On the Containment of Forbidden Patterns Problems

Florent Madelaine

LIMOS, UMR CNRS 6158, Campus des Cézeaux,
IUT informatique, B.P. 86, 63172 AUBIERE, France
fmadelaine@u-clermont1.fr

Abstract. Forbidden patterns problems are a generalisation of (finite) constraint satisfaction problems which are definable in Feder and Vardi's logic MMSNP [1]. In fact, they are examples of infinite constraint satisfaction problems with nice model theoretic properties introduced by Bodirsky [2]. In previous work [3], we introduced a normal form for these forbidden patterns problems which allowed us to provide an effective characterisation of when a problem is a finite or infinite constraint satisfaction problem. One of the central concepts of this normal form is that of a *recolouring*. In the presence of a recolouring from a forbidden patterns problem Ω_1 to another forbidden patterns problem Ω_2, containment of Ω_1 in Ω_2 follows. The converse does not hold in general and it remained open whether it did in the case of problems being given in our normal form. In this paper, we prove that this is indeed the case. We also show that the recolouring problem is Π_2^p-hard and in Σ_3^p.

Keywords: Constraint Satisfaction, Graph Homomorphism, Logic in Computer Science, Monadic Second Order Logic, Computational Complexity.

1 Introduction

Feder and Vardi [1] conjectured nearly 20 years ago that the class of non-uniform constraint satisfaction problems (CSP) has a dichotomy, that is that every problem in this class is either tractable or NP-complete. In contrast, it is believed that NP does not have the dichotomy property, as by Ladner's theorem [4], if $\mathsf{P} \neq \mathsf{NP}$, then there are problems in NP which are neither in P nor NP-complete. The dichotomy conjecture remains open though progress has been made using the central notion of polymorphisms in the mid nineties by Cohen, Jeavons and others and at the turn of the century great progress followed from Bulatov's powerful algebraic approach involving tame congruence theory (see [5] for a recent survey).

Descriptive complexity theory seeks to classify problems, *i.e.*, classes of finite structures, as to whether they can be defined using formulae of some specific logic, in relation to their computational complexity. One of the seminal results in descriptive complexity is Fagin's theorem [6] which states that a problem

D. Cohen (Ed.): CP 2010, LNCS 6308, pp. 345–359, 2010.

can be defined in existential second-order logic (ESO) if, and only if, it is in the complexity class NP. In their influential paper [1], Feder and Vardi also introduced the logic MMSNP, a syntactic fragment of monotone monadic ESO which is intimately linked to CSP. It is thought to be the largest such fragment to exhibit a dichotomy[1] and the derandomisation by Kun [7] of a lemma used by Feder and Vardi implies that MMSNP exhibits a dichotomy if, and only if, the dichotomy conjecture for CSP holds.

The logic MMSNP does not capture CSP: every problem in CSP can be defined in MMSNP but there are problems in MMSNP which are not in CSP [1,8]. In previous work with Iain Stewart [9,3], we provided an effective method to decide given a sentence of MMSNP whether it defines a problem in CSP or not. It turns out that these problems in MMSNP that are not in CSP are actually *constraint satisfaction problems with an infinite domain*, whose templates have nice model theoretic properties, introduced by Bodirsky [10]. So our previous result provides in fact a decision procedure that can tell whether a sentence of MMSNP defines a finite or an infinite CSP problem. In contrast, when the input of a problem definable in MMSNP is restricted to be of bounded degree, or from a proper minor closed class or more generally of bounded expansion, the restricted problem becomes a restricted *finite* CSP [11]. It is important to note that though there are infinite CSP *à la* Bodirsky which are not definable in MMSNP, this logic defines a large infinite class of natural infinite CSP which are worth studying in their own rights. For example, the complexity of problems in MMSNP have recently been investigated in the special case of monochromatic and loopless forbidden patterns [12].

Combinatorially, a problem in CSP can be seen as a homomorphism problem represented by a finite structure $\mathcal{T}$, the so-called template. It is well known that the containment of CSP corresponds exactly to the existence of a homomorphism from one template to another. More precisely, the CSP with template $\mathcal{T}_1$ is contained in the CSP with template $\mathcal{T}_2$ if, and only if, there is a homomorphism from $\mathcal{T}_1$ to $\mathcal{T}_2$. Therefore the category of relational structures and homomorphisms crops up naturally in the study of CSP [13].

Combinatorially, a problem in MMSNP can be represented by a finite set of coloured obstructions, the so-called *forbidden patterns*, and an instance is accepted if, and only if, it can be coloured while avoiding the presence of these patterns. The key ingredient of our previous result was to refine Feder and Vardi's normal form of MMSNP to take into account the fact that some colours might actually be redundant in the representation of the problem. To formalise this, we introduced the notion of a recolouring from a forbidden patterns problem Ω_1 to another forbidden patterns problem Ω_2 and showed that in the presence of such a recolouring, containment of the problem Ω_1 in the problem Ω_2 followed. The converse does not hold in general and it remained open whether it did in the case of problems being given in our normal form. In this paper, we prove that this is indeed the case. It follows that representations of forbidden patterns problems given in a normal form and recolourings provide us with the right category in

[1] Feder and Vardi showed that monotone monadic SNP with $\neq$ does not have a dichotomy.

the context of MMSNP. It would be interesting to settle the complexity of the containment of forbidden patterns problems. We investigate as a first step the complexity of the recolouring problem and show that it is in Σ_3^p and that it is Π_2^p-hard.

This paper is organised as follows. In the next section, as the reader may not be familiar with MMSNP, we shall introduce key concepts informally, mostly by discussing examples, prove simple cases of our main result to illustrate our method, and finally state our main result. In Section 3, we detail and adapt the computational equivalence between CSP and MMSNP given by Feder and Vardi. In Section 4, we prove our main result. We conclude with a discussion of the complexity of some related problems.

2 Preliminaries

Existential Second Order Logic. Fagin's theorem equates definability in ESO with membership in the complexity class NP. For example, the class of 3-colourable graphs can be defined using a sentence of the following form.

$$\begin{aligned}\Phi_1 := \exists R, G, B,\ &\text{three sets partitioning the vertices}\\ &\forall x, y,\ \neg\big(E(x,y) \wedge R(x) \wedge R(y)\big) \wedge \neg\big(E(x,y) \wedge G(x) \wedge G(y)\big)\\ &\quad \wedge \neg\big(E(x,y) \wedge B(x) \wedge B(y)\big)\end{aligned}$$

A graph is represented as a relational structure whose domain consists of vertices equipped with a single binary predicate E representing the edge relation. The above sentence has two kinds of quantifiers: second-order variables (always upper-case) are interpreted as relations, like R which is interpreted as a set of vertices, and first-order variables (always lower case), like x, which is interpreted as a vertex. The three second order predicates R, G and B stand for three colours, say red, green and blue and the sentence asserts that the vertices may be coloured with these three colours in such a way that for every edge in the graph, the extremities have different colours.

In this paper, we shall only need second-order predicates that are sets, the so-called *monadic* predicates, and we shall only allow them to be existentially quantified as in the above example. Note that finitely many sets of vertices correspond essentially to a partition of the vertices in distinct *colours*. In combinatorial terms this means that in order to check a property we have to guess some colours for each vertex before verifying some first-order property over the coloured graph. Let us clarify this with another example.

$$\begin{aligned}\Phi_2 := \exists M, N\ \forall x, y,\ &\neg\big(\neg M(x) \wedge \neg N(x)\big)\\ &\wedge \neg\big(E(x,y) \wedge M(x) \wedge N(x) \wedge M(y) \wedge N(y)\big)\\ &\wedge \neg\big(E(x,y) \wedge \neg M(x) \wedge N(x) \wedge \neg M(y) \wedge N(y)\big)\\ &\wedge \neg\big(E(x,y) \wedge M(x) \wedge \neg N(x) \wedge M(y) \wedge \neg N(y)\big)\end{aligned}$$

There are two monadic predicates M and N in Φ_2 and for a given vertex x there are four cases to consider: x is in both M and N ($M(x) \wedge N(x)$ holds), x is in M but not in N ($M(x) \wedge \neg N(x)$ holds) etc. So the above sentence disallows one of the colour (with the conjunct $\neg(\neg M(x) \wedge \neg N(x))$) and states for the three other colours that an edge can not have both extremities of the same colour. In other words, this sentence defines also the fact that a graph is 3-colourable.

Monotone Monadic Strict NP *without inequalities.* The two sentences Φ_1 and Φ_2 have a particular syntactic form: $\exists$ monadic predicates, $\forall$ variables ranging over vertices, followed by a quantifier-free first-order formula. Such sentences form the fragment SNP of ESO. It turns out that many combinatorial problems are definable in SNP, in particular every problem in CSP can be defined by a SNP sentence. For example, in the case of 3-colourability, we may use the sentence Φ_2. Let us explain in a bit more detail how we may build this sentence in a systematic fashion. Recall first that for a CSP with template $\mathcal{T}$, a structure $\mathcal{A}$ is a yes-instance if, and only if, there exists a *homomorphism* from $\mathcal{A}$ to $\mathcal{T}$. That is, a mapping h from the domain of $\mathcal{A}$ to that of $\mathcal{T}$ such that every arc in $\mathcal{A}$ is mapped to an arc in $\mathcal{T}$ (assuming we deal with digraphs for now for the sake of simplicity). The 3-colourability problem, recast as a digraph problem, has as template $\mathcal{T}$ the digraph with 3 vertices and all possible arcs that are not self-loops. Viewing the 3 elements of $\mathcal{T}$ as colours, we have readily explained how to use 2 monadic predicates M and N and one forbidden combination of them $\neg(\neg M(x) \wedge \neg N(x))$ to encode three colours. In order to enforce a homomorphism, we now encode the non-arcs of $\mathcal{T}$ by adding negated conjuncts, one for each non-arc. For example, if $M(x) \wedge N(x)$ stands for the colour corresponding to the first vertex of $\mathcal{T}$ and since there is no self-loop around this vertex, we add the following negated conjunct to the sentence:

$$\neg\big(E(x,y) \wedge M(x) \wedge N(x) \wedge M(y) \wedge N(y)\big).$$

Doing this with every non-arc, we obtain the sentence Φ_2 given above. It is important to note that the sentence we build this way uses only *monadic* predicates. Furthermore, the first-order part is a conjunction of negated conjuncts; and, in every negated conjunct atoms from the input (the edge symbol E in our examples) appears always positively. This means that the sentence is *monotone*. Finally, we never use the symbol $\neq$. We have therefore built a sentence of SNP that is *monadic*, *monotone* and *without inequality*. The sentences of SNP satisfying these three restrictions form the logic MMSNP introduced by Feder and Vardi. As we may build such a sentence for every template, we now know that

$$\text{CSP} \subseteq \text{MMSNP}.$$

Some sentences of MMSNP give rise to problems that are not in CSP and are in fact constraint satisfaction problems with an infinite template. For example,

$$\Psi_1 := \forall x, \forall y, \forall z,\ \neg\big(E(x,y) \wedge E(y,z) \wedge E(z,x)\big)$$

expresses that there are no oriented 3-cycles in a digraph (and also no self-loop as the variables may be equal). It is not difficult to see that this problem is not in CSP. Assume for contradiction that there exists a template $\mathcal{T}$ with n elements for this problem. We may build a yes-instance $\mathcal{A}$ for ψ_1 as follows: take $n+1$ vertices and add between any pair of distinct vertices a directed path of length 3. By assumption, there exists a homomorphism from $\mathcal{A}$ to $\mathcal{T}$. This homomorphism must identify two distinct elements joined by a directed 3-path. Hence, $\mathcal{T}$ contains a loop or an oriented 3-cycle and is a no-instance which is absurd as the template is always a yes-instance.

The problem defined by Ψ_1 is in fact a CSP with an infinite template. It is not difficult to construct an infinite template for this problem: simply take as a template the disjoint union of its yes-instances[2]. This infinite template is not particularly interesting, however, we may also construct for this problem an infinite template that has a nice model theoretic property called *ω-categoricacity.* From now on, by infinite CSP, we mean a problem with such a nice template. This property means in particular that the Galois-connection used in the finite case can be successfully adapted and that some logico-algorithmic results such as those involving Datalog still hold. We will refrain from going into more details and refer to Bodirsky's survey [2] on his pioneering work on infinite CSP.

Obstructions and containment. Note that the negated conjunct

$$\neg\big(E(x,y) \wedge E(y,z) \wedge E(z,x)\big)$$

in Ψ_1 essentially forbids the occurrence of an oriented 3-cycle. However, since the variables x, y and z may take the same value, this means in fact that we forbid the existence of a homomorphism from the oriented 3-cycle to the instance. Hence, the problem defined by Ψ_1 can be seen as a dual problem to a CSP. Whereas in the case of CSP we ask whether there is a homomorphism from the instance $\mathcal{A}$ to the template $\mathcal{T}$, we will ask here whether there is no homomorphism from an obstruction $\mathcal{F}$ to the instance $\mathcal{A}$. In the case of more than one obstruction, we have essentially the fragment of MMSNP that has no monadic predicate (sentences of MMSNP that are also first-order). In general, such a problem is known to be an infinite CSP [14]. Let us consider two such problems Ω_1 and Ω_2 given by two sets of obstructions $\mathscr{F}_1$ and $\mathscr{F}_2$. We will insist for simplicity for the obstructions to be connected[3]. We say that the problem Ω_1 is *contained* in Ω_2 if, and only if, for any instance $\mathcal{A}$, if $\mathcal{A}$ is a yes-instance of Ω_1 then $\mathcal{A}$ is a yes-instance of Ω_2. When is Ω_1 contained in Ω_2? A simple criteria defined in terms of existence of homomorphisms between the obstructions characterises containment in this simple case.

Proposition 1 ([9], see also [15]). *Ω_1 is contained in Ω_2 if, and only if, for every obstruction $\mathcal{F}_2$ in $\mathscr{F}_2$ there exists an obstruction $\mathcal{F}_1$ in $\mathscr{F}_1$ such that there is a homomorphism from $\mathcal{F}_1$ to $\mathcal{F}_2$.*

[2] This is true in general for any monotone problem that is closed under disjoint union.

[3] This is not a strong hypothesis as a problem with a disconnected obstruction is in fact the disjoint union of problems with connected obstructions.

The main result of this paper is a generalisation of the above result to the case where the obstructions are coloured, that is when the corresponding MMSNP sentences are no longer first-order sentences.

Forbidden patterns problems. Another example of a problem that is in MMSNP but not in CSP is:

$$\begin{aligned}\Psi_2 := \exists M,\ \forall x, y, z,\ &\neg\big(E(x,y) \wedge E(y,z) \wedge E(z,x) \wedge M(x) \wedge M(y) \wedge M(z)\big)\\ &\wedge \neg\big(E(x,y) \wedge E(y,z) \wedge E(z,x) \wedge \neg M(x) \wedge \neg M(y) \wedge \neg M(z)\big).\end{aligned}$$

We have a single monadic predicate which encodes two colours, say white and black. The two negated conjuncts forbid two vertex-coloured structures, namely a white oriented 3-cycle $\mathcal{F}'_1$ and a black oriented 3-cycle $\mathcal{F}'_2$. Thus, the problem defined by Ψ_2 accepts an instance $\mathcal{A}$ whenever its vertices can be coloured in white and black into a structure $\mathcal{A}'$ such that there is neither a homomorphism from $\mathcal{F}'_1$ to $\mathcal{A}'$ nor a homomorphism from $\mathcal{F}'_2$ to $\mathcal{A}'$.

In general a *forbidden patterns problem* Ω is given by a finite set of coloured structures. We insist that each structure is *connected* and *contains at least one tuple*. It makes sense to formalise the (vertex-)colouring of a structure by a homomorphism into some structure describing the colours. So Ω is given by a structure $\mathcal{T}$ representing the colours and a set $\mathscr{F}'$ of $\mathcal{T}$-coloured structures, the so-called *forbidden patterns*.

A $\mathcal{T}$-coloured structure is a pair $(\mathcal{F}, f)$ where f is a homomorphism from $\mathcal{F}$ to $\mathcal{T}$ which describes the colouring. The notion of structure homomorphism generalises naturally to coloured structures: given two $\mathcal{T}$-coloured structures $(\mathcal{F}, f)$ and $(\mathcal{G}, g)$, a homomorphism h from $(\mathcal{F}, f)$ to $(\mathcal{G}, g)$ is simply a homomorphism from $\mathcal{F}$ to $\mathcal{G}$ that preserves the colours, that is such that $f = g \circ h$.

An instance $\mathcal{A}$ of the problem Ω is a *yes-instance* if, and only if, there exists a homomorphism h from $\mathcal{A}$ to $\mathcal{T}$ such that there is no homomorphism from any forbidden pattern $(\mathcal{F}, f)$ in $\mathscr{F}'$ to $(\mathcal{A}, h)$.

When h is not a homomorphism or when there is a homomorphism from some forbidden pattern $(\mathcal{F}, f)$ in $\mathscr{F}'$ to $(\mathcal{A}, h)$, we say that $(\mathcal{A}, h)$ is *not valid* w.r.t. Ω. We denote by FPP the class of forbidden patterns problems. Forbidden patterns problems are known to be infinite CSP [10] and every sentence in MMSNP captures a finite union of problems in FPP [3].

Recolouring. A recolouring is a homomorphism which states how the colours of a problem Ω_1 can be transformed into colours of a problem Ω_2. Let us recall the formal definition before looking at an example.

Definition 2 (recolouring [9,3]). *Let Ω_1 (respectively, Ω_2) be a forbidden patterns problem given by $\mathcal{T}_1$ and a set $\mathscr{F}'_1$ of $\mathcal{T}_1$-coloured forbidden patterns (respectively, $\mathcal{T}_2$ and $\mathscr{F}'_2$).*

A recolouring *from Ω_1 to Ω_2 is a homomorphism r from $\mathcal{T}_1$ to $\mathcal{T}_2$ such that for every $(\mathcal{F}_2, f_2)$ forbidden by Ω_2, any of its inverse image $(\mathcal{F}_2, f_1)$ under r is not valid w.r.t. Ω_1. In other words, for every $\mathcal{T}_2$-coloured pattern $(\mathcal{F}_2, f_2)$ in $\mathscr{F}'_2$,*

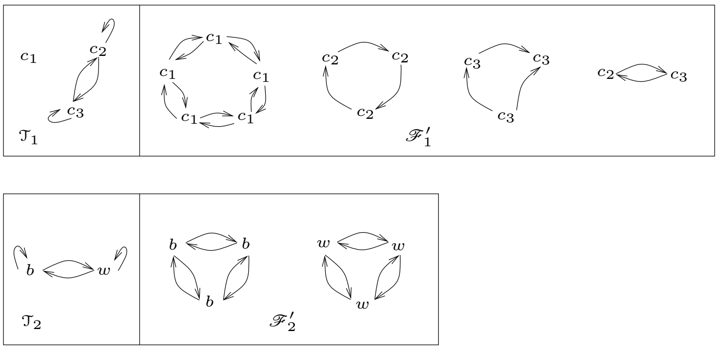

Fig. 1. Two forbidden patterns problems

and for any $\mathcal{T}_1$*-coloured structure* $(\mathcal{F}_2, f_1)$ *such that* $f_2 = r \circ f_1$*, there exists a forbidden pattern* $(\mathcal{G}_1, g_1)$ *in* $\mathscr{F}_1'$ *and a homomorphism* h *from* $(\mathcal{G}_1, g_1)$ *to* $(\mathcal{F}_2, f_1)$*.*

Note how this definition generalises the condition between the obstructions given in Proposition 1. We already know that the existence of a recolouring implies containment.

Proposition 3 ([9,3]). *If there is a recolouring* r *from a forbidden patterns problem* Ω_1 *to a forbidden patterns problem* Ω_2 *then* Ω_1 *is contained in* Ω_2*.*

Example 4. We consider the two forbidden patterns given on Figure 1 (note how the colours of the vertices of a forbidden pattern are simply given by labelling a vertex with its colour). The problem represented by $\mathcal{T}_2$ and $\mathscr{F}_2'$ is a variant of the problem defined by Ψ_2 in which triangles have arcs in both directions. Let r be the mapping from the colours of the first problem, namely $\{c_1, c_2, c_3\}$ to those of the second problem, namely $\{b, w\}$, that maps c_1, c_2 and c_3 to b. Note that r is indeed a homomorphism from $\mathcal{T}_1$ to $\mathcal{T}_2$. The only forbidden pattern of the second problem whose colours are in the image of r is the black triangle (the first forbidden pattern of $\mathscr{F}_2'$ listed on the figure). We need to show that every triangle whose vertices is coloured via r^{-1} are invalidated by the first problem. This can happen in two ways: the colouring may not be a homomorphism to $\mathcal{T}_1$, or some forbidden pattern in $\mathscr{F}_1'$ invalidates it. If the colours of the three vertices of the triangle are replaced by c_1, then the 5-cycle (the first forbidden pattern of $\mathscr{F}_1'$ listed on the figure) invalidates this choice of colours. Similarly, if the vertices are all coloured by c_2 only or c_3 only then the two next forbidden patterns on the figure invalidate these choices. If the colours of the three vertices of the triangle are replaced by c_1 and other colours then the colouring is not a homomorphism to $\mathcal{T}_1$. If the colours are replaced by c_2 and c_3 but not c_1 then the last forbidden pattern listed on the figure invalidates this choice. This shows that r is a recolouring from the first problem to the second problem.

Normal Form for Forbidden Patterns Problems. In this paper, we prove that the converse of Proposition 3 holds, when the two problems are given in the *normal form*. Note that this can always be done.

Theorem 5 ([3]). *Every forbidden patterns problem can be given by a representation in the normal form.*

We shall recall shortly what conditions this normal form entails. Let us first introduce some vocabulary. We say that a coloured structure is *weakly valid* w.r.t. a forbidden patterns problem if there is no *injective* homomorphism from a forbidden pattern into it. A forbidden patterns that consists of a coloured structure with a single tuple that mentions each element exactly once[4] is said to be *conform*. When a forbidden pattern is conform, we may drop it from the list of forbidden patterns and enforce its constraint by amending the structure $\mathcal{T}$ accordingly (by removing the corresponding tuple from $\mathcal{T}$). A forbidden patterns problem Ω is given by a structure $\mathcal{T}$ and a set of forbidden $\mathcal{T}$-coloured structures $\mathscr{F}'$. The pair $(\mathcal{T}, \mathscr{F}')$ is called a *representation* of Ω. If every recolouring from $(\mathcal{T}, \mathscr{F}')$ to itself is an automorphism of $\mathcal{T}$ then we say that the representation $(\mathcal{T}, \mathscr{F}')$ is a *core*.

Definition 6 (Normal Form [3]). *A representation $(\mathcal{T}, \mathscr{F}')$ of a forbidden patterns problem Ω is said to be* in the normal form *if, and only if it satisfies the following six conditions.*

- *($\mathfrak{p}1$). An instance is valid if, and only if, it is weakly valid.*
- *($\mathfrak{p}2$). Every pattern of $\mathscr{F}'$ is a core (as a coloured structure).*
- *($\mathfrak{p}3$). It is not the case that $(\mathcal{F}_1, f_1)$ is a substructure of $(\mathcal{F}_2, f_2)$, for any distinct patterns $(\mathcal{F}_1, f_1)$ and $(\mathcal{F}_2, f_2)$ in $\mathscr{F}'$.*
- *($\mathfrak{p}4$). No pattern of $\mathscr{F}'$ is conform.*
- *($\mathfrak{p}5$). Every forbidden pattern is biconnected.*
- *($\mathfrak{p}6$). The representation $(\mathcal{T}, \mathscr{F}')$ is a core.*

Example 7. Let Ω_4 be the problem given on the top of Figure 2. We shall discuss briefly how its normal form is computed without explaining why the obtained problem is equivalent, for further details please refer to [3].

First we enforce $\mathfrak{p}_1$ to $\mathfrak{p}_3$ simply by taking the homomorphic image of the forbidden pattern, keeping only the minimal ones with respect to injective homomorphisms. Note that $\mathfrak{p}_4$ holds also in the representation of the problem we obtain this way which is given in the second row on the figure.

Next, we enforce $\mathfrak{p}_5$ by splitting the path of length two along its articulation point and copying its colour c into two new colours b and w, one for the substructure to the left of this articulation point, one for the substructure to the right of this articulation point. Replacing elsewhere the colour c by w and b in all possible ways and simplifying again by keeping the minimal patterns to enforce $\mathfrak{p}_3$, we obtain the representation which is given in the fourth row of the figure. Note that it no longer satisfies $\mathfrak{p}_4$.

[4] Self-loops and their generalisation like $R(x, x, y)$ are not conform.

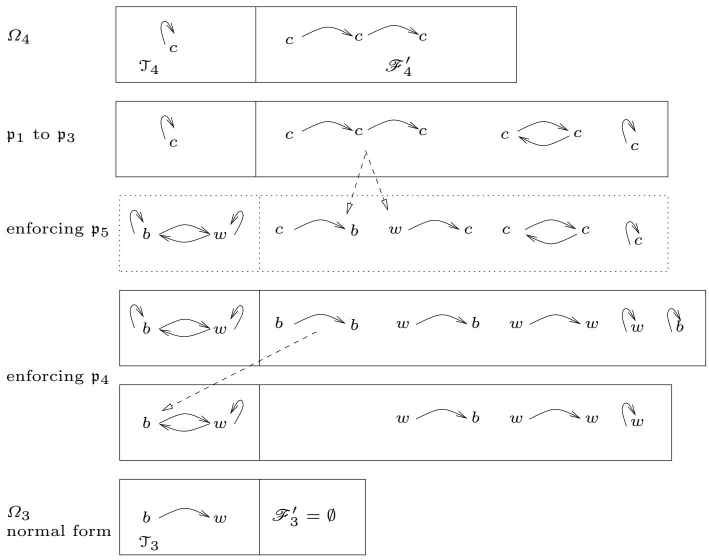

Fig. 2. Computing the normal form

We enforce progressively $\mathfrak{p}_4$ by removing the conform forbidden patterns and removing the corresponding tuple in the structure describing the colours. We also remove any forbidden pattern that is no longer a coloured structure. We finally obtain this way the problem Ω_3 given in the last row on the figure.

The mapping r which sends w and b to the single colour c of Ω_4 is a recolouring from Ω_3.

Conversely, there is no recolouring from Ω_4 to Ω_3 as there is no homomorphism from $\mathcal{T}_4$ to $\mathcal{T}_3$, since the former is a self-loop and the latter has no self-loop.

Note that the two problems Ω_3 and Ω_4 coincide and that Ω_3 is given in the normal form but that Ω_4 is not (its only forbidden pattern fails to be biconnected).

We are now ready to state the main result of this paper.

Theorem 8 (main result). *Let Ω_1 and Ω_2 be two forbidden patterns problems* given in the normal form *over the relational signature σ. Ω_1 is contained in Ω_2 if, and only if, there is a recolouring from Ω_1 to Ω_2.*

Another case where it is not too hard to see that the converse of Proposition 1 holds is when Ω_2 is in CSP. Though this case is subsumed by our main result, its proof will serve as a good warm-up. In particular, it will allow us to introduce a key ingredient which is a generalisation by Feder and Vardi of a result due to Erdös.

Proposition 9. *Let Ω_1 and Ω_2 be two forbidden patterns problems. If both problems are given in the normal form and Ω_2 is in* CSP *then Ω_1 is included in Ω_2 if, and only if, there is a recolouring r from Ω_1 to Ω_2.*

Recall that the *girth* of a structure is the length of its shortest cycle (and so if there are no cycles then the structure has infinite girth).

Lemma 10 (Erdös lemma [1]). *Fix two positive integers r and s. For every structure $\mathcal{B}$, there exists a structure $\mathcal{D}$ such that: the girth of $\mathcal{D}$ is greater than r; there is a homomorphism from $\mathcal{D}$ to $\mathcal{B}$; and for every structure $\mathcal{C}$ of size at most s, there is a homomorphism from $\mathcal{B}$ to $\mathcal{C}$ if, and only if, there is a homomorphism from $\mathcal{D}$ to $\mathcal{C}$.*

Proof (of Proposition 9). As Ω_2 is in the normal form and in CSP, this means that that $\mathscr{F}'_2 = \emptyset$ [3]. Thus, in this case a recolouring is nothing other than a homomorphism from $\mathcal{T}_1$ to $\mathcal{T}_2$. In particular if $\mathcal{T}_1$ is a yes-instance of Ω_1 then we are done. However, this is in fact not true in general.

By assumption Ω_1 is given in the normal form. This means that $\mathcal{T}_1$ is a no-instance of Ω_1 unless $\mathscr{F}'_1 = \emptyset$ [3]. We use Erdös Lemma: we choose r greater than the largest forbidden patterns in $\mathscr{F}'_1$; s to be $|\mathcal{T}_2|$, the size of $\mathcal{T}_2$; and $\mathcal{B} := \mathcal{T}_1$.

We claim that the structure $\mathcal{D}$ obtained from the lemma in this way is in fact a yes-instance of Ω_1. This is because the homomorphism, say d_1, given by the lemma from $\mathcal{D}$ to $\mathcal{B} = \mathcal{T}_1$ gives us a valid colouring w.r.t. Ω_1. To see this, we use the fact that Ω_1 is given in the normal form: it suffices to show that $(\mathcal{D}, d_1)$ is weakly valid; and, for every forbidden pattern $(\mathcal{F}_1, f_1)$, the structure $\mathcal{F}_1$ is biconnected and must contain a cycle, so it can not occur as a substructure of $\mathcal{D}$ which has a girth greater than the size of any forbidden patterns.

By containment of Ω_1 in Ω_2 it follows that $\mathcal{D}$ is a yes-instance of Ω_2 and that there is a homomorphism from $\mathcal{D}$ to $\mathcal{T}_2$. Hence, by construction of $\mathcal{D}$ this means that there is a homomorphism from $\mathcal{B} = \mathcal{T}_1$ to $\mathcal{T}_2$ and that we are done. □

3 From Forbidden Patterns Problem to CSP and Back

The following result is an adaptation of the ideas of Feder and Vardi's reduction of MMSNP to CSP [1] to forbidden patterns problems. We shall only sketch the proof. A detailed proof using the same notation is available in [9]. There is a small difference here, as the signature of the CSP is now parameterised by a set of patterns that must include the patterns from the forbidden patterns problem considered but may include more. This result is one of the ingredient of the proof of our main result. We denote by CSP$(-, \mathcal{T})$ the (non-uniform) constraint satisfaction problem with template $\mathcal{T}$ and by CSP$(\text{girth} > \gamma, \mathcal{T})$ its restriction to input of girth greater than γ.

Theorem 11. *Let Ω be a forbidden patterns problem given in the normal form over the relational signature σ. Let $\mathscr{F}$ be a set of biconnected σ-structures that includes all structures involved in patterns forbidden by Ω. Let γ be a fixed integer greater than the largest structure in $\mathscr{F}$.*

There exists a relational signature τ, a τ-structure $\mathcal{T}_\Omega$, and two first-order interpretations Π and Π^{-1} such that:

– *τ extends σ with new symbols, one symbol $R_\mathcal{F}$ of arity $|\mathcal{F}|$ for each $\mathcal{F}$ in $\mathscr{F}$;*

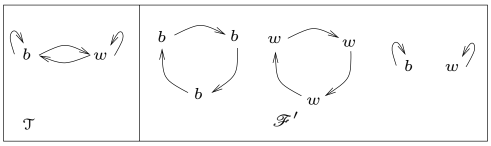

Fig. 3. No-Monochromatic-Triangle

- Π *is a quantifier-free first-order interpretation using conjunction only;*
- Π^{-1} *is a first-order interpretation;*
- $\Pi^{-1} \circ \Pi$ *is the identity over* σ*-structures;*
- Ω *reduces to* CSP$(-, \mathcal{T}_\Omega)$ *via* Π*; and,*
- CSP$(girth > \gamma, \mathcal{T}_\Omega)$ *reduces to* Ω *via* Π^{-1}.

We sketch the proof of this result in the remaining of this section, providing an example to help the reader understand the main ideas[5].

Example 12. We consider the forbidden patterns problem defined by the sentence Ψ_2 in the introduction. It is a variant of the well-known NP-complete problem **No-Monochromatic-Triangle**. It is given in its normal form on Figure 3. The signature of this problem is $\sigma = \langle E \rangle$ where E is binary which we extend to a *new signature* $\tau = \langle E, R, S \rangle$ where R is ternary and S unary (R encodes the 3-cycles and S the self-loops). The *interpretation* Π *from* σ *to* τ is given by: $\varphi_R(y_1, y_2, y_3) := E(y_1, y_2) \wedge E(y_2, y_3) \wedge E(y_3, y_1)$, $\varphi_S(y_1) := E(y_1, y_1)$ and $\varphi_E(y_1, y_2) := E(y_1, y_2)$. The *interpretation* Π^{-1} *from* τ *to* σ is given by the formula ψ_E which is as follows:

$$\big(E(y_1, y_2)\big) \vee \big(y_1 = y_2 \wedge S(y_1)\big) \vee \big(\exists x\ R(y_1, y_2, x) \vee R(x, y_1, y_2) \vee R(y_2, x, y_1)\big).$$

The structure $\mathcal{T}_\Omega$ has two elements b and w and, relations $E := \{b, w\}^2$, $S := \emptyset$ and $T := \{b, w\}^3 \setminus \{(b, b, b)\ (w, w, w)\}$. □

Signature of the CSP. The problem Ω is represented by a σ-structure $\mathcal{T}$ and a list of forbidden $\mathcal{T}$-coloured structures $\{(\mathcal{F}_1, f_1), (\mathcal{F}_2, f_2), \ldots, (\mathcal{F}_n, f_n)\}$. Let $\mathscr{F}$ be the set of the σ-structures that consists of the structures $\mathcal{F}_i$ considered up to isomorphism. For every $\mathcal{F}$ in $\mathscr{F}$, we introduce a new symbol $R_{\mathcal{F}}$ of arity $|\mathcal{F}|$. Let τ be the signature that consists of the symbol of σ together with the new symbols $R_{\mathcal{F}}$.

Interpretation from the forbidden patterns problem to the CSP. Let $\varphi_{\mathcal{F}}$ be the quantifier-free part of the canonical conjunctive query of $\mathcal{F}$, that is:

$$\varphi_{\mathcal{F}} := \bigwedge_{R \in \sigma} \ \bigwedge_{R^{\mathcal{F}}(\bar{x}) \text{ holds}} R(\bar{x})$$

[5] We advise the reader to go through the proof and progress in parallel on the example.

Let $\varphi_R := R(\bar{x})$. Let Π be the interpretation from σ to τ given by the formulae $\varphi_{\mathcal{F}}$ and the formulae φ_R. Note that Π is a quantifier-free interpretation of width one using only conjunction.

Interpretation from the CSP *to the forbidden patterns problem.* Let Π^{-1} be the interpretation from τ to σ given by reversing in a natural way the interpretation Π:

$$\psi_R := R(\bar{y}) \vee \bigvee_{\mathcal{F} \in \mathscr{F}} \bigvee_{R^{\mathcal{F}}(\bar{y}) \text{ holds}} \exists \tilde{x} R_{\mathcal{F}}(\tilde{x}, \bar{y}) \wedge \epsilon(\tilde{x}, \bar{y})$$

In the above sentence $\tilde{x}$ represent the elements of $\mathcal{F}$ not present among $\bar{y}$ and in $R_{\mathcal{F}}(\tilde{x}, \bar{y})$, the reader should understand that the variables $\tilde{x}, \bar{y}$ are reordered in a suitable fashion. The sentence ϵ is a conjunction of equalities between variables among $\tilde{x}, \bar{y}$.

By construction, $\Pi^{-1} \circ \Pi$ is the identity over σ-structures.

Construction of the template of the CSP. We build the τ-structure $\mathcal{T}_\Omega$ as an extension of the σ-structure $\mathcal{T}$ describing the colours of the forbidden patterns problem Ω. So on σ both structures agree and for every n-ary new symbol $R_{\mathcal{F}}$ and for every n-tuples of colours $c_1, c_2, \ldots, c_n$ we set $R_{\mathcal{F}}(c_1, c_2, \ldots, c_n)$ to hold unless,

- it is explicitly forbidden by a pattern $(\mathcal{F}, f)$ where $f(x_i) = c_i$; or,
- $\star$ the coloured structure $(\mathcal{F}, f)$ is implicitly forbidden by $(\mathcal{G}, g)$ in $\mathscr{F}'$ where $\mathcal{G}$ is a substructure of $\mathcal{F}$ and g agrees with $f(x_i) = c_i$ where defined[6].

Computational equivalence. By construction, the forbidden patterns problem Ω reduces to $\text{CSP}(-, \mathcal{T}_\Omega)$ via the interpretation Π. The converse interpretation Π^{-1} is not a reduction in general. It is a reduction for the τ-structures that will "not change too much" under $\Pi \circ \Pi^{-1}$. More formally, let $\mathcal{B}$ be the image of a τ-structure $\mathcal{A}$ under $\Pi \circ \Pi^{-1}$. The monotonic nature of the interpretations means that $\mathcal{A}$ is necessarily a substructure of $\mathcal{B}$ and that we only need to show that if $\mathcal{A}$ is a yes-instance then so is $\mathcal{B}$. The colouring certificate for $\mathcal{A}$ will validate $\mathcal{B}$ provided that if a new tuple involving $R_{\mathcal{G}}$ appeared in $\mathcal{B}$ it is a consequence of a larger tuple $R_{\mathcal{F}}$ where $\mathcal{F}$ and $\mathcal{G}$ are patterns in $\mathscr{F}$ and $\mathcal{G}$ is a substructure of $\mathcal{F}$. This holds because of the condition $\star$ in the construction of $\mathcal{T}_\Omega$.

In particular, we can guarantee that $\Pi \circ \Pi^{-1}$ will not change too much a τ-structure $\mathcal{A}$ if it is of sufficiently high girth, say a girth higher than γ, the number of elements of the largest pattern in $\mathscr{F}$ (this is because all patterns in $\mathscr{F}$ are biconnected). This proves that Π^{-1} is a reduction for instances of girth greater or equal to γ.

Note that we may extend $\mathscr{F}$ with any biconnected σ-structure without affecting the constructions or the result. This concludes the proof. □

[6] This second case $\star$ allows to channel constraints from one symbol in τ to another as all information regarding the relationship between the forbidden patterns is lost in the new signature τ.

4 Recolouring Captures Containment

This section is a proof of our main result (Theorem 8).

Let $\mathscr{F}$ be the set of biconnected structures involved as patterns in both Ω_1 and Ω_2. Let γ be the size of the largest structure in $\mathscr{F}$. We use Theorem 11 for each problem, using $\mathscr{F}$ as a parameter, and obtain a τ-structure $\mathcal{T}_{\Omega_1}$ for Ω_1 and a τ-structure $\mathcal{T}_{\Omega_2}$ for Ω_2.

Lemma 13. *If Ω_1 is contained in Ω_2 then* $\text{CSP}(girth > \gamma, \mathcal{T}_{\Omega_1})$ *is contained in* $\text{CSP}(girth > \gamma, \mathcal{T}_{\Omega_2})$.

Proof. Let $\mathcal{A}$ be a τ-structure of girth greater than γ such that there is a homomorphism from $\mathcal{A}$ to $\mathcal{T}_{\Omega_1}$. Since Π^{-1} is a reduction to Ω_1, it follows that $\Pi^{-1}(\mathcal{A})$ is a yes-instance of Ω_1. By inclusion of Ω_1 in Ω_2 it follows that $\Pi^{-1}(\mathcal{A})$ is also a yes-instance of Ω_2. Since Π is a reduction from Ω_2 to $\text{CSP}(-, \mathcal{T}_{\Omega_2})$, the structure $\mathcal{B} := \Pi \circ \Pi^{-1}(\mathcal{A})$ is a yes-instance of $\text{CSP}(-, \mathcal{T}_{\Omega_2})$. Hence, there is a homomorphism from $\mathcal{B}$ to $\mathcal{T}_{\Omega_2}$. Since $\mathcal{A}$ is a substructure of $\mathcal{B}$ by monotonicity of the interpretations, it follows that there is a homomorphism from $\mathcal{A}$ to $\mathcal{T}_{\Omega_2}$. ☐

Using Erdös Lemma we will derive the following.

Lemma 14. *The following are equivalent.*

(i) $\text{CSP}(girth > \gamma, \mathcal{T}_{\Omega_1})$ *is contained in* $\text{CSP}(girth > \gamma, \mathcal{T}_{\Omega_2})$.
(ii) $\text{CSP}(-, \mathcal{T}_{\Omega_1})$ *is contained in* $\text{CSP}(-, \mathcal{T}_{\Omega_2})$.
(iii) There is a homomorphism from $\mathcal{T}_{\Omega_1}$ to $\mathcal{T}_{\Omega_2}$.

Proof. The equivalence between (*ii*) and (*iii*) is easy and well known. The implication from (*ii*) to (*i*) holds trivially.

We prove that (*i*) implies (*iii*). Let $\mathcal{D}$ be the structure obtained from Erdös Lemma from $\mathcal{B} := \mathcal{T}_{\Omega_1}$ with $s := |\mathcal{T}_{\Omega_2}|$ and $g := \gamma$. We know that there is a homomorphism from $\mathcal{D}$ of girth greater than γ to $\mathcal{T}_{\Omega_1}$. It follows from our assumption (*i*) that there is also a homomorphism from $\mathcal{D}$ to $\mathcal{T}_{\Omega_2}$. Appealing to Erdös Lemma again for $\mathcal{C} := \mathcal{T}_{\Omega_1}$, we finally have that there is a homomorphism from $\mathcal{B} = \mathcal{T}_{\Omega_1}$ to $\mathcal{C} = \mathcal{T}_{\Omega_2}$. ☐

Lemma 15. *If r is a homomorphism from $\mathcal{T}_{\Omega_1}$ to $\mathcal{T}_{\Omega_2}$ then r is a recolouring from Ω_1 to Ω_2.*

Proof. Recall that $\mathcal{T}_1$ (respectively $\mathcal{T}_2$) the structure used to colour the forbidden patterns of Ω_1 (respectively Ω_2) is by construction the σ-reduct of $\mathcal{T}_{\Omega_1}$ (respectively, $\mathcal{T}_{\Omega_2}$). Hence, r is readily a homomorphism from $\mathcal{T}_{\Omega_1}$ to $\mathcal{T}_{\Omega_2}$.

It remains to show that for any $\mathcal{T}_2$-coloured pattern $(\mathcal{F}_2, f_2)$ forbidden by Ω_2, any of its inverse image under r– that is a $\mathcal{T}_1$-coloured structure $(\mathcal{F}_2, f_1)$ such that $f_2 = r \circ f_1$– is not valid w.r.t. Ω_1. Let $(\mathcal{F}_2, f_2)$ and $(\mathcal{F}_2, f_1)$ be as above. By construction of $\mathcal{T}_{\Omega_2}$, the tuple $R_{\mathcal{F}_2}(f_2(\bar{x}))$ does not hold in $\mathcal{T}_{\Omega_2}$. Since r is a homomorphism such that $f_2 = r \circ f_1$, the tuple $R_{\mathcal{F}_2}(f_1(\bar{x}))$ does not hold in $\mathcal{T}_{\Omega_1}$. By construction of $\mathcal{T}_{\Omega_1}$, this is because either a coloured pattern $(\mathcal{G}_1, g_1)$ forbidden by Ω_1 with pattern $\mathcal{F}_2$ or a substructure of $\mathcal{F}_2$ disallowed this tuple. In any case, we have that $(\mathcal{G}_1, g_1)$, which is forbidden by Ω_1 occurs in $(\mathcal{F}_2, f_1)$. This shows that $(\mathcal{F}_2, f_1)$ is not valid w.r.t. Ω_1. ☐

Our main result follows directly from the three previous lemmas.

Proof (of the main result). The definition of a recolouring implies containment as proved in Proposition 3. We now prove the converse. Suppose that Ω_1 is contained in Ω_2. By Lemma 13, it follows that CSP(girth $> \gamma, \mathcal{T}_{\Omega_1}$) is contained in CSP(girth $> \gamma, \mathcal{T}_{\Omega_2}$). By Lemma 15, it follows that there is a homomorphism r from $\mathcal{T}_{\Omega_1}$ to $\mathcal{T}_{\Omega_2}$. Finally, by Lemma 15 it follows that r is a recolouring from Ω_1 to Ω_2. □

We can strengthen our main result by relaxing some hypothesis as follows[7].

Corollary 16. *Let Ω_1 and Ω_2 be two forbidden patterns problems over the relational signature σ. If Ω_1 is given in a form that satisfies properties $\mathfrak{p}_1$ to $\mathfrak{p}_5$ then Ω_1 is contained in Ω_2 if, and only if, there is a recolouring from Ω_1 to Ω_2.*

5 Closing remarks

Feder and Vardi argued that MMSNP containment is decidable [1]. However a precise complexity was not given. Every sentence of MMSNP captures a finite union of forbidden patterns problems [3] so this motivates us to reformulate the question in terms of forbidden patterns problems.

FPP-**Containment:**

- Input: forbidden patterns problems Ω_1 and Ω_2 given by $(\mathcal{T}_1, \mathscr{F}'_1)$ and $(\mathcal{T}_2, \mathscr{F}'_2)$.
- Question: is Ω_1 contained in Ω_2?

It is not difficult to see that the problem is at least NP-hard. Indeed, in the restricted case when the problems have no forbidden patterns, we have in fact the CSP-containment problem (also known as the uniform constraint satisfaction problem) which is NP-complete. In the restricted case when Ω_1 is given by a representation $(\mathcal{T}_1, \mathscr{F}'_1)$ which satisfies properties $\mathfrak{p}_1$ to $\mathfrak{p}_5$, the question is equivalent to the following decision problem (see Corollary 16).

Recolouring:

- Input: forbidden patterns problems Ω_1 and Ω_2 given by $(\mathcal{T}_1, \mathscr{F}'_1)$ and $(\mathcal{T}_2, \mathscr{F}'_2)$.
- Question: is there a recolouring from Ω_1 to Ω_2?

The complexity of this problem is at most in Σ_3^p . This third level of the polynomial hierarchy is obtained directly from the definition of a recolouring. *Guess a homomorphism r, for every inverse image of every forbidden pattern, guess that it is non valid.* There are not many known complete problems in the third level of the polynomial hierarchy to choose from. There are however a myriad of problems in the second level. Using **Generalised Graph Colouring** [16] we can easily show that.

Proposition 17. *The restriction of **Recolouring** where Ω_2 has a single colour is Π_2^p-complete. Consequently, **Recolouring** is Π_2^p-hard.*

[7] Note that this is the best we can do as we may not do without property $\mathfrak{p}_5$ as example 7 shows.

In future work, we will try to pinpoint more accurately the complexity of **Recolouring**, which should be complete for Σ_3^p. Our hope is that a suitable generalisation of recolouring will enable us to derive that the complexity of FPP-**Containment** and **Recolouring** are the same. The long term aim is to classify the complexity of MMSNP containment. Though the translation to FPP is exponential, we hope that the insight gained in the combinatorial world of forbidden patterns problems can be used to solve the problem in the logical world of MMSNP.

References

1. 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, 57–104 (1999); An initial version of this paper appeared in STOC 1993
2. Bodirsky, M.: Constraint satisfaction problems with infinite templates. In: Complexity of Constraints, pp. 196–228 (2008)
3. Madelaine, F., Stewart, I.A.: Constraint satisfaction, logic and forbidden patterns. SIAM Journal on Computing 37(1), 132–163 (2007)
4. Ladner, R.E.: On the structure of polynomial time reducibility. J. Assoc. Comput. Mach. 22, 155–171 (1975)
5. Bulatov, A.A., Valeriote, M.: Recent results on the algebraic approach to the CSP. In: Complexity of Constraints, pp. 68–92 (2008)
6. Fagin, R.: Generalized first-order spectra and polynomial-time recognizable sets. In: Karp, R.M. (ed.) Complexity of Computation, pp. 43–73 (1974)
7. Kun, G.: Constraints, MMSNP and expander relational structures. arXiv:0706.1701v1 (June 2007)
8. Madelaine, F., Stewart, I.A.: Some problems not definable using structures homomorphisms. Ars Combinatoria LXVII (2003)
9. Madelaine, F.: Constraint satisfaction problems and related logic. PhD thesis, University of Leicester, Department of Maths and Computer Science (2003)
10. Bodirsky, M., Dalmau, V.: Datalog and constraint satisfaction with infinite templates. In: Durand, B., Thomas, W. (eds.) STACS 2006. LNCS, vol. 3884, pp. 646–659. Springer, Heidelberg (2006)
11. Madelaine, F.R.: Universal structures and the logic of forbidden patterns. Logical Methods in Computer Science 5(2) (2009)
12. Bodirsky, M., Chen, H., Feder, T.: On the complexity of MMSNP (November 2009)
13. Hell, P., Nešetřil, J.: Graphs and homomorphisms. Oxford University Press, Oxford (2004)
14. Cherlin, G., Shelah, S., Shi, N.: Universal graphs with forbidden subgraphs and algebraic closure. Adv. Appl. Math. 22(4), 454–491 (1999)
15. Foniok, J., Nesetril, J., Tardif, C.: Generalised dualities and maximal finite antichains in the homomorphism order of relational structures. Eur. J. Comb. 29(4), 881–899 (2008)
16. Schaefer, M., Umans, C.: Completeness in the polynomial-hierarchy: part I. SIGACT news 33(3), 32–49 (2002); SIGACT news complexity theory column 37, Guest column, introduced by Lane A. Hemaspaandra