

The Core of a Countably Categorical Structure

Manuel Bodirsky

Humboldt-Universität zu Berlin
bodirsky@informatik.hu-berlin.de

An extended abstract of this article is published
in the proceedings of STACS'05, LNCS 3404, Springer Verlag.

Abstract. A relational structure is a *core*, if all its endomorphisms are embeddings. This notion is important for the classification for the computational complexity of constraint satisfaction problems. It is a fundamental fact that every finite structure S has a core, i.e., S has an endomorphism e such that the structure induced by $e(S)$ is a core; moreover, the core is unique up to isomorphism.

We prove that this result remains valid for ω -categorical structures, and prove that every ω -categorical structure has a core, which is unique up to isomorphism, and which is finite or ω -categorical. We also show that the core of an ω -categorical structure Γ is model complete, and therefore $\forall\exists$ -axiomatizable. If Γ contains all primitive positive definable relations, then the core of Γ admits quantifier elimination. We discuss consequences for constraint satisfaction with ω -categorical templates.

1 Introduction

Let Γ and Δ be relational structures with the same relational signature τ . A mapping $f : \Gamma \rightarrow \Delta$ is called a *homomorphism*, if for all relations $R \in \tau$ and $x_1, \dots, x_n \in \Gamma$ the relation $R(f(x_1), \dots, f(x_n))$ holds in Δ whenever $R(x_1, \dots, x_n)$ holds in Γ . A homomorphism is called *strong*, if $R(x_1, \dots, x_n)$ holds in Γ *if and only if* $R(f(x_1), \dots, f(x_n))$ holds in Δ . A injective strong homomorphism is also called an *embedding*. A homomorphism from Γ to Γ is called an *endomorphism* of Γ , and a bijective strong endomorphism of Γ is called an *automorphism* of Γ .

Definition 1. A (finite or infinite) structure Γ is a *core* if every endomorphism of Γ is an embedding. A core Γ is called a *core* of Δ if Γ is the image of an endomorphism of Δ .

Homomorphisms that are not embeddings are called *strict*. The above definition says that cores do not have strict endomorphisms. For *finite* cores it clearly holds that every endomorphism is an automorphism. The following is well-known and easy to prove.

Proposition 1. *Every finite relational structure has a core, which is unique up to isomorphism.*

Therefore, we speak of *the core* of a finite relational structure Γ . As we will see, the notion of a core is central for the complexity study of constraint satisfaction problems; for its general rôle in structural combinatorics see e.g. [21].

1.1 Cores of Infinite Structures

Various core-like properties of infinite structures were studied by Bauslaugh [5, 6]. In general, infinite structures might not have a core in the sense introduced above, and the core might not be unique, see [5, 6].

In this article we focus on structures that are ω -categorical (also called *countably categorical*, or \aleph_0 -categorical), i.e., countably infinite structures with a first-order theory that has only one countable model, up to isomorphism. This is for instance the case for the dense linear order of the rational numbers $(\mathbb{Q}; <)$. Clearly, $(\mathbb{Q}; <)$ is a core. We state one of the main results.

Theorem 1. *Every ω -categorical structure has a core, which is unique up to isomorphism, and which is finite or ω -categorical.*

If we add all primitive positive formulas to the signature of an ω -categorical core, the resulting relational structure admits quantifier elimination. We also show that an ω -categorical structure Γ is a core if and only if every first-order formula has in Γ an existential positive definition. In particular, ω -categorical cores are model complete; see Section 5.

1.2 Constraint Satisfaction

The notion of a core has applications in the theory of constraint satisfaction. Let Γ be a structure with relational signature τ . The constraint satisfaction problem for the so-called *template* Γ is the following computational problem:

CSP(Γ)

INSTANCE: A finite structure S of the same relational signature τ as the template Γ .

QUESTION: Is there a homomorphism $h : S \rightarrow \Gamma$?

We want to stress that Γ is not part of the input; each Γ defines a computational problem. Note that the image of Γ under an endomorphism has the same constraint satisfaction problem as Γ .

Proposition 2. *Let Γ be a relational structure. Then $\text{CSP}(\Gamma)$ can be formulated as a constraint satisfaction problem with a finite template if and only if Γ has a finite core.*

Proof. Clearly, if Γ has a finite core T , then $\text{CSP}(\Gamma)$ is equivalent to $\text{CSP}(T)$. Conversely, suppose $\text{CSP}(\Gamma)$ equals $\text{CSP}(T)$ for a finite template T . This implies that every finite substructure of Γ homomorphically maps to T . A standard compactness argument shows that there is a homomorphism from Γ to T , and therefore Γ has a finite core. \square

For a finite template T , the computational problem $\text{CSP}(T)$ is clearly contained in NP. A classification of tractable and hard constraint satisfaction problems with a finite template is intensively studied, but still not complete. See [10,17,20], just to mention a few highlights on that subject. In all these approaches, the authors make use of the assumption that the templates of the constraint satisfaction problems under consideration are cores.

The class of constraint satisfaction problems with an *infinite* template was not yet studied systematically. It turns out that many interesting computational problems can be formulated with templates that are ω -categorical, as demonstrated by the following list of well-known computational problems.

- Allen’s interval algebra, and all its fragments [4, 22, 26, 29]
- Problems in phylogenetic analysis [19, 30]
- Tree description constraints in computational linguistics [8, 9, 14]
- Computational problems in the theory of relation algebras [16, 23, 28]
- All problems in monotone monadic SNP without inequality [7, 17]

In particular, every constraint satisfaction problem with a finite template T can also be formulated with an ω -categorical template. To see this, add for each vertex v in T a countably infinite number of copies v_1, v_2, \dots , such that for all $i \geq 1$ the relation $R(\dots, v_i, \dots)$ holds in the resulting structure Γ if and only if $R(\dots, v, \dots)$ holds in T . It is not hard to see that the structure Γ is ω -categorical, and that the core of Γ is isomorphic to T . Clearly, there are constraint satisfaction problems with ω -categorical templates that can not be formulated with finite templates: all the classes of computational problems mentioned above contain examples of such problems.

1.3 Examples

To illustrate the concepts we have seen so far, we formulate several well-known computational problems as constraint satisfaction problems. With Theorem 2 in the next section it will be easy to check that the corresponding templates are all ω -categorical. Three more examples follow at the end of Section 2, since we need the concept of amalgamation to define them conveniently. In all these examples, it is fairly easy to check that the chosen template is a core.

Betweenness. An important NP-hard problem is Betweenness [18], since the hardness of many fragments of Allen’s Interval Algebra [4, 26] can be proven easily by reduction from Betweenness. Given a finite set V , and a collection C of ordered triples (x, y, z) of distinct elements from V , the computational question is whether there is an injective function $f : V \rightarrow \{1, \dots, |V|\}$ such that, for each $(a, b, c) \in C$, we have either $f(a) < f(b) < f(c)$ or $f(c) < f(b) < f(a)$. The formulation as a constraint satisfaction problem is straightforward, using for instance the rational numbers as the base set of the template.

Switching-Acyclicity. Given a digraph $D = (V; E)$, can we partition the vertices V into two parts, such that the graph that arises from D by switching all arcs between the two parts is acyclic? To formulate this as a constraint satisfaction problem with an ω -categorical template, consider a dense subset X of \mathbb{Q} , and switch the order $<$ between the elements of X and $\mathbb{Q} - X$, and leave the edges within X and within $\mathbb{Q} - X$ unchanged [9]. The resulting structure is called $S(2)$ and is isomorphic for all choices of dense sets X , see e.g. [13]. The constraint satisfaction problem of $S(2)$ is the problem described above. For equivalent definitions of $S(2)$ and an hardness-proof of its constraint satisfaction problem, see [7, 9].

Partial tree descriptions. Our next example was studied in computational linguistics [14], and the first polynomial time algorithm can be found in [8]. Let D be a digraph with two types of arcs, called *ancestorship* and *non-ancestorship* arcs. The question is whether D is a consistent partial tree description, i.e., whether we can find a forest with oriented edges on the vertex set of D , such that for every ancestor arc in D there is a directed path in the forest, and for every non-ancestor arc there is no directed path in the forest.

To formulate this problem as a constraint satisfaction problem, we choose the following ω -categorical *dense proper semilinear order* [2, 12, 15]

as a template. The domain of the structure is the set A of all non-empty finite sequences $a = (q_0, q_1, \dots, q_{n-1})$ of rational numbers. Let $a < b$ if either

- b is a proper initial subsequence of a , or
- $b = (q_0, \dots, q_{n-1}, q_n)$ and $a = (q_0, \dots, q_{n-1}, q'_n, q_{n+1}, \dots, q_m)$, where the rational number q_n is smaller than q'_n .

The relation $<$ corresponds to ancestorship edges in partial tree descriptions. The set of all ordered pairs of distinct points that are not in $<$, denoted by $\not<$, corresponds to the non-ancestorship edges. $\text{CSP}((A; <, \not<))$ is the constraint satisfaction problem we were looking for.

Non-cores. Of course, there are plenty of ω -categorical structures that are not cores, for instance the Random graph \mathbf{R} [11, 24], whose core is the complete graph K_ω on a countably infinite set of vertices (the constraint satisfaction problem of \mathbf{R} and K_ω is trivial). Another example is the structure $(A; <)$, i.e., the template for partial tree descriptions introduced above without the relation $\not<$. Here the core is isomorphic to $(\mathbb{Q}; <)$.

2 Countably Categorical Structures

Finite structures are up to isomorphism determined by their first-order theory. We can not expect this for infinite structures: by the theorem of Löwenheim-Skolem, every consistent theory with a model of cardinality λ has models of arbitrary cardinality $\geq \lambda$. However, it might still be the case that all models of a certain cardinality are isomorphic. If this is the case for the countably infinite models, we call the theory ω -categorical. A countably infinite structure Γ is called ω -categorical, if its first-order theory $\text{Th}(\Gamma)$ (i.e., the set of all first-order sentences that hold in Γ , where the atomic formulas are built from the symbols in τ and equality) is ω -categorical. Throughout the paper we only consider relational and at most countable structures and signatures. Despite the powerful theorems quoted below, the class of ω -categorical structures remains somewhat mysterious, and all classification results require some additional properties (stability in e.g. [27], or homogeneity in [13]). All notions used here are standard and can be found e.g. in [24].

Theorem 2 (Engeler, Ryll-Nardzewski, Svenonius). *The following properties of a structure Γ are equivalent:*

1. *the structure Γ is ω -categorical;*

2. for each $n \geq 1$, there are finitely many orbits of n -tuples in the automorphism group of Γ ;
3. for each $n \geq 1$, there are finitely many inequivalent formulas with n free variables over Γ .

Permutation groups with the second property in Theorem 2 are called *oligomorphic* [11]. We need another concept, which is of a more combinatorial nature, and links ω -categoricity via *homogeneity* and Fraïssé's theorem to *amalgamation classes*.

A structure is *homogeneous* (sometimes also called *ultra-homogeneous*) if every isomorphism between finite substructures of Γ can be extended to an automorphism (in this paper, substructure always means *induced substructure*, as in [24]). A structure Γ admits *quantifier elimination*, if every first-order formula has in Γ a quantifier-free definition. An ω -categorical structure has quantifier elimination if and only if it is homogeneous (2.22 in [11]). A homogeneous τ -structure is ω -categorical if and only if the signature τ contains finitely many relation symbols of arity k , for all $k \geq 1$.

An example of an ω -categorical structure that is not homogeneous is $(\mathbb{A}; <, \neq)$ [15]. For an example of a homogeneous structure that is not ω -categorical, consider the expansion of a countably infinite structure Γ by unary singleton predicates for each element in Γ . This structure is homogeneous, since there are no distinct isomorphic substructures in Γ , and it is not ω -categorical, since the number of orbits in the automorphism group of Γ is infinite.

The next theorem asserts that a countable homogeneous structure is up to isomorphism characterized by its *age*, i.e., the set of its finite substructures. A class of finite relational structures \mathcal{C} is an *amalgamation class* if \mathcal{C} is nonempty, closed under isomorphisms and taking substructures, and has the *amalgamation property*, which says that for all $A, B_1, B_2 \in \mathcal{C}$ and embeddings $e_1 : A \rightarrow B_1$ and $e_2 : A \rightarrow B_2$ there exists $C \in \mathcal{C}$ and embeddings $f_1 : B_1 \rightarrow C$ and $f_2 : B_2 \rightarrow C$ such that $f_1 e_1 = f_2 e_2$.

Theorem 3 (Fraïssé). *A countable class \mathcal{C} of finite relational structures with countable signature is the age of a countable homogeneous structure if and only if \mathcal{C} is an amalgamation class. In this case the homogeneous structure is up to isomorphism unique and called the Fraïssé-limit of \mathcal{C} .*

The following templates of well-known constraint satisfaction problems are easily defined with amalgamation classes.

Triangle-free colorings. The class of all triangle-free graphs is an amalgamation class. Let us denote its Fraïssé-limit by \triangleleft . Clearly, $\text{CSP}(\triangleleft)$ is

tractable; but it can not be formulated with a *finite* template. The structure $[\mathcal{A}, \mathcal{A}]$, i.e., the structure that consists of two copies of \mathcal{A} , where all vertices between the two copies are linked, has an interesting constraint satisfaction problem, which can be formulated as follows: Given a graph, can we partition its vertices into two parts such that both parts do not contain a triangle? This problem is a rather typical example from the class *monotone monadic SNP without inequality (MMSNP)*, a fragment of existential second-order logic introduced in [17] in the context of constraint satisfaction. A general result on so-called *G-free colorability* implies its NP-hardness [1].

Quartet compatibility. The next example is an important structure in the theory of infinite permutation groups [11]. A *boron tree* is a finite tree in which all vertices have degree one (*hydrogen atoms*) or degree three (*boron atoms*). On the hydrogen atoms of a boron tree we can define a quaternary relation $xy|uv$ that holds when the paths joining x to y and u to v are disjoint. The class of all structures \mathcal{D} with a quaternary relation that stem from a boron tree as defined above is an amalgamation class [2]. Let D be the Fraïssé-limit of \mathcal{D} . Then $\text{CSP}(D)$ is a well-known NP-hard problem [30] that was independently studied in phylogenetic analysis (without any reference to constraint satisfaction), and is called *quartet-compatibility*: Given a collection C of quartets $xy|uv$ over a set X , is there some tree with leaf set X such that for each quadruple $xy|uv$ in C the paths from x to y and from u to v do not have common vertices?

Rooted triple consistency. The next problem is studied in phylogenetic analysis, again without notice that the problem can be stated as a constraint satisfaction problem. If we fix a point a in the previous structure D and consider the ternary relation ‘:’ defined by $x : yz \Leftrightarrow ax|yz$, we again obtain an ω -categorical structure (this is a *C-set* in [2]). The age of this structure now contains the finite structures T that come from finite *rooted* trees, and the relation $x : yz$ says that the least common ancestor of y and z is strictly below the least common ancestor of x, y , and z in the tree T . The corresponding constraint satisfaction problem is known as the *rooted triple consistency problem* [30], and tractable. The first polynomial time algorithm for this problem goes back to [3], motivated by a question in database theory.

3 Primitive Positive Expansions

A formula ϕ is *primitive* (*primitive positive*), if it is of the form

$$\exists \bar{x}. \psi_1 \wedge \cdots \wedge \psi_k$$

where ψ_i are literals (atomic formulas) that might include the equality relation. It is called *existential* (*existential positive*), if it is of the form $\exists \bar{x}. \Psi$ where Ψ is quantifier-free (and negation-free). The strongest of these four syntactic restrictions, *primitive positivity*, is important for constraint satisfaction, since the expansion of a template with a primitive positive definable (short, *p.p.-definable*) relation does not change the complexity of the corresponding constraint satisfaction problem. This is an easy observation, see e.g. [25]. But primitive positive definitions are also important for purely model theoretic questions.

Recall that a structure Γ admits *quantifier elimination*, if every first-order formula has in Γ a quantifier-free definition. For example, consider modules, and add all p.p.-definable relations to the signature. The theorem of Baur and Monk says that the resulting structure admits quantifier elimination (see e.g. [24]). As we will see at the end of this section, cores behave similarly in this respect. The following well-known fact can be applied to eliminate *negations* in existential formulas.

Proposition 3. *Let T be a first-order theory such that every homomorphism between models of T is an embedding. Then every existential formula is equivalent to an existential positive formula with respect to T .*

Proof. A formula is equivalent to an existential positive formula with respect to a theory T if and only if it is preserved by all homomorphisms between models of T ; this is e.g. Exercise 2 in Section 5.5 in [24]. Let f be a homomorphism between two models of T . By assumption, f is an embedding, and therefore clearly preserves all existential formulas. \square

We start with a proposition on primitive expansions.

Proposition 4. *If we expand a structure by all primitive definable relations, the resulting structure Γ is homogeneous.*

Proof. Let \bar{a} be a tuple of elements from Γ , let B_1, B_2 be induced substructures of Γ and $e_1 : \bar{a} \rightarrow B_1$ and $e_2 : \bar{a} \rightarrow B_2$ be embeddings. Since there are relation symbols for every primitive formula in the signature, there is a relation R_1 that holds on the tuple $e_1(\bar{a})$ and corresponds to the

primitive formula for the structure B_1 where the points from $B_1 - e_1(\bar{a})$ are existentially quantified. We also have a relation R_2 corresponding to B_2 where the points from $B_2 - e_2(\bar{a})$ are existentially quantified. Since e_1 and e_2 are embeddings, these relations also hold on \bar{a} , and they assert that we can find an extension C of \bar{a} with embeddings $f_1 : B_1 \rightarrow C$, $f_2 : B_2 \rightarrow C$ such that $f_1 e_1 = f_2 e_2$. Thus, the age of Γ has the amalgamation property, and Theorem 3 implies that Γ is homogeneous. \square

The following is an useful consequence of Proposition 3.

Proposition 5. *Let Γ be a τ -structure where τ contains a relation symbol for each primitive positive definable relation in Γ . If all homomorphisms between models of $\text{Th}(\Gamma)$ are embeddings, then Γ is homogeneous.*

Proof. If we expand Γ by all primitive definable relations, Proposition 4 shows that the resulting structure Γ' is homogeneous. Hence, each orbit of k -tuples is uniquely determined by the literals that hold on a k -tuple from the orbit. To show that also Γ is homogeneous, let R be an orbit of k -tuples in Γ . Since R has a quantifier-free definition in Γ' , R is in Γ equivalent to a boolean combination of primitive formulas. By Theorem 3, we know that primitive formulas have a primitive positive definition in Γ . Since τ contains a relation symbol for each of them, R has a quantifier-free definition, and therefore is homogeneous. \square

4 The Core of a Countably Categorical Structure

We prove that every countably categorical structure has a core, which is again ω -categorical and unique up to isomorphism. We start with a proposition that states the existence of a ‘youngest’ endomorphic image of an ω -categorical structure. The proof employs a typical technique for ω -categorical structures.

Proposition 6. *Let Γ be an ω -categorical relational τ -structure. Then there exists an endomorphism c of Γ such that for every other endomorphism g , all finite substructures of $c(\Gamma)$ embed into $g(\Gamma)$. This is, there exists an endomorphic image of Γ of smallest age.*

Proof. Let \mathcal{S} be the set of all finite τ -structures S such that there is an endomorphism g of Γ such that S is not a substructure of $g(\Gamma)$. We have to show that there is an endomorphism c such that $c(\Gamma)$ does not contain any substructure from \mathcal{S} . For the construction of c we consider the following tree. Let a_1, a_2, \dots be an enumeration of Γ . The vertices on

level n of the tree are equivalence classes of *good* homomorphisms from $\{a_1, \dots, a_n\}$ to Γ . A homomorphism h is good, if $h(\{a_1, \dots, a_n\})$ does not contain any substructure from \mathcal{S} . Two homomorphisms g_1 and g_2 are equivalent, if there exists an automorphism α of Γ such that $g_1 = g_2\alpha$. Clearly, if a homomorphism is good, then all equivalent homomorphisms and all restrictions are also good. A vertex u on level $n + 1$ in the tree is connected to a vertex v on level n , if some homomorphism from u is the restriction of some homomorphism from v . Because of ω -categoricity, the tree is finitely branching. We want to show that the tree has vertices on each level n , and iteratively construct a sequence h_1, h_2, \dots, h_k of homomorphisms from $\{a_1, \dots, a_n\}$ to Γ , where the last endomorphism h_k induces a good homomorphism. Initially, if the structure induced by $\{a_1, \dots, a_n\}$ does not have a substructure from \mathcal{S} , we can choose the identity as a good homomorphism. Otherwise, there is a substructure $S \in \mathcal{S}$ on the elements $\{a_1, \dots, a_n\}$ and an endomorphism e such that $e(\Gamma)$ does not contain S . Hence, $h_1 := e|_{\{a_1, \dots, a_n\}}$ is a strict homomorphism.

In step i , if the structure induced by $h_i(\{a_1, \dots, a_n\})$ does not have a substructure from \mathcal{S} , then h_i is a good homomorphism, and we are again done. Otherwise there is an endomorphism e of Γ and a structure $S \in \mathcal{S}$ on elements from $h_i(\{a_1, \dots, a_n\})$, such that $e(\Gamma)$ does not contain S . We can then define a strict homomorphism $h_{i+1} : \{a_1, \dots, a_n\} \rightarrow \Gamma$ by $h_{i+1}(x) := e(h_i(x))$. Since in the sequence of structures induced by $h_1(\{a_1, \dots, a_n\}), h_2(\{a_1, \dots, a_n\}), \dots$ either the number of vertices decreases or the number of tuples in relations increases, and since Γ is ω -categorical, the sequence has to be finite. Hence, there exists a good homomorphism from $\{a_1, \dots, a_n\}$ to Γ , for all $n \geq 0$. By König's tree lemma, there exists an infinite path in the tree. Since adjacency in the tree was defined by restriction between homomorphisms, this path defines an endomorphism c of Γ . By construction, $c(\Gamma)$ does not contain a substructure from \mathcal{S} . \square

It follows that all cores of Γ have the same age as $c(\Gamma)$.

Theorem 4. *Let Γ be an ω -categorical τ -structure. If τ contains a relation symbol for each primitive positive definable relation, then it has a homogeneous core, which is unique up to isomorphism. Moreover, the core is finite or ω -categorical.*

Proof. Let $\Gamma^* := c(\Gamma)$ be the structure induced by the endomorphism c constructed in Proposition 6. We first show

- (*) Every homomorphism between two models Γ_1 and Γ_2 of $Th(\Gamma^*)$ is strong and injective.

In particular, this holds for endomorphisms of Γ^* , and therefore Γ^* is a core. To prove (*), first observe that both Γ_1 and Γ_2 embed into Γ . They both have the same first-order theory and thus the same age as Γ^* , which is contained in the age of Γ . Since Γ is ω -categorical, there are embeddings of Γ_1 and Γ_2 in Γ , and we can thus assume that $\Gamma_1 \subseteq \Gamma$ and $\Gamma_2 \subseteq \Gamma$. Now suppose for contradiction that $f : \Gamma_1 \rightarrow \Gamma_2$ is a homomorphism that does not preserve the formulas $\neg R(\bar{u})$ or $u_1 = u_2$ for some k -ary relation R on Γ_1 and some tuple $\bar{u} = (u_1, \dots, u_k)$ of elements in Γ_1 . We will then construct an endomorphism h of Γ such that $h(\Gamma)$ does not contain a copy of the substructure S induced by \bar{u} in Γ_1 . This is a contradiction: On the one hand S is a substructure of Γ^* , since Γ_1 has the same theory and thus the same age as Γ^* . On the other hand, since S is not a substructure of $h(\Gamma)$, Proposition 6 says that S is not a substructure of Γ^* .

To construct this homomorphism h we consider an infinite but finitely branching tree. The vertices on level n in this tree will be equivalence classes of *good* homomorphisms from $\{a_1, a_2, \dots, a_n\}$ to Γ , where a_1, a_2, \dots is an enumeration of Γ . A homomorphism g on level n is *good*, if the structure induced by $g(\{a_1, \dots, a_n\})$ does not contain an induced copy of S . Two homomorphisms g_1 and g_2 are equivalent if there exists an automorphism α of Γ such that $g_1 = g_2\alpha$. Adjacency is defined by restriction; this is, two nodes on level n and $n + 1$ are adjacent in the tree if there are representatives g_1 and g_2 of the nodes such that g_1 is a restriction of g_2 . Clearly, all restrictions of a good homomorphism are again good homomorphisms, and all homomorphisms in an equivalence class are good, or all are not good. By ω -categoricity of Γ , the tree is finitely branching. The crucial step is that the tree contains vertices on every level, i.e., there exists a good homomorphism $h_n : \{a_1, \dots, a_n\} \rightarrow \Gamma$ for each $n \geq 1$. We show this in the following; and here we use the assumption that all p.p.-definable relations are in the signature of Γ .

To find h_n for each $n \geq 1$, we consider a sequence $(h_n^i)_{i \geq 0}$ of homomorphisms into Γ , where the domain of h_n^0 is $\{a_1, \dots, a_n\}$, and the domain of h_n^{i+1} equals the image of h_n^i . Hence, we can define the following composed homomorphism $h_n^{(i)} : \{a_1, \dots, a_n\} \rightarrow \Gamma$ by $h_n^{(i)}(x) := h_n^i(\dots h_n^1(h_n^0(x)) \dots)$. We now define the sequence $(h_n^i)_{i \geq 0}$. If the structure induced by the domain of h_n^i does not contain an induced copy of S , we are done, because then $h_n^{(i-1)}$ is a good homomorphism from $\{a_1, \dots, a_n\}$ to Γ . Otherwise, there are elements (b_1^i, \dots, b_k^i) in the domain of h_n^i that induce a structure isomorphic to the structure S induced by (u_1, \dots, u_k) in Γ . We now define $h_n^i(b_j^i) := f(u_j)$ for $1 \leq j \leq k$, and want to extend this mapping to a (strict) homomorphism h_n^i on the other

elements b_{k+1}^i, \dots, b_m^i , $m \leq n$, in the domain of h_n^i . Consider the formula $\phi := \exists x_{k+1}^i, \dots, x_m^i \psi$, where ψ is a conjunction of atomic formulas corresponding via $x_j^i \leftrightarrow b_j^i$ to the structure induced by b_{k+1}^i, \dots, b_m^i in Γ , and x_1^i, \dots, x_k^i are the free variables of ϕ . This formula ϕ clearly holds for b_1^i, \dots, b_k^i .

Since the signature contains a relation symbol for all p.p.-definable relations, the primitive positive formula ϕ also holds on u_1, \dots, u_k , since these vertices induce the same structure as b_1^i, \dots, b_k^i . Since f preserves primitive positive formulas, ϕ also holds on $f(u_1), \dots, f(u_k)$. We thus can find representatives r_{k+1}, \dots, r_m in Γ for the variables x_{k+1}^i, \dots, x_m^i of the existential quantifiers in ϕ , and extend h_i by $h_n^i(b_j^i) := r_j$ for $k+1 \leq j \leq m$. Then h_n^i clearly is a homomorphism to Γ , which is also strict, because it does not preserve some of the inequalities or negated relations that hold on b_1^i, \dots, b_k^i . Therefore the sequence $(h_n^{(i)})_i$ of homomorphisms must be finite, because ω -categoricity of Γ implies that there are only finitely many non-isomorphic homomorphic images of the structure induced by $\{a_1, \dots, a_n\}$ in Γ . Let $h_n^{(i_0)}$ be the last homomorphism in this sequence. By construction, this function is a good homomorphism h_n for every $n \geq k$.

Therefore, the constructed tree contains vertices on all levels, and König's tree lemma asserts that the tree contains an infinite path. Since adjacency is defined by restriction, this path defines an infinite endomorphism h of Γ . The image $h(\Gamma)$ does not contain an induced copy of S . This contradicts the minimality property of Γ^* formulated in Proposition 6. Hence, every homomorphism from Γ_1 to Γ_2 is an embedding. This completes the proof of (*).

Since τ contains a relation symbol for each p.p.-definable relation, and since a tuple in Γ^* satisfies a primitive positive formula if and only if it satisfies the formula in Γ , the signature contains a relation symbol also for all p.p.-definable relations in Γ^* . Since Γ^* also has Property (*), it satisfies the conditions of Proposition 5, and is homogeneous. By Fraïssé's theorem Γ^* is uniquely described by its age. We proved in Proposition 6 that every core has the same age as Γ^* , up to isomorphism, and therefore all cores of Γ are isomorphic.

Since Γ is ω -categorical, there are only finitely many inequivalent n -ary primitive positive formulas in Γ , and thus the signature of Γ^* contains finitely many relation symbols of each arity. Together with homogeneity, this implies that Γ^* is finite or ω -categorical. \square

We can now prove Theorem 1 that was already stated in the introduction.

Theorem 1. *Every countably categorical structure has a core, which is again ω -categorical and unique up to isomorphism.*

Proof. Let Γ be an ω -categorical τ -structure. To find the core of Γ , we first expand Γ by all p.p.-definable relations. By Theorem 4, there exists an endomorphism c of the expanded structure Γ^+ , such that $c(\Gamma^+)$ is an ω -categorical core. We denote the restriction of the core to the signature τ by Γ_0 . Clearly, $c(\Gamma) = \Gamma_0$ is also ω -categorical, since reducts of ω -categorical structures are ω -categorical. We claim that Γ_0 is a core: Suppose f is an endomorphism of Γ_0 that is not injective or not strong. Then f is also a strict endomorphism of Γ_0 with the signature that is expanded by the p.p.-definable relations; this contradicts that $c(\Gamma^+)$ is a core.

Next we show that Γ_0 is unique up to isomorphism. Suppose Γ_1 is another core of Γ . Then the expansion of Γ_1 by all p.p.-formulas is a core of Γ^+ , and by Theorem 4 isomorphic to $c(\Gamma^+)$. Hence the τ -restrictions Γ_1 and Γ_0 of these structures are also isomorphic. Thus the core of an ω -categorical structure is unique up to isomorphism. \square

Proposition 7. *Let Γ be an ω -categorical core, and let R be an orbit of k -tuples in $\text{Aut}(\Gamma)$. Then R has a primitive positive definition in Γ .*

Proof. Let Γ' be the expansion of Γ by all p.p.-definable relations. Since R is an orbit also in Γ' , all k -tuples in R are isomorphic to some substructure S of Γ' . By Theorem 4, Γ' is homogeneous, and all k -tuples in Γ' that are isomorphic to S are contained in R . Thus, R has a definition as a conjunction φ of atomic formulas. We replace all relation symbols in φ that are contained in the signature of Γ' , but not in the signature of Γ , by their primitive positive definition. The resulting formula is equivalent to a primitive positive definition of R in Γ . \square

This result yields alternative characterizations of cores of ω -categorical structures. An embedding of a τ -structures Γ in a τ -structure Δ is called *elementary*, if it preserves all first-order τ -formulas (this is a standard notion in model-theory [24]). A set of functions F from Γ to Γ *locally generates* a function g , if for every finite subset A of Γ there is a function $f \in F$ such that $g(a) = f(a)$ for all $a \in A$ (this is a standard notion in universal algebra [31]).

Theorem 5. *Let Γ be an ω -categorical structure. Then the following are equivalent.*

1. Γ is a core.
2. Every first-order formula has in Γ an existential positive definition.
3. Every endomorphism of Γ is elementary.
4. The automorphism group of Γ locally generates the endomorphism monoid of Γ .

Proof. Let Γ be an ω -categorical core. Every first-order definable k -ary relation R in Γ is the union of a finite number of orbits of k -tuples of $\text{Aut}(\Gamma)$. With Proposition 7 we can find a primitive positive definition for each orbit, and by commuting existential quantifiers with a finite disjunction we obtain an existential positive definition for R . Hence, 1 implies 2.

Since endomorphisms clearly preserve existential positive formulas, 2 implies that every endomorphism is elementary.

A function from Γ to Γ is in the local closure of the automorphism group of Γ if and only if it preserves all first-order definable relations. If we assume that every endomorphism is elementary, then the automorphism group locally generates all endomorphisms of Γ .

To prove that 4 implies 1, assume that Γ has a strict endomorphism f , i.e., there is a tuple (u_1, \dots, u_k) in Γ such that $f(u_1) = f(u_2)$, or $R(f(u_1), \dots, f(u_k))$ and not $R(u_1, \dots, u_k)$. Clearly, such a function f cannot be locally generated by automorphisms. \square

A theory T is called *model-complete*, if all embeddings between models of T are elementary. In the case that T is the theory of an ω -categorical structure Γ , we call also Γ model complete, as usual. It is well-known that an ω -categorical structure is model complete if and only if its first-order theory has an $\forall\exists$ -axiomatization, i.e., is equivalent to a set of sentences of the form $\forall\bar{x}\exists\bar{y}\phi(\bar{x}, \bar{y})$, where ϕ is quantifier-free (see e.g. Theorem 7.3.3f in [24]). The following follows immediately from Theorem 5.

Corollary 1. *The core of an ω -categorical structure is model-complete, and therefore its first-order theory has a $\forall\exists$ -axiomatization.*

5 Adding Constants to the Signature

One of the main results in [10] says that if Γ is a finite core, then adding a singleton-relation does not increase the complexity of the corresponding constraint satisfaction problem. We show that the same holds for constraint satisfaction problems where the template is an ω -categorical core. Note that this directly applies to all the computational problems presented in the introduction and Section 2.

Theorem 6. *Let Γ be an ω -categorical core, and Γ' be the expansion of Γ by a unary singleton relation $C = \{c\}$. If $\text{CSP}(\Gamma)$ is tractable, then so is $\text{CSP}(\Gamma')$. (If $\text{CSP}(\Gamma')$ is NP-hard, then so is $\text{CSP}(\Gamma)$.)*

Proof. We show how to solve $\text{CSP}(\Gamma')$ in polynomial time, under the assumption that $\text{CSP}(\Gamma)$ can be solved in polynomial time. Let S' be an instance of $\text{CSP}(\Gamma')$. Let P be the orbit of c in the automorphism group of Γ . By Proposition 7, P is p.p.-definable in Γ . Thus we can assume without loss of generality that Γ and Γ' contain the relation P . Replace all occurrences of the relation C in S' by the relation P . Solve the resulting instance S of $\text{CSP}(\Gamma)$; by assumption this is possible in polynomial time. If S is not satisfiable, then in particular S' was not satisfiable. On the other hand, if there is a homomorphism h from S to Γ , we claim that there is a homomorphism from S' to Γ' . Since P is the orbit of the element c , there is an automorphism α of Γ such that $h\alpha$ is a solution of the instance S' of $\text{CSP}(\Gamma')$. \square

6 Discussion

We showed that every ω -categorical structure Γ has a core, which is unique up to isomorphism. Since the core of Γ has the same constraint satisfaction problem as Γ , and since the core is again finite or ω -categorical, we reduced the classification of constraint satisfaction with ω -categorical templates to the classification of constraint satisfaction problems where the template is a finite or ω -categorical core.

The complexity of a constraint satisfaction problem does not change if we expand the template by a p.p.-definable relation. If we expand the core by all p.p.-definable relations, the resulting structure admits quantifier elimination. Finally we proved that a result known for constraint satisfaction with finite templates [10] remains valid for ω -categorical structures: if we expand an ω -categorical core by a singleton relation, then the resulting constraint satisfaction problem has the same complexity.

7 Acknowledgements

I am grateful to Julia Böttcher, Peter Cameron, Gregory Cherlin, Daniel Král, and Martin Ziegler for discussions.

References

1. D. Achlioptas. The complexity of G -free colourability. *Discrete Mathematics*, 165:21–30, 1997.

2. S. Adeleke and P. M. Neumann. Structure of partially ordered sets with transitive automorphism groups. *AMS Memoir*, 57(334), 1985.
3. A. Aho, Y. Sagiv, T. Szymanski, and J. Ullman. Inferring a tree from lowest common ancestors with an application to the optimization of relational expressions. *SIAM Journal on Computing*, 10(3):405–421, 1981.
4. J. F. Allen. Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11):832–843, 1983.
5. B. Bauslaugh. Core-like properties of infinite graphs and structures. *Disc. Math.*, 138(1):101–111, 1995.
6. B. Bauslaugh. Cores and compactness of infinite directed graphs. *Journal of Combinatorial Theory, Series B*, 68(2):255–276, 1996.
7. M. Bodirsky. Constraint satisfaction with infinite domains. Dissertation an der Humboldt-Universität zu Berlin, 2004.
8. M. Bodirsky and M. Kutz. Pure dominance constraints. In *Proceedings of STACS'02*, pages 287–298, 2002.
9. M. Bodirsky and J. Nešetřil. Constraint satisfaction with countable homogeneous templates. In *Proceedings of CSL'03*, pages 44–57, Vienna, 2003.
10. A. Bulatov, A. Krokhin, and P. G. Jeavons. Classifying the complexity of constraints using finite algebras. *To appear in SIAM Journal on Computing*, 2004.
11. P. J. Cameron. *Oligomorphic Permutation Groups*. Cambridge Univ. Press, 1990.
12. P. J. Cameron. The random graph. *R. L. Graham and J. Nešetřil, Editors, The Mathematics of Paul Erdős*, 1996.
13. G. Cherlin. The classification of countable homogeneous directed graphs and countable homogeneous n -tournaments. *AMS Memoir*, 131(621), January 1998.
14. T. Cornell. On determining the consistency of partial descriptions of trees. In *Proceedings of the ACL*, pages 163–170, 1994.
15. M. Droste. Structure of partially ordered sets with transitive automorphism groups. *AMS Memoir*, 57(334), 1985.
16. I. Düntsch. Relation algebras and their application in temporal and spatial reasoning. *Artificial Intelligence Review*, 23:315–357, 2005.
17. T. Feder and M. 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.
18. M. Garey and D. Johnson. *A guide to NP-completeness*. CSLI Press, 1978.
19. D. Gusfield. *Algorithms on strings, trees, and sequences. Computer Science and Computational Biology*. Cambridge University Press, New York, 1997.
20. P. Hell and J. Nešetřil. On the complexity of H -coloring. *Journal of Combinatorial Theory, Series B*, 48:92–110, 1990.
21. P. Hell and J. Nešetřil. *Graphs and Homomorphisms*. Oxford University Press, 2004.
22. R. Hirsch. Relation algebras of intervals. *Artificial Intelligence Journal*, 83:1–29, 1996.
23. R. Hirsch. Expressive power and complexity in algebraic logic. *Journal of Logic and Computation*, 7(3):309 – 351, 1997.
24. W. Hodges. *A shorter model theory*. Cambridge University Press, 1997.
25. P. Jeavons, D. Cohen, and M. Gyssens. Closure properties of constraints. *JACM*, 44(4):527–548, 1997.
26. P. Jeavons, P. Jonsson, and A. A. Krokhin. Reasoning about temporal relations: The tractable subalgebras of Allen's interval algebra. *JACM*, 50(5):591–640, 2003.

27. A. H. Lachlan. Stable finitely homogeneous structures: A survey. In *Algebraic Model Theory, NATO ASI Series*, volume 496, pages 145–159, 1996.
28. P. B. Ladkin and R. D. Maddux. On binary constraint problems. *Journal of the Association for Computing Machinery*, 41(3):435–469, 1994.
29. B. Nebel and H.-J. Bürckert. Reasoning about temporal relations: A maximal tractable subclass of Allen’s interval algebra. *JACM*, 42(1):43–66, 1995.
30. M. Steel. The complexity of reconstructing trees from qualitative characters and subtrees. *Journal of Classification*, 9:91–116, 1992.
31. A. Szendrei. *Clones in universal Algebra*. Seminaire de mathematiques superieures. Les Presses de L’Universite de Montreal, 1986.