then Γ is homomorphically equivalent to a reduct of $(H_n, =)$.*

In the second case of Proposition 9, it turns out that all polymorphisms preserve the relations E, N, and $\neq$, by the following lemma and Theorem 5.

▶ **Lemma 11.** *Let Γ be such that $\mathrm{End}(\Gamma) = \overline{\mathrm{Aut}(H_n, E)}$. Then E, N, and $\neq$ have primitive positive definitions in Γ.*

Before moving on to binary polymorphisms, we observe the following corollary of Proposition 9, first mentioned in [33].

▶ **Corollary 12.** *For every $n \geq 3$, the permutation group $\mathrm{Aut}(H_n, E)$ is a maximal closed subgroup of the full symmetric group on H_n, i.e., every closed subgroup of the full symmetric group containing $\mathrm{Aut}(H_n, E)$ either equals $\mathrm{Aut}(H_n, E)$ or the full symmetric group.*

3.2 Binary polymorphisms

We investigate binary functions preserving E, N, and $\neq$. A finitary operation $f(x_1, \ldots, x_n)$ on a set is *essential* if it does not depend on only one of its arguments x_i.

▶ **Lemma 13.** *Every essential function $f\colon H_n^k \to H_n$ that preserves E, N, and $\neq$ generates a binary injection.*

By Proposition 9, the remark thereafter, Lemma 11 and Lemma 13, we may assume that $\mathrm{Pol}(\Gamma)$ contains a binary injection f, as otherwise the complexity of $\mathrm{CSP}(\Gamma)$ is known. After an analysis of the possible behaviours of f, we can make further assumptions on the binary injection in $\mathrm{Pol}(\Gamma)$.

▶ **Lemma 14.** *Let $f\colon H_n^k \to H_n$ be an essential function that preserves E, N, and $\neq$. Then f generates one of the following binary canonical injections: of behaviour* min *and N-dominated; or of behaviour p_1, balanced in the first, and N-dominated in the second argument.*

We conclude this section by summarising the results we have so far.

▶ Proposition 15. Let Γ be a reduct of (H_n, E), where $n \geq 3$. Then either

(1) Γ is homomorphically equivalent to a reduct of $(H_n, =)$, or
(2) Γ pp-defines E, N, and $\neq$.

In the latter case we have that either

(2a) every function in $\mathrm{Pol}(\Gamma)$ is essentially unary, or
(2b) $\mathrm{Pol}(\Gamma)$ contains one of the two binary canonical injections of Lemma 14.

Note that if item (1) holds then $\mathrm{CSP}(\Gamma)$ is either in P or NP-complete [7], and if item (2a) holds then $\mathrm{CSP}(\Gamma)$ is NP-complete (Theorem 10 in [5]). In case (2b), when $\mathrm{Pol}(\Gamma)$ contains a binary canonical injection of behaviour min which is N-dominated then $\mathrm{CSP}(\Gamma)$ is in P, as we will discuss later. It thus remains to further consider the second case of Lemma 14, which we do in the next subsection.

3.3 The relation H

We investigate Case (2b) of Proposition 15. The following relation characterises the NP-complete cases in this situation.

▶ **Definition 16.** We define a 6-ary relation $H(x_1, y_1, x_2, y_2, x_3, y_3)$ on H_n by

$$\begin{aligned}
&\bigwedge_{i,j\in\{1,2,3\},i\neq j,u\in\{x_i,y_i\},v\in\{x_j,y_j\}} N(u,v) \\
&\wedge \big((E(x_1,y_1) \wedge N(x_2,y_2) \wedge N(x_3,y_3)) \\
&\quad \vee\ (N(x_1,y_1) \wedge E(x_2,y_2) \wedge N(x_3,y_3)) \\
&\quad \vee\ (N(x_1,y_1) \wedge N(x_2,y_2) \wedge E(x_3,y_3))\big)\ .
\end{aligned}$$

The importance of the relation H is reflected in the following proposition, which states that if Γ is a reduct of (H_n, E) with E and N primitive positive definable in Γ, then either H has a primitive positive definition in Γ, in which case CSP(Γ) is NP-complete, or Pol(Γ) has a certain canonical polymorphism which will imply tractability of the CSP. NP-completeness and tractability for those cases will be discussed in Sections 4 and 5.

▶ Proposition 17. Let Γ be a reduct of (H_n, E) with E and N primitive positive definable in Γ. Then at least one of the following holds:

(a) There is a primitive positive definition of H in Γ.
(b) Pol(Γ) contains a canonical binary injection of behaviour min.

4 CSPs over Henson graphs

We now explain why any reduct of (H_n, E) which has H among its relations, and hence by Lemma 4 every reduct which pp-defines H, has an NP-hard CSP. While it would be possible to show NP-hardness of CSP(H_n, H) directly by reduction of, say, the NP-hard problem positive 1-in-3-SAT, we will use results from [15], and in fact a recent strengthening thereof from [2], to prove hardness more elegantly via a structural property of Pol(H_n, H).

▶ **Definition 18.** Let Γ be a structure. A *projective clone homomorphism* of Γ (or Pol(Γ)) is a mapping from Pol(Γ) onto its projections which: preserves arities; fixes each projection; and preserves composition.

A *projective strong h1 clone homomorphism* of Γ is a mapping as above, where the third condition is weakened to preservation of composition with projections.

▶ **Theorem 19** (from [2]). *Let Γ be a countable ω-categorical structure in a finite relational language which has a uniformly continuous strong h1 clone homomorphism. Then* CSP(Γ) *is NP-hard.*

▶ Proposition 20. The structure (H_n, H) has a uniformly continuous strong h1 clone homomorphism. Consequently, CSP(H_n, H) is NP-hard.

5 Summary for the Henson graphs

We can restate Theorem 1 in a more detailed fashion as follows.

▶ **Theorem 21.** *Let Γ be a reduct of a Henson graph (H_n, E). Then one of the following holds.*

(1) *Γ has an endomorphism whose image induces an independent set, and is homomorphically equivalent to a reduct of $(H_n, =)$.*
(2) Pol(Γ) *has a uniformly continuous projective clone homomorphism.*
(3) Pol(Γ) *contains a binary canonical injection which is of behaviour* min *and N-dominated.*

Items (2) and (3) cannot simultaneously hold, and when Γ has a finite relational signature, then (2) *implies NP-completeness and (3) implies tractability of its CSP.*

The first statement of Theorem 21 follows directly from the proof of Theorem 1, with the additional observation that the strong h1 clone homomorphism defined in Proposition 20 is in fact a clone homomorphism. When (3) holds for a reduct, then (2) cannot hold, because (3) implies the existence of $f(x, y) \in$ Pol(Γ) and $\alpha \in \overline{\mathrm{Aut}(\Gamma)}$ such that $f(x, y) = \alpha f(y, x)$ holds, and equation impossible to satisfy by projections. In fact, by further analysing case (1), one can easily show that it also implies either (2) or (3), so that we have the following.

▶ **Corollary 22.** *For every reduct* Γ *of a Henson graph* (H_n, E)*, precisely one of the following holds:* $\mathrm{Pol}(\Gamma)$ *has a uniformly continuous projective clone homomorphism; or* $\mathrm{Pol}(\Gamma)$ *contains* $f(x,y) \in \mathrm{Pol}(\Gamma)$ *and* $\alpha \in \overline{\mathrm{Aut}(\Gamma)}$ *such that* $f(x,y) = \alpha f(y,x)$*.*

When Γ *has a finite relational signature, then the first case possesses NP-completeness and the second case the tractability of its CSP.*

6 Reducts of homogeneous equivalence relations

We now investigate polymorphisms of reducts of the graphs (C_n^s, E), for $2 \leq n, s \leq \omega$, with precisely one of n, s equal to ω. Recall from the preliminaries that we write Eq for the reflexive closure of E.

Similarly to the case of the Henson graphs, we start with unary polymorphisms in Section 6.1, reducing the problem to model-complete cores.

We then turn to higher-arity polymorphisms; here, the organisation somewhat differs from the case of the Henson graphs. The role of the NP-hard relation H from the Henson graphs is now taken by the two sources of NP-hardness mentioned in the introduction: the first source being that factoring by the equivalence relation Eq yields a structure with an NP-hard problem, and the second source being that restriction to some equivalence class yields a structure with an NP-hard problem. In Section 6.2, we show that in fact, one of the two sources always applies for model-complete cores when $2 < n < \omega$ or $2 < s < \omega$. Consequently, only the higher-arity polymorphisms of the reducts of (C_2^ω, E) and (C_ω^2, E) require deeper investigation using Ramsey theory; this will be dealt with in Sections 6.3 and 6.4, respectively.

6.1 The unary case: model-complete cores

▶ Proposition 23. Let Γ be a reduct of (C_n^s, E), where $1 \leq n, s \leq \omega$, and at least one of n, s equals ω. Then $\mathrm{End}(\Gamma) = \overline{\mathrm{Aut}(\Gamma)} = \overline{\mathrm{Aut}(C_n^s, E)}$, or $\mathrm{End}(\Gamma)$ contains an endomorphism onto a clique or an independent set.

In the following sections, we investigate essential polymorphisms of reducts Γ of (C_n^s, E) which are model-complete cores, i.e., $\mathrm{End}(\Gamma) = \overline{\mathrm{Aut}(C_n^s, E)}$. The following proposition implies that in that situation, the equivalence relation Eq is invariant under $\mathrm{Pol}(\Gamma)$.

▶ Proposition 24. Let Γ be a reduct of (C_n^s, E), where $1 \leq n, s \leq \omega$. If $\mathrm{End}(\Gamma) = \overline{\mathrm{Aut}(C_n^s, E)}$, then E, N and Eq are preserved by the polymorphisms of Γ.

Therefore, in the above situation Eq is an equivalence relation which is invariant under $\mathrm{Pol}(\Gamma)$, and so $\mathrm{Pol}(\Gamma)$ acts naturally on the equivalence classes of Eq. Moreover, if we fix any $c \in C_n^s$ and expand the structure Γ by the constant c, then the equivalence class C of c has a primitive positive definition in that expansion (Γ, c), since Eq and c do. Hence, C is invariant under $\mathrm{Pol}(\Gamma, c)$, and so $\mathrm{Pol}(\Gamma, c)$ acts naturally on C via restriction. In the following sections, we analyse these actions.

6.2 The case $2 < n < \omega$ or $2 < s < \omega$

It turns out that in these cases, one of the sources of hardness always applies. We will use the following fact about function clones on a finite domain.

▶ Proposition 25 (from [23]). Every function clone on a finite domain of at least three elements which contains all permutations as well as an essential function contains a unary constant function.

▶ Proposition 26. Let Γ be a reduct of (C_n^ω, E), where $2 < n < \omega$, such that $\mathrm{End}(\Gamma) = \overline{\mathrm{Aut}(C_n^\omega, E)}$. Then the action of $\mathrm{Pol}(\Gamma)$ on the equivalence classes of Eq has no essential and no constant operation.

▶ Proposition 27. Let Γ be a reduct of (C_ω^s, E), where $2 < s < \omega$, such that $\mathrm{End}(\Gamma) = \overline{\mathrm{Aut}(C_\omega^s, E)}$. Then for any $c \in C_\omega^s$, the action of $\mathrm{Pol}(\Gamma, c)$ on the equivalence class of c has no essential and no constant operation.

6.3 The case of two infinite classes: $n = 2$ and $s = \omega$

The following proposition states that either one of the two sources of hardness applies, or $\mathrm{Pol}(\Gamma)$ contains a ternary canonical function with a certain behaviour.

▶ Proposition 28. Let Γ be a reduct of (C_2^ω, E) such that $\mathrm{End}(\Gamma) = \overline{\mathrm{Aut}(C_2^\omega, E)}$. Then one of the following holds:

- the action of $\mathrm{Pol}(\Gamma)$ on the classes of Eq has no essential function;
- the action of $\mathrm{Pol}(\Gamma, c)$ on the equivalence class of c has no essential function, for some $c \in C_2^\omega$;
- $\mathrm{Pol}(\Gamma)$ contains a canonical ternary injection of behaviour minority which is hyperplanely of behaviour E-dominated projection.

To prove the proposition, we need to recall a special case of Post's classical result about function clones acting on a two-element set, as well as a result on function clones on a countable set which contain all permutations. Comparing this statement with Proposition 25 sheds light on why the case of this section is more involved than the cases of the preceding section.

▶ Proposition 29 (Post [29]). Every function clone with domain $\{0, 1\}$ containing both permutations of $\{0, 1\}$ as well as an essential function contains a unary constant operation or the ternary addition modulo 2.

▶ Proposition 30 (from [7]). Every closed function clone on a countably infinite set which contains all permutations as well as an essential operation contains a binary injection.

▶ Proposition 31. Let Γ be preserved by a ternary injection h of behaviour minority which is hyperplanely an E-dominated projection. Then $\mathrm{CSP}(\Gamma)$ can be solved in polynomial time.

6.4 The case of infinitely many classes of size two: $n = \omega$ and $s = 2$

As in the preceding section, we show that either one of the two sources of hardness applies, or $\mathrm{Pol}(\Gamma)$ contains a ternary canonical function of a certain behaviour.

▶ Proposition 32. Let Γ be a reduct of (C_ω^2, E) such that $\mathrm{End}(\Gamma) = \overline{\mathrm{Aut}(C_\omega^2, E)}$. Then one of the following holds:

- the action of $\mathrm{Pol}(\Gamma)$ on the classes of Eq has no essential function;
- the action of $\mathrm{Pol}(\Gamma, c)$ on the equivalence class of c has no essential function, for some $c \in C_\omega^2$;
- $\mathrm{Pol}(\Gamma)$ contains a ternary canonical function h with $h(N, \cdot, \cdot) = h(\cdot, N, \cdot) = h(\cdot, \cdot, N) = N$ and which behaves like a minority on $\{E, =\}$.

To prove the proposition, we are again going to make use of Propositions 29 and 30, and the following lemma.

▶ **Lemma 33.** *Let Γ be a reduct of (C_ω^2, E) such that $\mathrm{End}(\Gamma) = \overline{\mathrm{Aut}(C_\omega^2, E)}$. If $\mathrm{Pol}(\Gamma)$ contains a ternary function which behaves like $x + y + z$ modulo 2 on some equivalence class,*

then it contains a ternary function which behaves like $x + y + z$ modulo 2 on all equivalence classes.

Let Γ be a reduct of (C^2_ω, Eq) where Eq is an equivalence relation with infinitely many classes of size two such that $\mathrm{Pol}(\Gamma)$ contains a ternary canonical function h as in item 3 of Proposition 32.

▶ Proposition 34. A relation with a first-order definition in (C^2_ω, Eq) is preserved by h if and only if it can be defined by a conjunction of formulas of the form

$$N(x_1, y_1) \vee \cdots \vee N(x_k, y_k) \vee Eq(z_1, z_2) \tag{1}$$

for $k \geq 0$, or of the form

$$N(x_1, y_1) \vee \cdots \vee N(x_k, y_k) \vee (|\{i \in S : x_i \neq y_i\}| \equiv_2 p) \tag{2}$$

where $p \in \{0, 1\}$ and $S \subseteq \{1, \ldots, k\}$.

▶ Proposition 35. There is a polynomial-time algorithm that decides whether a given set Φ of formulas as in the statement of Proposition 34 is satisfiable.

▶ **Corollary 36.** *Let Γ be a reduct of (C^2_ω, Eq) with finite signature and such that $\mathrm{Pol}(\Gamma)$ contains the operation h. Then $\mathrm{CSP}(\Gamma)$ is in P.*

We close the section with a more detailed variant of Theorem 2.

▶ **Theorem 37.** *Let (C^s_n, E) be an infinite graph whose reflexive closure Eq is an equivalence relation with n classes of size s, where $1 \leq n, s \leq \omega$. Let Γ be a reduct of (C^s_n, E). Then one of the following holds.*

(1) Γ has an endomorphism whose image induces a clique or an independent set, and is homomorphically equivalent to a reduct of $(C^s_n, =)$.

(2) Γ is a model complete core and $\mathrm{Pol}(\Gamma, c)$ has a uniformly continuous projective clone homomorphism for some $c \in (C^s_n, E)$.

(3) $n = 2, s = \omega$, Γ is a model complete core, and $\mathrm{Pol}(\Gamma)$ contains a canonical ternary injection of behaviour minority which is hyperplanely of behaviour E-dominated projection.

(4) $n = \omega, s = 2$, Γ is a model complete core, and $\mathrm{Pol}(\Gamma)$ contains a ternary canonical function h with $h(N, \cdot, \cdot) = h(\cdot, N, \cdot) = h(\cdot, \cdot, N) = N$ and which behaves like a minority on $\{E, =\}$.

Neither items (2) and (3), nor items (2) and (4) can simultaneously hold, and when Γ has a finite relational signature, then (2) implies NP-completeness and both (3) and (4) imply tractability of its CSP.

Outlook

We have classified the computational complexity of CSPs for reducts of the infinite homogeneous graphs. Our proof shows that the scope of the classification method from [14] is much larger than one might expect at first sight. The general research goal here is to identify larger and larger classes of infinite-domain CSPs where systematic complexity classification is possible; a general dichotomy conjecture is open for CSPs of reducts of finitely bounded homogeneous structures [16, 2]. The next step in this direction might be to show a general complexity dichotomy for reducts of homogeneous structures whose age is finitely bounded and has the *free amalgamation property* (the Henson graphs provide natural examples for such structures). The present paper indicates that this problem might be within reach.

References

1 Libor Barto, Marcin Kozik, and Todd Niven. 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), 2009.

2 Libor Barto, Jakub Opršal, and Michael Pinsker. The wonderland of reflections. Preprint arXiv:1510.04521, 2015.

3 Manuel Bodirsky. Cores of countably categorical structures. *Logical Methods in Computer Science*, 3(1):1–16, 2007.

4 Manuel Bodirsky. Complexity classification in infinite-domain constraint satisfaction. Mémoire d'habilitation à diriger des recherches, Université Diderot – Paris 7. Available at arXiv:1201.0856, 2012.

5 Manuel Bodirsky, Hubie Chen, Jan Kára, and Timo von Oertzen. Maximal infinite-valued constraint languages. *Theoretical Computer Science (TCS)*, 410:1684–1693, 2009. A preliminary version appeared at ICALP'07.

6 Manuel Bodirsky, Peter Jonsson, and Trung Van Pham. The complexity of phylogeny constraint satisfaction. In *Proceedings of the Symposium on Theoretical Aspects of Computer Science (STACS)*, 2016. Preprint arXiv:1503.07310.

7 Manuel Bodirsky and Jan Kára. The complexity of equality constraint languages. *Theory of Computing Systems*, 3(2):136–158, 2008. A conference version appeared in the proceedings of Computer Science Russia (CSR'06).

8 Manuel Bodirsky and Jan Kára. The complexity of temporal constraint satisfaction problems. *Journal of the ACM*, 57(2):1–41, 2009. An extended abstract appeared in the Proceedings of the Symposium on Theory of Computing (STOC'08).

9 Manuel Bodirsky, Barnaby Martin, and Antoine Mottet. Constraint satisfaction problems over the integers with successor. In *Automata, Languages, and Programming - 42nd International Colloquium, ICALP 2015, Kyoto, Japan, July 6-10, 2015, Proceedings, Part I*, pages 256–267, 2015.

10 Manuel Bodirsky and Antoine Mottet. Reducts of finitely bounded homogeneous structures, and lifting tractability from finite-domain constraint satisfaction. Submitted. Preprint available under ArXiv:1601.04520, 2016.

11 Manuel Bodirsky and Jaroslav Nešetřil. Constraint satisfaction with countable homogeneous templates. *Journal of Logic and Computation*, 16(3):359–373, 2006.

12 Manuel Bodirsky and Michael Pinsker. Reducts of Ramsey structures. *AMS Contemporary Mathematics, vol. 558 (Model Theoretic Methods in Finite Combinatorics)*, pages 489–519, 2011.

13 Manuel Bodirsky and Michael Pinsker. Minimal functions on the random graph. *Israel Journal of Mathematics*, 200(1):251–296, 2014.

14 Manuel Bodirsky and Michael Pinsker. Schaefer's theorem for graphs. *Journal of the ACM*, 62(3):Article no. 19, 1–52, 2015. A conference version appeared in the Proceedings of STOC 2011, pages 655–664.

15 Manuel Bodirsky and Michael Pinsker. Topological Birkhoff. *Transactions of the American Mathematical Society*, 367:2527–2549, 2015.

16 Manuel Bodirsky, Michael Pinsker, and András Pongrácz. Projective clone homomorphisms. Preprint arXiv:1409.4601, 2014.

17 Manuel Bodirsky, Michael Pinsker, and Todor Tsankov. Decidability of definability. *Journal of Symbolic Logic*, 78(4):1036–1054, 2013. A conference version appeared in the Proceedings of LICS 2011, pages 321–328.

18 Manuel Bodirsky and Michał Wrona. Equivalence constraint satisfaction problems. In *Proceedings of Computer Science Logic*, volume 16 of *LIPICS*, pages 122–136. Dagstuhl Publishing, September 2012.

19 Andrei A. Bulatov. A dichotomy theorem for constraint satisfaction problems on a 3-element set. *Journal of the ACM*, 53(1):66–120, 2006.

20 Andrei A. Bulatov, Andrei A. Krokhin, and Peter G. Jeavons. Classifying the complexity of constraints using finite algebras. *SIAM Journal on Computing*, 34:720–742, 2005.

21 Tomás Feder and Moshe Y. Vardi. 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.

22 Martin Goldstern and Michael Pinsker. A survey of clones on infinite sets. *Algebra Universalis*, 59:365–403, 2008.

23 L. Haddad and Ivo G. Rosenberg. Finite clones containing all permutations. *Canadian Journal of Mathematics*, 46(5):951–970, 1994.

24 Pavol Hell and Jaroslav Nešetřil. On the complexity of H-coloring. *Journal of Combinatorial Theory, Series B*, 48:92–110, 1990.

25 Peter Jeavons, David Cohen, and Marc Gyssens. Closure properties of constraints. *Journal of the ACM*, 44(4):527–548, 1997.

26 Alistair H. Lachlan and Robert E. Woodrow. Countable ultrahomogeneous undirected graphs. *Transactions of the AMS*, 262(1):51–94, 1980.

27 Jaroslav Nešetřil and Vojtěch Rödl. The partite construction and Ramsey set systems. *Discrete Mathematics*, 75(1-3):327–334, 1989.

28 András Pongrácz. Reducts of the Henson graphs with a constant. Preprint, 2011.

29 Emil L. Post. The two-valued iterative systems of mathematical logic. *Annals of Mathematics Studies*, 5, 1941.

30 Thomas J. Schaefer. The complexity of satisfiability problems. In *Proceedings of the Symposium on Theory of Computing (STOC)*, pages 216–226, 1978.

31 Ágnes Szendrei. *Clones in universal algebra*. Séminaire de Mathématiques Supérieures. Les Presses de l'Université de Montréal, 1986.

32 Katrin Tent and Martin Ziegler. *A course in model theory*. Lecture Notes in Logic. Cambridge University Press, 2012.

33 Simon Thomas. Reducts of the random graph. *Journal of Symbolic Logic*, 56(1):176–181, 1991.