

Containment for Guarded Monotone Strict NP

Alexey Barsukov 

Charles University, Prague, Czech Republic

Michael Pinsker 

Technische Universität Wien, Austria

Jakub Rydval 

Technische Universität Wien, Austria

Abstract

Guarded Monotone Strict NP (GMSNP) extends Monotone Monadic Strict NP (MMSNP) by guarded existentially quantified predicates of arbitrary arities. We prove that the containment problem for GMSNP is decidable, thereby settling an open question of Bienvenu, ten Cate, Lutz, and Wolter, later restated by Bourhis and Lutz. Our proof also comes with a 2NEXPTIME upper bound on the complexity of the problem, which matches the lower bound for containment of MMSNP due to Bourhis and Lutz. In order to obtain these results, we significantly improve the state of knowledge of the model-theoretic properties of GMSNP. Bodirsky, Knäuer, and Starke previously showed that every GMSNP sentence defines a finite union of CSPs of ω -categorical structures. We show that these structures can be used to obtain a reduction from the containment problem for GMSNP to the much simpler problem of testing the existence of a certain map called recolouring, albeit in a more general setting than GMSNP; a careful analysis of this yields said upper bound. As a secondary contribution, we refine the construction of Bodirsky, Knäuer, and Starke by adding a restricted form of homogeneity to the properties of these structures, making the logic amenable to future complexity classifications for query evaluation using techniques developed for infinite-domain CSPs.

2012 ACM Subject Classification Theory of computation \rightarrow Logic; Theory of computation \rightarrow Computational complexity and cryptography

Keywords and phrases guarded, monotone, SNP, forbidden patterns, query containment, recolouring, decidability, computational complexity, ω -categoricity, constraint satisfaction, homogeneity, amalgamation property, Ramsey property, canonical function

Digital Object Identifier 10.4230/LIPIcs.ICALP.2025.140

Category Track B: Automata, Logic, Semantics, and Theory of Programming

Related Version Full Version: <https://arxiv.org/abs/2310.01254> [2]

Funding This research was funded in whole or in part by the Austrian Science Fund (FWF) [P 32337, I 5948]. For the purpose of Open Access, the authors have applied a CC BY public copyright licence to any Author Accepted Manuscript (AAM) version arising from this submission. This research is also funded by the European Union (ERC, POCOCOP, 101071674). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

Acknowledgements The first author is thankful to Florent Madelaine for making him aware of the containment problem for GMSNP.

1 Introduction

1.1 The logic GMSNP

Guarded Monotone Strict NP (GMSNP) is a syntactic fragment of existential second-order logic describing problems of the form

Is there a colouring of the relational tuples in a given finite relational structure avoiding a fixed set of finitely many forbidden colour-patterns?



© Alexey Barsukov, Michael Pinsker, and Jakub Rydval;
licensed under Creative Commons License CC-BY 4.0

52nd International Colloquium on Automata, Languages, and Programming (ICALP 2025).

Editors: Keren Censor-Hillel, Fabrizio Grandoni, Joël Ouaknine, and Gabriele Puppis

Article No. 140; pp. 140:1–140:20



Leibniz International Proceedings in Informatics

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany



The logic GMSNP was introduced in [6] as a generalisation of the logic *Monotone Monadic Strict NP without inequality* (MMSNP) of Feder and Vardi [19], which captures similar problems concerning colourings of vertices instead of relational tuples. However, its true origins actually go back to an earlier work of Madelaine [28], who was studying GMSNP under the names MMSNP₂ and FPP (standing for *Forbidden Pattern Problems*).

Formally, a GMSNP sentence Φ is of the form $\exists X_1, \dots, X_n \forall \bar{x}. \phi(\bar{x})$, for a CNF-formula ϕ over a finite relational signature $\tau \cup \{X_1, \dots, X_n\}$, satisfying the following two conditions:

- τ -atoms and equalities may only appear negatively in clauses of ϕ , i.e., the only positive atoms in clauses of ϕ are of the form $X_i(\bar{x})$ for some $i \in [n]$; (*Monotonicity*)
- for every clause ψ of ϕ and every positive atom $X_i(\bar{x})$ in ψ there exists a negative atom $\neg R(\bar{y})$ in ψ with $R \in \tau \cup \{X_1, \dots, X_n\}$ such that $\bar{x} \subseteq \bar{y}$. (*Guarding*)

Unless stated otherwise, we will denote the set $\{X_1, \dots, X_n\}$ of the second-order variables in Φ by σ ; we refer to σ as the *existential* symbols and τ as the *input* symbols in Φ . We also refer to the first-order $(\tau \cup \sigma)$ -sentence $\forall \bar{x}. \phi(\bar{x})$ as the *first-order part* of Φ . By a simple application of De Morgan's laws, we can rewrite ϕ as a formula of the form $\bigwedge_i \neg \phi_i$, where each ϕ_i is a conjunction of τ -atoms and possibly negated σ -atoms. We refer to the conjuncts ϕ_i as the *forbidden patterns* of Φ .

A typical example of a problem that can be defined in GMSNP is whether a given undirected simple graph, say with edge relation denoted by E , can be edge-2-coloured while avoiding monochromatic triangles [21]:

$$\begin{aligned} \exists B, R \forall x, y, z (&\neg E(x, y) \vee \neg E(y, z) \vee \neg E(z, x) \vee \neg B(x, y) \vee \neg B(y, z) \vee \neg B(z, x)) \\ &\wedge (\neg E(x, y) \vee \neg E(y, z) \vee \neg E(z, x) \vee \neg R(x, y) \vee \neg R(y, z) \vee \neg R(z, x)) \\ &\wedge (\neg E(x, y) \vee B(x, y) \vee R(x, y)) \wedge (\neg E(x, y) \vee \neg B(x, y) \vee \neg R(x, y)). \end{aligned} \quad (1)$$

See Figure 1 for illustration.



■ **Figure 1** An edge-2-colouring of the complete graph on 5 vertices avoiding monochromatic triangles (outer edges red, inner blue). Note that there is no vertex-2-colouring with this property.

The forbidden patterns of this problem are the two monochromatic triangles, overlapping colours, and uncoloured edges. In MMSNP, we can formulate a similar problem where the task is to find a vertex-2-colouring instead of an edge-2-colouring:

$$\begin{aligned} \exists B, R \forall x, y, z (&\neg E(x, y) \vee \neg E(y, z) \vee \neg E(z, x) \vee \neg B(x) \vee \neg B(y) \vee \neg B(z)) \\ &\wedge (\neg E(x, y) \vee \neg E(y, z) \vee \neg E(z, x) \vee \neg R(x) \vee \neg R(y) \vee \neg R(z)) \\ &\wedge (\neg B(x) \vee \neg R(x)) \wedge (B(x) \vee R(x)). \end{aligned} \quad (2)$$

Strictly speaking, GMSNP does *not* fully contain MMSNP because the latter does not require the *guarding* axiom (it instead requires the *monadicity* axiom). However, the containment does hold up to a small syntactic transformation, modulo which GMSNP is in fact strictly more expressive than MMSNP [6]. Namely, every MMSNP sentence Φ can be converted into a GMSNP sentence $\Gamma(\Phi)$ by expanding the base signature τ by a fresh unary “domain” symbol D and adding negative atoms $\neg D(v)$ to each clause, for every universally quantified

first-order variable v appearing in that clause. Given a τ -structure \mathfrak{A} , we have that Φ holds in \mathfrak{A} if and only if $\Gamma(\Phi)$ holds in \mathfrak{A}^+ , where \mathfrak{A}^+ is the $(\tau \cup \{D\})$ -structure obtained from \mathfrak{A} by including every domain element in the relation interpreting D . On the other hand, given a $(\tau \cup \{D\})$ -structure \mathfrak{A} , we have that $\Gamma(\Phi)$ holds in \mathfrak{A} if and only if Φ holds in \mathfrak{A}^- , where \mathfrak{A}^- is the τ -structure obtained from \mathfrak{A} by forgetting D and removing all domain elements which were not contained in the relation interpreting D . For example, the last two clauses in (2) do not satisfy the guarding axiom, but we can obtain a polynomial-time equivalent GMSNP sentence by padding (2) with the predicate $D \in \tau$:

$$\begin{aligned} & \exists B, R \forall x, y, z \\ & (\neg D(x) \vee \neg D(y) \vee \neg D(z) \vee \neg E(x, y) \vee \neg E(y, z) \vee \neg E(z, x) \vee \neg B(x) \vee \neg B(y) \vee \neg B(z)) \\ & \wedge (\neg D(x) \vee \neg D(y) \vee \neg D(z) \vee \neg E(x, y) \vee \neg E(y, z) \vee \neg E(z, x) \vee \neg R(x) \vee \neg R(y) \vee \neg R(z)) \\ & \wedge (\neg D(x) \vee B(x) \vee R(x)) \wedge (\neg D(x) \vee \neg B(x) \vee \neg R(x)). \end{aligned}$$

1.2 A brief history of GMSNP

In 2014, Bienvenu, ten Cate, Lutz, and Wolter [6] discovered an interesting correspondence that allows one to translate complexity classification results for certain well-behaved fragments of the logic SNP [26, 34] to analogous statements about the complexity of ontology-mediated queries. In the special case of MMSNP, the above-mentioned result from [6] implies that there is a dichotomy between P and NP-completeness if and only if there is a dichotomy between P and coNP-completeness for the evaluation of unions of conjunctive queries mediated by ontologies specified in the description logic \mathcal{ALC} [1]. The significance of this equivalence lies in the fact that the computational complexity of MMSNP had already been intensively investigated within the programme initiated by Feder and Vardi in 1998 [19].

Feder and Vardi proved that every problem described by an MMSNP sentence is equivalent under polynomial-time randomised reductions to the *Constraint Satisfaction Problem* (CSP) of a structure with a finite domain, i.e., the problem of testing whether a given conjunctive query is satisfiable in that structure. They moreover conjectured that every finite-domain CSP is in P or NP-complete, which was eventually confirmed twenty years later by Bulatov and Zhuk [17, 39] in what can be viewed as the culmination of decades of work on connections between algorithms for CSPs and universal algebra. This implied dichotomies for MMSNP and, in turn, for the above-mentioned queries up to randomised reductions. Although the originally randomised reduction of MMSNP to CSPs by Feder and Vardi was derandomised by Kun in 2013 [27], it still relied on a probabilistic proof of the existence of expanders of large girth. The first truly constructive proof of the dichotomy was obtained in 2018 by Bodirsky, Madelaine, and Mottet [12] using methods from infinite-domain constraint satisfaction, i.e., CSPs of infinite structures. Roughly speaking, they start with the observation from [9] that every MMSNP sentence defines a finite union of CSPs of (infinite) ω -categorical structures. They then gradually upgrade these structures until a state is reached where they can show that their CSPs admit a polynomial-time many-one reduction to a finite-domain CSP which is polynomial-time tractable unless one of the ω -categorical structures can simulate the 3-colouring problem via a *pp-construction* [4] (in which case the sentence defines a NP-complete problem).

The question whether GMSNP also exhibits a P/NP-dichotomy was left open in [6]; a positive answer would yield a P/coNP-dichotomy for the evaluation of unions of conjunctive queries mediated by ontologies specified in guarded first-order logic. It was observed by Bodirsky, Knäuer, and Starke [11] that similarly to MMSNP, the logic GMSNP can be studied using methods from infinite-domain constraint satisfaction. Using this approach,

the existence of a dichotomy has already been confirmed in some specific cases, e.g., for graph orientations with forbidden tournaments [10, 7, 20]. However, at the moment there is no clear road map for how exactly the question of the existence of a P/NP-dichotomy for GMSNP should be approached: it is apparent that the increase in the arity of the existentially quantified predicates introduces new obstacles which were not present with MMSNP. Hope is sparked by a recently announced result of Guzmán-Pro [22] who used the sparse-incomparability lemma [27] to show that the homomorphism-sandwiching method [16] for obtaining hardness reductions from finite-domain *promise* CSPs fails for CSPs definable in GMSNP. This is in fact a good plausibility argument for the applicability of methods from infinite-domain constraint satisfaction to GMSNP, because several NP-hardness results for finite-domain promise CSPs (Proposition 10.1 in [3] and Theorem 2.2 in [38]) are inconsistent with the algebraic dichotomy conjecture for infinite-domain CSPs [5].

1.3 The containment problem

Besides query evaluation, two other computational decision problems for ontology-mediated queries were studied in [6]: first-order/datalog rewritability and containment. As with their counterparts in automata theory, the complexity of such problems tends to be significantly higher. This raises the question of whether, in studying the complexity of these problems, one can perhaps avoid the fine-grained analysis that makes the proof of complexity dichotomies so intricate. In the case of MMSNP, the precise complexity of first-order/datalog rewritability and containment is known: both are 2NEXPTIME-complete [31, 32, 15]. While rewritability seems to be intimately linked to the reduction from MMSNP to finite-domain CSPs [31, 32], for the containment problem, the link seems to be much weaker [15, 12]. This makes the study of the containment problem for GMSNP a perfect starting point for the development of tools that will later be useful for approaching the topics of rewritability and query evaluation.

In the proof of our main result, it will be important that we formulate the containment problem in the more general setting of *SNP sentences*, which are obtained by leaving out the monotonicity and the guarding axioms. For an SNP sentence Φ , we denote the class of all finite models of Φ by $fm(\Phi)$. Formally, the *containment problem* can be stated as follows:

Containment for SNP

INSTANCE: A pair (Φ_1, Φ_2) of SNP sentences in a common signature τ .

QUESTION: Is it true that $fm(\Phi_1) \subseteq fm(\Phi_2)$?

The containment problem for SNP is undecidable in general; in fact, it is already undecidable for the *Datalog* fragment [37]. The question whether it is decidable for GMSNP was left open in [6, 15]; the only known result regarding the complexity of this problem is the lower bound of Bourhis and Lutz for the containment within MMSNP. To see that this lower bound also applies for GMSNP, note that the map Γ introduced in Section 1.1 is not only a polynomial-time reduction from MMSNP to GMSNP, but it also acts as a fully faithful covariant functor with respect to containment. More specifically, for a pair (Φ_1, Φ_2) of MMSNP τ -sentences, we have $fm(\Phi_1) \subseteq fm(\Phi_2)$ if and only if $fm(\Gamma(\Phi_1)) \subseteq fm(\Gamma(\Phi_2))$.

► **Theorem 1** (Theorem 3 in [15]). *Containment for GMSNP is 2NEXPTIME-hard.*

1.4 Contributions

The contributions of the present article are twofold. On the one hand, we confirm that the containment problem for GMSNP is indeed decidable. Our proof of decidability also comes with an upper bound on the complexity matching the lower bound in Theorem 1.

► **Theorem 2.** *The containment problem for GMSNP is in 2NEXPTIME.*

► **Corollary 3.** *The containment problem for GMSNP is 2NEXPTIME-complete.*

The proof of Theorem 2 is based on the recolouring method, first used implicitly by Feder and Vardi [19, Theorem 7] in their proof of decidability of containment for MMSNP.¹ Roughly said, the strategy is to modify the input sentences by adding new clauses and existentially quantified symbols until a state is reached where the containment problem becomes equivalent to the much simpler problem of testing the existence of a mapping between the existential symbols of the two sentences satisfying a certain condition (a recolouring). In contrast to MMSNP, in order to achieve the promised 2NEXPTIME upper bound on the complexity of containment for GMSNP using the recolouring method, we have to deviate from the GMSNP framework (to SNP) and significantly generalise the notion of a recolouring.

The first preprocessing step, common across most approaches to the containment problem for MMSNP, is a reduction to the connected case [12, 15, 19, 29]. Intuitively, an MMSNP sentence is *connected* if its forbidden colour-patterns are connected in the graph-theoretic sense. This reduction can be generalised to GMSNP, as shown in [11]. Bodirsky, Knäuer, and Starke [11, Proposition 1] showed that every GMSNP sentence is logically equivalent to a finite disjunction of connected GMSNP sentences. An inspection of their proof reveals that if the complexity of checking containment is in 2NEXPTIME for connected GMSNP sentences, then so it is for the entire class GMSNP.

The second preprocessing step (in the approaches to the containment for MMSNP) is typically a reduction to the biconnected case [15, 12]. Intuitively, an MMSNP sentence is *biconnected* if no forbidden colour-pattern can be disconnected by deleting a single vertex. The reduction to biconnected MMSNP is the point where a significant amount of work can be outsourced to a general combinatorial result, e.g., a lemma of Erdős in the case of [15] or a theorem of Hubička and Nešetřil in the case of [12]. The procedure for attaining biconnectedness from [12, Lemma 4.4] transforms a given connected MMSNP sentence to a logically equivalent biconnected MMSNP sentence by iteratively selecting a forbidden colour-pattern which can be disconnected by deleting a single vertex and splitting it into two forbidden colour-patterns (corresponding to the two parts); the split is then marked using a fresh unary existential predicate. The GMSNP-analogue of this procedure does not terminate if the arities of the existentially quantified predicates are greater than 2, and we do not see any way how this issue could be fixed. Therefore, to progress further, we must understand how exactly the said general combinatorial results are used in the complexity analysis of the containment problem; in the present paper, we focus specifically on the approach from [12] using the mentioned theorem of Hubička and Nešetřil.

Roughly said, the theorem implies that every connected MMSNP τ -sentence Φ can be assigned a highly symmetric infinite structure \mathfrak{C}_Φ in a signature extending $\tau \cup \sigma$ such that the $(\tau \cup \sigma)$ -reducts of the finite substructures of \mathfrak{C}_Φ are up to isomorphism precisely the finite models of the first-order part of Φ ; this was first observed by Bodirsky and Dalmau [9]. Bodirsky, Madelaine, and Mottet [12] proved that if Φ is biconnected, then the $(\tau \cup \sigma)$ -reduct of \mathfrak{C}_Φ retains a certain amount of the original symmetry, and this fact is crucial in their take on the recolouring method. Later on, Bodirsky, Knäuer, and Starke [11] showed that also connected GMSNP sentences can be assigned said highly symmetric infinite structures; they did not comment on whether this fact is helpful in analysing the complexity of the containment problem for GMSNP. We show that the highly symmetric infinite structures \mathfrak{C}_Φ associated

¹ Recolourings were first explicitly mentioned in the work of Madelaine and Stewart [30] (see also [29]).

with GMSNP sentences Φ can be described by SNP sentences $\Delta(\Phi)$ of size doubly exponential in the size of Φ . Subsequently, we show that the containment $fm(\Phi_1) \subseteq fm(\Phi_2)$ is equivalent to the existence of a particular mapping from \mathfrak{C}_{Φ_1} to \mathfrak{C}_{Φ_2} which is uniquely determined by the images of substructures whose size is bounded by the maximal arity of a symbol in Φ_1 or Φ_2 . From this fact, we extract our more general version of recolourings for SNP. By carefully analysing the total complexity of computing the SNP sentences $\Delta(\Phi_1)$ and $\Delta(\Phi_2)$ stemming from the theorem of Hubička and Nešetřil and of the subsequent check for the existence of a recolouring between such sentences, we conclude that the containment problem for GMSNP is in 2NEXPTIME. We remark that our methods are, in theory, applicable in a more general setting than GMSNP; a fragment of SNP which is closely related to such applicability is *Amalgamation SNP* introduced in [11]. We also remark that the SNP sentences $\Delta(\Phi_1)$ and $\Delta(\Phi_2)$ define a linear order over the domain of a given structure, which is at the core of our argument why the recolouring method works in our case. It is provably not possible to define a linear order in GMSNP [11, Example 6], which indicates that this method requires us to leave the GMSNP framework.

As our secondary contribution, we analyse to what extent it is possible to prove the decidability of the containment problem for GMSNP while staying as close as possible to the original recolouring method for MMSNP. A (vertex)-recolouring ξ between two MMSNP sentences Φ_1 and Φ_2 is simply a mapping from the (unary) existential symbols of Φ_1 to those of Φ_2 whose application to the relations denoted by these symbols of any model of the first-order part of Φ_1 yields a model of the first-order part of Φ_2 . That is, the mapping ξ does not introduce any vertex-colour patterns forbidden by Φ_2 into structures which originally did not contain any patterns forbidden by Φ_1 (see [12, Definition 4.25]). If Φ_1 and Φ_2 assert that the existential relations form a partition of all vertices (cf. the notion of *normal form* in [12]), then this is what one would intuitively understand under a recolouring of vertices – hence the name. But the true intention of this concept, which does not require any additional assumptions on Φ_1 and Φ_2 (such as the vertices being partitioned by the existential predicates) is to capture a very specific case of the containment $fm(\Phi_1) \subseteq fm(\Phi_2)$, which is witnessed uniformly across all structures in $fm(\Phi_1)$. The natural extension of vertex-recolourings to GMSNP would be *edge-recolourings*, i.e., mappings

$$\xi: R(x_1, \dots, x_n) \wedge \alpha_1(x_1, \dots, x_n) \mapsto R(x_1, \dots, x_n) \wedge \alpha_2(x_1, \dots, x_n), \quad (3)$$

where $R \in \tau$ and α_i is an atomic σ_i -formula ($i \in [2]$), such that ξ induces a mapping between models of the first-order parts of the two sentences. The notion of a recolouring for SNP we use in the proof of Theorem 2 is considerably more general and abstract. Hence, it makes sense to ask whether one can also prove decidability of containment for GMSNP by simply refining the input sentences Φ_1 and Φ_2 long enough, while staying in GMSNP, until the containment $fm(\Phi_1) \subseteq fm(\Phi_2)$ becomes equivalent to the existence of an edge-recolouring from Φ_1 to Φ_2 .

The central notion in analysing this possibility is *recolouring-readiness*, which is comparable to the notion of a *simple program* of Bourhis and Lutz [15] or the normal form of Bodirsky, Madelaine, and Mottet [12]. Instead of relying purely on syntactic preprocessing, in the definition of recolouring-readiness, we draw our inspiration from the proof of Theorem 2. Intuitively, a GMSNP τ -sentence Φ is recolouring-ready if it can be assigned a highly symmetric infinite structure \mathfrak{C}_Φ as in our proof of Theorem 2 such that the $(\tau \cup \sigma)$ -reduct of \mathfrak{C}_Φ retains a sufficient amount of the original symmetry for a reduction from containment to edge-recolouring to work. We show that every connected GMSNP sentence Φ is logically equivalent to a connected recolouring-ready GMSNP sentence $\Omega(\Phi)$. However, the size of

the smallest such $\Omega(\Phi)$ that we are able to obtain surpasses the upper bound provided in Theorem 2. We leave it as an open question whether the precise complexity of containment for GMSNP can be determined purely from within the GMSNP framework.

► **Theorem 4.** *For every connected GMSNP τ -sentence Φ there exists a logically equivalent recolouring-ready connected GMSNP τ -sentence $\Omega(\Phi)$ of size triple-exponential in the size of Φ such that the maximal arity of the symbols in Φ and in $\Omega(\Phi)$ is the same. Moreover, $\Omega(\Phi)$ can be computed from Φ in nondeterministic triple-exponential time.*

1.5 Outline

In Section 2, we provide some background knowledge necessary for the presentation of our results. Section 3 contains the proof of Theorem 2, where the individual steps are presented in the order given in the introduction. Section 4 contains a proof sketch for Theorem 4.

2 Preliminaries

We use the bar notation for tuples, and the set $\{1, \dots, n\}$ is denoted by $[n]$. We extend the containment relation on sets to tuples by ignoring the ordering on the entries. For example, we might write $X \subseteq \bar{t}$ for a set X and a tuple \bar{t} .

Structures. A (relational) signature τ is a set of relation symbols, each $R \in \tau$ is associated with a natural number called *arity*. A (relational) τ -structure \mathfrak{A} consists of a set A (the *domain*) together with the relations $R^{\mathfrak{A}} \subseteq A^k$ for each $R \in \tau$ with arity k . An *expansion* of \mathfrak{A} is a σ -structure \mathfrak{B} with $A = B$ such that $\tau \subseteq \sigma$ and $R^{\mathfrak{B}} = R^{\mathfrak{A}}$ for each relation symbol $R \in \tau$. Conversely, we then call \mathfrak{A} a *reduct* of \mathfrak{B} . We denote the reduct of \mathfrak{B} to a subset τ of its signature by \mathfrak{B}^τ . A *linear-order expansion* of a τ -structure \mathfrak{A} is an expansion by a single linear order denoted by $<$ (assuming $< \notin \tau$); we denote such expansion by $(\mathfrak{A}, <)$. We often do not distinguish between the symbol $<$ and the associated linear order. Given a class of τ -structures \mathcal{K} , we denote by $\mathcal{K}^<$ the class of all $\tau \cup \{<\}$ -structures which are linear-order expansions of structures from \mathcal{K} . The *union* of two τ -structures \mathfrak{A} and \mathfrak{B} is the τ -structure $\mathfrak{A} \cup \mathfrak{B}$ with domain $A \cup B$ and relations $R^{\mathfrak{A} \cup \mathfrak{B}} := R^{\mathfrak{A}} \cup R^{\mathfrak{B}}$ for every $R \in \tau$. A structure is *connected* if it is not the union of two structures with disjoint non-empty domains.

A *homomorphism* $h: \mathfrak{A} \rightarrow \mathfrak{B}$ for τ -structures $\mathfrak{A}, \mathfrak{B}$ is a mapping $h: A \rightarrow B$ that *preserves* each relation of τ , i.e., whenever $\bar{t} \in R^{\mathfrak{A}}$ for some relation symbol $R \in \tau$, then $h(\bar{t})$ (computed componentwise) is an element of $R^{\mathfrak{B}}$. We write $\mathfrak{A} \rightarrow \mathfrak{B}$ if \mathfrak{A} maps homomorphically into \mathfrak{B} . The *Constraint Satisfaction Problem* (CSP) of \mathfrak{A} , denoted by $\text{CSP}(\mathfrak{A})$, is defined as the class of all finite structures which homomorphically map into \mathfrak{A} . An *embedding* is an injective homomorphism $h: \mathfrak{A} \rightarrow \mathfrak{B}$ that additionally satisfies the following condition: for every k -ary relation symbol $R \in \tau$ and $\bar{t} \in A^k$ we have $h(\bar{t}) \in R^{\mathfrak{B}}$ only if $\bar{t} \in R^{\mathfrak{A}}$. We write $\mathfrak{A} \hookrightarrow \mathfrak{B}$ if \mathfrak{A} embeds into \mathfrak{B} . The *age* of \mathfrak{A} , denoted by $\text{age}(\mathfrak{A})$, is the class of all finite structures which embed into \mathfrak{A} . A *substructure* of \mathfrak{A} is a structure \mathfrak{B} over $B \subseteq A$ such that the inclusion map $i: B \rightarrow A$ is an embedding. An *isomorphism* is a surjective embedding. A *partial isomorphism* on a structure \mathfrak{A} is an isomorphism between two substructures of \mathfrak{A} . Two structures \mathfrak{A} and \mathfrak{B} are *isomorphic* if there exists an isomorphism from \mathfrak{A} to \mathfrak{B} . An *automorphism* of \mathfrak{A} is an isomorphism from \mathfrak{A} to itself. The set of all automorphisms of \mathfrak{A} is denoted by $\text{Aut}(\mathfrak{A})$.

The *orbit* of a tuple $\bar{t} \in B^k$ in \mathfrak{B} is the set $\{g(\bar{t}) \mid g \in \text{Aut}(\mathfrak{B})\}$. For two structures \mathfrak{A} and \mathfrak{B} , a function $f: A \rightarrow B$ is called *canonical from \mathfrak{A} to \mathfrak{B}* if, for every $k \geq 1$, the componentwise action of f induces a well-defined function from the orbits of k -tuples in \mathfrak{A} to the orbits of k -tuples in \mathfrak{B} ; that is, if any two k -tuples belonging to the same orbit, say O_1 , are mapped into the same orbit, say O_2 ; we can then write $f(O_1) = O_2$. In other words, for every $k \geq 1$, every $\bar{t} \in A^k$, and every $\alpha \in \text{Aut}(\mathfrak{A})$, there exists $\beta \in \text{Aut}(\mathfrak{B})$ such that $f(\alpha(\bar{t})) = \beta(f(\bar{t}))$. A countable structure \mathfrak{B} is ω -categorical if, for every $k \geq 1$, there are only finitely many orbits of k -tuples in \mathfrak{B} . The following lemma can be shown by a standard compactness argument, e.g., using König's tree lemma.

► **Lemma 5** (Lemma 4.1.7 in [8]). *Let \mathfrak{A} and \mathfrak{B} be countable relational structures such that \mathfrak{B} is ω -categorical. Then $\mathfrak{A} \hookrightarrow \mathfrak{B}$ if and only if every finite substructure of \mathfrak{A} embeds into \mathfrak{B} .*

Logic. We assume that the reader is familiar with classical *first-order* logic as well as with basic preservation properties of first-order formulas, e.g., that every first-order formula ϕ is preserved by isomorphisms; by embeddings if ϕ is existential, and by homomorphisms if ϕ is existential positive. We assume that equality $=$ is always available when building first-order formulas. We say that a first-order formula ϕ is k -ary if it has k free variables; we use the notation $\phi(\bar{x})$ to indicate that the free variables of ϕ are among \bar{x} . This does not mean that the truth value of ϕ depends on each entry in \bar{x} .

In the present article, *atomic τ -formulas*, or τ -atoms for short, over a relational signature τ are of the form $R(\bar{x})$ for some $R \in \tau$ and a tuple \bar{x} of first-order variables matching the arity of R . For technical reasons, formulas of the form $x = y$ built using the default equality predicate are not considered atomic. For a finite conjunction ϕ of τ -atoms, the *canonical database* of ϕ is the structure \mathfrak{A} whose domain consists of the variables of ϕ and such that $\mathfrak{A} \models R(\bar{t})$ holds if and only if ϕ contains the τ -atom $R(\bar{t})$ as a conjunct. The canonical database of ϕ admits a homomorphism to a τ -structure \mathfrak{B} if and only if ϕ is satisfiable in \mathfrak{B} [18]. For a finite τ -structure \mathfrak{A} , the *canonical query* of \mathfrak{A} is defined as the formula $\bigwedge_{R \in \tau} \bigwedge_{\bar{t} \in R^{\mathfrak{A}}} R(\bar{t})$. The canonical query of \mathfrak{A} is satisfiable in a τ -structure \mathfrak{B} if and only if there exists a homomorphism from \mathfrak{A} to \mathfrak{B} [18].

Structural Ramsey theory. A relational structure \mathfrak{B} is *homogeneous* if, for every $k \geq 1$, two tuples $\bar{t}_1, \bar{t}_2 \in B^k$ lie in the same orbit if and only if the function mapping the i -th coordinate in \bar{t}_1 to the i -th coordinate in \bar{t}_2 for all $1 \leq i \leq k$ is an isomorphism between two substructures of \mathfrak{B} . In other words, a structure is homogeneous if every partial isomorphism between two finite substructures extends to an automorphism of the entire structure. Every homogeneous structure over a finite relational signature is ω -categorical.

Homogeneous structures arise as limit objects of certain classes of finite structures. Let \mathcal{K} be a class of finite structures in a finite relational signature τ closed under isomorphisms and substructures. We say that \mathcal{K} has the *amalgamation property* (AP) if, for all $\mathfrak{A}, \mathfrak{B} \in \mathcal{K}$ whose substructures induced on $A \cap B$ are identical, there exist $\mathfrak{W} \in \mathcal{K}$ and embeddings $e: \mathfrak{A} \hookrightarrow \mathfrak{W}$, $f: \mathfrak{B} \hookrightarrow \mathfrak{W}$ such that $e|_{A \cap B} = f|_{A \cap B}$.

► **Theorem 6** (Fraïssé, Theorem 6.1.2 in [23]). *For a class \mathcal{K} of finite structures in a finite relational signature τ , the following are equivalent:*

- \mathcal{K} is the age of a countable homogeneous τ -structure; this structure is necessarily unique up to isomorphism and called the *Fraïssé-limit* of \mathcal{K} ;
- \mathcal{K} is closed under isomorphisms, substructures, and has the AP.

For structures \mathfrak{A} and \mathfrak{W} , we denote by $\left(\frac{\mathfrak{W}}{\mathfrak{A}}\right)$ the set of all embeddings of \mathfrak{A} into \mathfrak{W} . A class \mathcal{K} of structures over a common signature τ has the *Ramsey property* (RP) if, for all $\mathfrak{A}, \mathfrak{B} \in \mathcal{K}$

and $k \in \mathbb{N}$, there exists $\mathfrak{W} \in \mathcal{K}$ such that, for every map $f: \binom{\mathfrak{W}}{\mathfrak{A}} \rightarrow [k]$, there exists $e \in \binom{\mathfrak{W}}{\mathfrak{B}}$ such that f is constant on the set $\{e \circ u \mid u \in \binom{\mathfrak{B}}{\mathfrak{A}}\} \subseteq \binom{\mathfrak{W}}{\mathfrak{A}}$. Following [12, Definition 5.3], in the present article, we call a (countable) ω -categorical structure \mathfrak{B} *Ramsey* if the age of the expansion of \mathfrak{B} by all first-order definable relations (which is always a homogeneous structure) has the Ramsey property. By the theorem of Engeler, Ryll-Nardzewski and Svenonius [23], this property only depends on the automorphism group $\text{Aut}(\mathfrak{B})$. It is not hard to see that every homogeneous ω -categorical structure whose age has the RP is Ramsey, since all of its first-order definable relations are in fact definable without quantifiers.

In the proof of Theorem 2, we employ structural Ramsey theory to effectively reduce the problem of containment between two infinite classes of finite structures to the existence of a certain computable function between two infinite structures. This reduction is achieved by first representing the two classes by infinite structures via Theorem 6 above, and then ensuring the computability of the function comparing the two infinite structures (if it exists) using the following statement, an immediate consequence of Theorem 5 in [13].

► **Theorem 7** ([14], Theorem 5 in [13]). *Let \mathfrak{A} , \mathfrak{B} be countable ω -categorical relational structures such that \mathfrak{A} is Ramsey, and let \mathfrak{A}' , \mathfrak{B}' be reducts of these structures. If there exists an embedding from \mathfrak{A}' to \mathfrak{B}' , then there also exists an embedding from \mathfrak{A}' to \mathfrak{B}' that is canonical viewed as a mapping from \mathfrak{A} to \mathfrak{B} .*

3 The containment problem

To accurately measure the size of SNP sentences Φ , we introduce the following four parameters:

- the *height* $\text{ht}(\Phi)$ denotes the number of relation symbols in Φ ;
- the *length* $\text{lh}(\Phi)$ denotes the number of clauses in Φ ;
- the *width* $\text{wd}(\Phi)$ denotes the the maximum number of variables per clause in Φ ;
- the *arity* $\text{ar}(\Phi)$ denotes the maximum arity among all relation symbols in Φ .

3.1 Connectedness

An SNP τ -sentence $\exists X_1, \dots, X_n \forall \bar{x}. \phi(\bar{x})$ is *connected* in the sense of [11, 8] if, for each clause ψ in ϕ , the following $(\tau \cup \sigma)$ -structure \mathfrak{C}_ψ is connected: the domain consists of all variables in ψ , and $\bar{t} \in R^{\mathfrak{C}_\psi}$ if and only if $\neg R(\bar{t})$ is a disjunct in ψ ($R \in \tau \cup \sigma$). For example, the sentences in equations (1) and (2) from the introduction are connected, while the sentence

$$\exists B, R \forall x, y, u, v (\neg E(x, y) \vee \neg E(u, v) \vee \neg R(x, y) \vee B(u, v))$$

is not because the variable sets $\{x, y\}$ and $\{u, v\}$ partition the structure \mathfrak{C}_ψ associated to the clause $\psi = (\neg E(x, y) \vee \neg E(u, v) \vee \neg R(x, y) \vee B(u, v))$ into two disjoint substructures.

The next proposition was essentially proved in [11] (Proposition 1), except that the authors did not comment on the complexity of the procedure arising from their proof.

► **Proposition 8** (Proposition 1 in [11]). *Every GMSNP sentence Φ is logically equivalent to a disjunction of connected GMSNP sentences $\Phi_1 \vee \dots \vee \Phi_\ell$ over the same input and existential signatures with the following properties: for all $i \in [\ell]$, $\text{lh}(\Phi_i)$ and $\text{wd}(\Phi_i)$ are polynomial in $\text{lh}(\Phi)$ and $\text{wd}(\Phi)$, respectively; ℓ is single-exponential in $\text{lh}(\Phi) \cdot \text{wd}(\Phi)$; and the entire disjunction can be computed from Φ in deterministic single-exponential time.*

Proposition 8 shows that when studying the complexity of the the containment problem for GMSNP, we may without loss of generality restrict ourselves to connected GMSNP.

► **Corollary 9.** *If the containment problem for connected GMSNP sentences is in 2NEXPTIME, then also the containment problem for general GMSNP sentences is in 2NEXPTIME.*

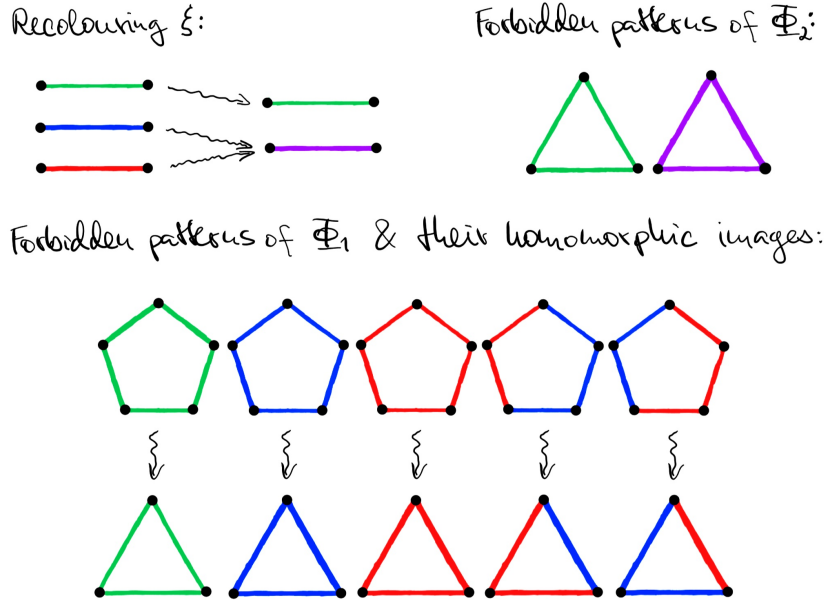
3.2 Recolourings

We call a relational structure *standard* if its domain is $[n]$ for some $n \in \mathbb{N}$. For an SNP τ -sentence Φ , we denote the class of all finite models of the first-order part of Φ (over the signature $\tau \cup \sigma$) by $\text{expfm}(\Phi)$. We define the n -colours of Φ , denoted $n\text{-colours}(\Phi)$, as the set of all standard structures from $\text{expfm}(\Phi)$ of size $\leq n$. A *recolouring* between two SNP τ -sentences Φ_1 and Φ_2 is a mapping ξ from $n\text{-colours}(\Phi_1)$ to $n\text{-colours}(\Phi_2)$, where $n := \max(\text{ar}(\Phi_1), \text{ar}(\Phi_2))$, with the two following properties. First, the τ -reducts of every n -colour of Φ_1 and of its ξ -image are identical. Secondly, for every $\mathfrak{A} \in \text{expfm}(\Phi_1)$, there is a structure $\xi'(\mathfrak{A}) \in \text{expfm}(\Phi_2)$ on the same domain such that for all $\mathfrak{N} \in n\text{-colours}(\Phi_1)$ and every embedding $e \in \binom{\mathfrak{A}}{\mathfrak{N}}$ we have $e \in \binom{\xi'(\mathfrak{A})}{\xi(\mathfrak{N})}$. In other words, the following extension ξ' of ξ is a well-defined mapping from $\text{expfm}(\Phi_1)$ to $\text{expfm}(\Phi_2)$:

For every $\mathfrak{A} \in \text{expfm}(\Phi_1)$, the structure $\xi'(\mathfrak{A})$ on the same domain as \mathfrak{A} is obtained by replacing for every $\mathfrak{N} \in n\text{-colours}(\Phi_1)$ and every embedding $e \in \binom{\mathfrak{A}}{\mathfrak{N}}$ the substructure $e(\mathfrak{N})$ of \mathfrak{A} by $e(\xi(\mathfrak{N}))$.

Note that the second formulation makes it clear that $\xi'(\mathfrak{A})$ is unique, because $n = \max(\text{ar}(\Phi_1), \text{ar}(\Phi_2))$. For the same reason, we have that $\mathfrak{A}^\tau = \xi'(\mathfrak{A})^\tau$. Also observe that in the second formulation the well-definedness of ξ' is a non-trivial property since the required replacements might be on overlapping substructures.

We remark that this definition of a recolouring is compatible with the definition of a recolouring for the logic MMSNP given in [29, 12]. There, the unary existential symbols were interpreted as colours, and a recolouring was simply a mapping between the existential symbols which does not introduce any forbidden patterns. It is noteworthy, however, that in that case the compatibility on overlapping structures is trivially satisfied.



■ **Figure 2** A recolouring from Φ_1 to Φ_2 in Example 10.

► **Example 10.** Consider the GMSNP $\{E\}$ -sentences (E is binary)

$$\begin{aligned}\Phi_1 &:= \exists R, G, B \forall x_1, x_2, x_3, x_4, x_5. \phi_1(x_1, x_2, x_3, x_4, x_5), \\ \Phi_2 &:= \exists P, G \forall x_1, x_2, x_3. \phi_2(x_1, x_2, x_3)\end{aligned}$$

whose forbidden colour patterns are as in Figure 2; we additionally forbid uncoloured edges and overlapping colours in order to ensure that the colours partition the edges.

$$\begin{aligned}& (\neg E(x_1, x_2) \vee R(x_1, x_2) \vee G(x_1, x_2) \vee B(x_1, x_2)) \\ & \wedge (\neg E(x_1, x_2) \vee \neg R(x_1, x_2) \vee \neg G(x_1, x_2)) \\ & \wedge (\neg E(x_1, x_2) \vee \neg G(x_1, x_2) \vee \neg B(x_1, x_2)) \\ & \wedge (\neg E(x_1, x_2) \vee \neg B(x_1, x_2) \vee \neg R(x_1, x_2))\end{aligned}$$

Note that in GMSNP we cannot enforce that the colours partition all pairs of elements of the domain, because every clause must satisfy the guarding axiom. Now consider a map ξ mapping the green colour of undirected edges to itself, while merging the red and the blue colours to purple. The remaining structures in $n\text{-colours}(\Phi_1)$, where $n = 2$, consist either of a single vertex or two non- E -related vertices, possibly related with $\{R, G, B\}$ -relations. They are mapped to $n\text{-colours}(\Phi_2)$ in an arbitrary way consistent with the first recolouring condition, i.e., that the $\{E\}$ -reducts are identical.

To show that ξ is a recolouring, we must show that it does not produce green or purple triangles in edge-coloured graphs that originally did not admit a homomorphism from any of the 5 forbidden colourings of pentagons. Clearly, ξ cannot produce a green triangle: the only way to produce a green triangle would be to start with one, and a green triangle admits a homomorphism from a green pentagon, which is forbidden. We can argue similarly in the case of a purple triangle, which can only be obtained with ξ from one of the 4 red/blue-coloured triangles at the bottom of the figure. All 4 of them admit a homomorphism from one of the forbidden coloured pentagons. Hence, ξ is a recolouring.

► **Lemma 11.** *The existence of a recolouring between two SNP sentences Φ_1 and Φ_2 can be tested non-deterministically in time $\mathcal{O}\left(\text{lh}(\Phi_1) \cdot \text{lh}(\Phi_2) \cdot 2^{\text{wd}(\Phi_1) \cdot \text{wd}(\Phi_2) \cdot \text{ht}(\Phi_1) \cdot \text{ht}(\Phi_2)} \cdot 2^{\text{ar}(\Phi_1) \cdot \text{ar}(\Phi_2)}\right)$.*

3.3 The theorem of Hubička and Nešetřil

A general construction method of ω -categorical homogeneous Ramsey structures was provided by Hubička and Nešetřil [25]. We use a variant of their result, which can be found in the appendix of [12] (Theorem A.4 in the print version, or Theorem A.3 in the arXiv version).

Pieces of structures. The following notion, originating from [24], is of essential importance in the present article. A *piece* of a structure \mathfrak{B} is a pair (\mathfrak{P}, \bar{t}) , where \mathfrak{P} is a proper substructure of \mathfrak{B} (with domain $P \subsetneq B$) and \bar{t} (called the *root* of (\mathfrak{P}, \bar{t})) is a tuple with pairwise distinct entries enumerating a non-empty subset T of P such that every tuple \bar{s} contained in a relation of \mathfrak{B} is completely contained either in P or in $T \cup (B \setminus P)$.

► **Remark 12.** The definition of a piece in the present paper is more general than the one in [24] because we do not require \bar{t} to be a *minimal separating cut* of \mathfrak{P} . However, this difference does not matter here since we employ the results from [24] in a restricted setting.

To every piece (\mathfrak{P}, \bar{t}) , we assign a relational symbol of arity equal to the length of \bar{t} , which for the sake of brevity will be denoted by the piece itself. For an atomic formula $(\mathfrak{P}, \bar{t})(\bar{x})$ using such a symbol, let $(\mathfrak{P}, \bar{t})(\bar{x})^{-1}$ be the τ -formula obtained from the canonical query of \mathfrak{P} by substituting \bar{x} for \bar{t} in the order in which both tuples are listed; that is, the formula asserts about \bar{x} all atomic τ -formulas (with parameters) that hold for \bar{t} in \mathfrak{P} . Finally, the primitive positive formula $(\mathfrak{P}, \bar{t})(\bar{x})^{-1\exists}$ is obtained from $(\mathfrak{P}, \bar{t})(\bar{x})^{-1}$ by existentially quantifying over all variables which are not contained in \bar{x} . That is, the formula asserts the existence of a homomorphic image of the piece (\mathfrak{P}, \bar{t}) around \bar{x} , with \bar{x} taking the place of \bar{t} .

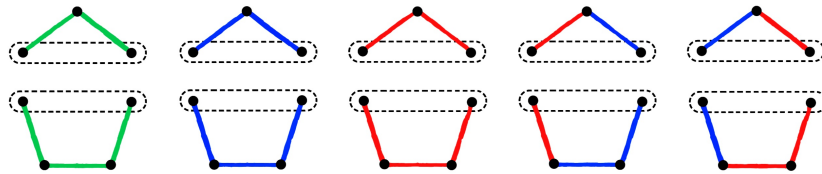
For a set \mathcal{F} of structures with a signature τ , we denote by $\text{Forb}_h(\mathcal{F})$ the class of all finite τ -structures which do not admit a homomorphism from any member of \mathcal{F} . The following theorem of Hubička and Nešetřil states that every class of the form $\text{Forb}_h(\mathcal{F})$, where each member of \mathcal{F} is connected, almost has the amalgamation property, up to taking an expansion where pieces of the structures in \mathcal{F} are marked using fresh predicates ρ ; moreover, the class of all linear-order expansions of structures from such a class has the Ramsey property.

► **Theorem 13** (Hubička and Nešetřil, Theorem A.4 in [12]). *Let \mathcal{F} be a finite set of finite connected structures over a common finite relational signature τ . Let ρ be the signature whose elements are the pieces (\mathfrak{P}, \bar{t}) of structures from \mathcal{F} . Consider the class $\mathcal{K}_{\text{HN}}(\mathcal{F})$ consisting of all substructures of those ρ -expansions of structures from $\text{Forb}_h(\mathcal{F})$ that satisfy*

$$\forall \bar{x} ((\mathfrak{P}, \bar{t})(\bar{x}) \iff (\mathfrak{P}, \bar{t})(\bar{x})^{-1\exists}). \quad (4)$$

Then the class $\mathcal{K}_{\text{HN}}^<(\mathcal{F}) := (\mathcal{K}_{\text{HN}}(\mathcal{F}))^<$ has the RP and the AP.

To illustrate the construction in Theorem 13, consider the two sets \mathcal{F}_1 and \mathcal{F}_2 consisting of the canonical databases of the forbidden patterns of Φ_1 and Φ_2 from Figure 2. It is not hard to see that $\text{Forb}_h(\mathcal{F}_2)$ already has the AP because the forbidden graphs are cliques; moreover, $\text{Forb}_h(\mathcal{F}_2)^<$ has the AP and the RP by a theorem of Nešetřil and Rödl [33]. Therefore, \mathcal{F}_2 represents a trivial case where adding ρ -predicates is not necessary. This is not true for \mathcal{F}_1 , since $\text{Forb}_h(\mathcal{F}_1)$ does not have the AP. In Figure 3, we provide a graphical representation of the pieces of structures that need to be stored using ρ -predicates in order to obtain a class of expansions of structures from $\text{Forb}_h(\mathcal{F}_1)$ with the AP and the RP.



■ **Figure 3** An illustration of some pieces of the canonical databases of the forbidden patterns of the SNP sentence Φ_1 from Example 10; the roots of the pieces are marked by the dashed lines.

3.4 A proof of Theorem 2

Besides the reduction to the connected case, the proof of Theorem 2 is a straightforward consequence of the combination of the following three intermediate results. First, Lemma 11 from Section 3.2 gives a (non-deterministic) double-exponential upper bound on the time complexity of finding a recolouring between two SNP sentences. Second, Proposition 14 below states that every connected GMSNP sentence is equivalent to an at most double-exponentially larger SNP sentence whose class of expanded finite models enjoys the AP

and the RP. Third, Proposition 15 states that, for SNP sentences with the latter property, containment is equivalent to the existence of a recolouring. Here we are in fact very lucky that the parameters on which Lemma 11 and Proposition 14 depend doubly-exponentially are different – this will become clear in the proof of Theorem 2.

► **Proposition 14.** *From every connected GMSNP τ -sentence Φ one can compute in deterministic double-exponential time an SNP τ -sentence $\Delta(\Phi)$ such that:*

1. $\text{wd}(\Delta(\Phi)) \in \mathcal{O}(\text{wd}(\Phi))$;
2. $\text{ar}(\Delta(\Phi)) \in \mathcal{O}(\text{ar}(\Phi) + \text{wd}(\Phi))$;
3. $\text{ht}(\Delta(\Phi)) \in \mathcal{O}(\text{ht}(\Phi) + \text{lh}(\Phi) \cdot 2^{\text{wd}(\Phi)})$
4. $\text{lh}(\Delta(\Phi)) \in \mathcal{O}(2^{(\text{ht}(\Phi) + \text{lh}(\Phi)) \cdot 2^{\text{wd}(\Phi) + \text{ar}(\Phi)}})$;
5. $\text{expfm}(\Delta(\Phi))$ has the AP and the RP;
6. $\text{fm}(\Delta(\Phi)) = \text{fm}(\Phi)$.

The proof of Proposition 14 below builds on the proof of [11, Theorem 5], so we omit some details that can be found in [11].

Proof sketch. Let Φ be an arbitrary connected GMSNP τ -sentence. We first obtain signatures τ' and σ' by introducing, for every $R \in \tau \cup \sigma$, two new symbols R_+ and R_- of the same arity. Next, we obtain an auxiliary GMSNP τ' -sentence Φ' from Φ as follows. For each forbidden pattern ϕ of Φ , the sentence Φ' contains the forbidden pattern ϕ' obtained by replacing each positive atom $R(\bar{x})$ with $R_+(\bar{x})$ and each negative atom $\neg R(\bar{x})$ with $R_-(\bar{x})$.

Let \mathcal{F} be the set of canonical databases for the forbidden patterns of Φ' (which are just conjunctions of positive $\tau' \cup \sigma'$ -atoms); define ρ from \mathcal{F} as in Theorem 13. Consider the τ' -sentence Φ'' with the additional existential symbols $\rho \cup \{<\}$ obtained from Φ' by adding the following new clauses:

- i. clauses defining the linear order using irreflexivity, transitivity, and totality:

$$\forall x, y, z \neg <(x, x) \wedge (<(x, y) \wedge <(y, z) \Rightarrow <(x, z)) \wedge (<(x, y) \vee (x = y) \vee <(y, x));$$

- ii. every clause ϕ in the signature $\tau' \cup \sigma' \cup \rho$ of the form $\psi(\bar{x}, \bar{y}) \Rightarrow (\mathfrak{P}', \bar{t}')(\bar{x})$ (for $(\mathfrak{P}', \bar{t}') \in \rho$) with at most $\text{wd}(\Phi)$ -many variables, where ψ is a conjunction of positive atoms with the following property: replacing every atomic subformula $(\mathfrak{P}, \bar{t})(\bar{u})$ in ψ by $(\mathfrak{P}, \bar{t})(\bar{u})^{-1}$ yields a $(\tau' \cup \sigma')$ -formula ψ^{-1} such that there exists a homomorphism h from the canonical database of $(\mathfrak{P}', \bar{t}')(\bar{x})^{-1}$ to the canonical database of ψ^{-1} satisfying $h(\bar{x}) = \bar{x}$;
- iii. every clause ϕ in the signature $\tau' \cup \sigma' \cup \rho$ with at most $\text{wd}(\Phi)$ -many variables whose forbidden pattern is a conjunction of positive atoms with the following property: replacing every atomic subformula $(\mathfrak{P}, \bar{t})(\bar{u})$ in ϕ by $(\mathfrak{P}, \bar{t})(\bar{u})^{-1}$ yields a formula whose canonical database admits a homomorphism from a member of \mathcal{F} .

Before we proceed further, let us assess the size of Φ'' in detail. The number of variables per clause remains in $\mathcal{O}(\text{wd}(\Phi))$. The arity of the symbols from ρ is in $\mathcal{O}(\text{ar}(\Phi) + \text{wd}(\Phi))$, because Φ'' inherits all symbols from Φ' and gains additional symbols whose arity is bounded by the domain size of structures in \mathcal{F} . The number of relation symbols increases exponentially in $\text{wd}(\Phi)$ and polynomially in the remaining parameters, more specifically it is in $\mathcal{O}(\text{ht}(\Phi) + \text{lh}(\Phi) \cdot 2^{\text{wd}(\Phi)})$; the reason is that the pieces of structures in \mathcal{F} are no more than all possible substructures of structures in \mathcal{F} . Regarding the number of clauses: Item (i) produces constant number of clauses and items (ii) and (iii) produce $\mathcal{O}(2^{(\text{ht}(\Phi) + \text{lh}(\Phi)) \cdot 2^{\text{wd}(\Phi) + \text{ar}(\Phi)}})$ many clauses. The reason for the double-exponential blow-up in item (iii) is that there are as many new clauses as there are structures of size at most $\text{wd}(\Phi)$ over the signature of Φ'' . Similarly, also item (ii) yields a double-exponential blow-up; here the number from item (iii) is additionally multiplied with the number of symbols in ρ (i.e., pieces of structures in \mathcal{F}), which is in $\mathcal{O}(\text{lh}(\Phi) \cdot 2^{\text{wd}(\Phi)})$.

Note that all structures in \mathcal{F} are connected since Φ is connected. By Theorem 13, the class $\mathcal{K}_{\text{HN}}^<(\mathcal{F})$ over the signature $\tau' \cup \sigma' \cup \rho \cup \{<\}$ has the AP and the RP. Theorem 6 then implies that there exists a homogeneous Ramsey structure \mathfrak{C} with $\text{age}(\mathfrak{C}) = \mathcal{K}_{\text{HN}}^<(\mathcal{F})$. By the construction of Φ'' , we have that $\text{age}(\mathfrak{C}) = \text{expfm}(\Phi'')$; below we provide an intuitive explanation of why this is the case. First, item (i) ensures that $<$ interprets as a linear order in each $\mathfrak{A} \in \text{expfm}(\Phi'')$ which, for lack of additional clauses with $<$, is independent of the remaining signature; the same is true in each $\mathfrak{A} \in \text{age}(\mathfrak{C})$. Next, consider the remaining part $\tau' \cup \sigma' \cup \rho$ of the signature. If $\mathfrak{A} \in \text{age}(\mathfrak{C})$, then by definition, it does not admit a homomorphism from any member of \mathcal{F} ; moreover, its ρ -predicates are obtained through equation (4) by definition, and hence it satisfies items (ii) and (iii). Whence, $\mathfrak{A} \in \text{expfm}(\Phi'')$. Conversely, if a finite structure \mathfrak{A} satisfies the first-order part of Φ'' , then extending it by new elements so that the truth of all ρ -predicates is witnessed as in equation (4) yields a structure in which the equivalence (4) holds (because of item (ii)) and which does not admit a homomorphism from any member from \mathcal{F} (because of item (iii)); hence $\mathfrak{A} \in \text{age}(\mathfrak{C})$.

We say that a set $S \subseteq C$ is *correctly labelled* if every tuple \bar{t} over S matching the arity of some $X \in \sigma$ is contained either in $X_+^{\mathfrak{C}}$ or in $X_-^{\mathfrak{C}}$, but not both. Consider the $(\tau \cup \sigma \cup \tau' \cup \sigma' \cup \rho \cup \{<\})$ -expansion \mathfrak{C}_{Φ} of \mathfrak{C} where each $R \in \tau \cup \sigma$ of arity k interprets as the k -ary relation defined by the quantifier-free first-order formula

$$R_+(x_1, \dots, x_k) \wedge \left(\{x_1, \dots, x_k\} \text{ is correctly labelled} \right). \quad (5)$$

As first-order definable relations are preserved by automorphisms, \mathfrak{C}_{Φ} is homogeneous Ramsey.

Finally, we obtain the sentence $\Delta(\Phi)$ from Φ'' by first adding clauses defining $(\tau \cup \sigma)$ -relations using equation (5) and then existentially quantifying over all symbols in $\tau' \cup \sigma'$. The length of $\Delta(\Phi)$ only increases by an additional multiplicative single exponential because we need to consider all subsets of $[\text{ar}(\Phi) + \text{wd}(\Phi)]$ for every relation symbol in σ ; its width, height, and arity remain unchanged. It follows from Theorem 13 that $\text{expfm}(\Delta(\Phi))$ has the AP and the RP. Moreover, we have $\text{age}(\mathfrak{C}_{\Phi}) = \text{expfm}(\Delta(\Phi))$. It remains to show that $\text{fm}(\Phi) = \text{fm}(\Delta(\Phi))$. This follows from the fact that $\text{CSP}(\mathfrak{C}_{\Phi}^{\tau}) = \text{fm}(\Phi) = \text{age}(\mathfrak{C}_{\Phi}^{\tau})$. ◀

The next proposition shows that the property in item (5) of Proposition 14 attained through the transformation Δ is sufficient for a reduction from containment to recolouring.

► **Proposition 15.** *For SNP τ -sentences Φ_1 and Φ_2 such that $\text{expfm}(\Phi_1)$ and $\text{expfm}(\Phi_2)$ both have the AP and the RP the following are equivalent.*

1. $\text{fm}(\Phi_1) \subseteq \text{fm}(\Phi_2)$.
2. *There exists a recolouring from Φ_1 to Φ_2 .*

Proof. Let $n := \max(\text{ar}(\Phi_1), \text{ar}(\Phi_2))$.

“(1) \Rightarrow (2)” By Theorem 6, there exist homogeneous Ramsey structures \mathfrak{C}_1 and \mathfrak{C}_2 such that $\text{expfm}(\Phi_1) = \text{age}(\mathfrak{C}_1)$ and $\text{expfm}(\Phi_2) = \text{age}(\mathfrak{C}_2)$. Then, $\text{age}(\mathfrak{C}_1^{\tau}) \subseteq \text{age}(\mathfrak{C}_2^{\tau})$. By the ω -categoricity of \mathfrak{C}_2^{τ} , Lemma 5 implies that there exists an embedding e from \mathfrak{C}_1^{τ} to \mathfrak{C}_2^{τ} . By Theorem 7, we may assume that e is canonical as a function from \mathfrak{C}_1 to \mathfrak{C}_2 since the former is an ω -categorical Ramsey structure. Let ξ' be the mapping from $\text{expfm}(\Phi_1)$ to $\text{expfm}(\Phi_2)$ defined as follows. Given $\mathfrak{A} \in \text{expfm}(\Phi_1)$, consider an arbitrary embedding i from \mathfrak{A} to \mathfrak{C}_1 . Then, we define $\xi'(\mathfrak{A}) \in \text{expfm}(\Phi_2)$ with domain A whose relations are defined through their preimages under $e \circ i$. Let ξ be the restriction of ξ' to $n\text{-colours}(\Phi_1)$. We claim that ξ' arises from ξ as in the definition of a recolouring.

First, the τ -reduct of any $\mathfrak{N} \in n\text{-colours}(\Phi_1)$ and its ξ -image are identical since $\xi(\mathfrak{N}) = \xi'(\mathfrak{N})$ is obtained by pulling back relations via the τ -embedding $e \circ i$. For the second part

of the definition pertaining to ξ' , let $\mathfrak{A} \in \text{expfm}(\Phi_1)$, $\mathfrak{N} \in n\text{-colours}(\Phi_1)$, and $f \in \binom{\mathfrak{A}}{\mathfrak{N}}$ be arbitrary; we claim that $f \in \binom{\xi'(\mathfrak{A})}{\xi(\mathfrak{N})}$. By definition, the relations of $\xi'(\mathfrak{A})$ and of $\xi(\mathfrak{N})$ are obtained by first embedding \mathfrak{A} and \mathfrak{N} into \mathfrak{C}_1 , then applying e , and then pulling back the relations from \mathfrak{C}_2 . This definition does not depend on the choice of the embeddings into \mathfrak{C}_1 , by the homogeneity of \mathfrak{C}_1 and the canonicity of f – the claim follows.

“(2) \Rightarrow (1)” Let $\mathfrak{A} \in \text{fm}(\Phi_1)$ be arbitrary; then there exists an expansion $\mathfrak{A}' \in \text{expfm}(\Phi_1)$. By the definition of a recolouring we have $\xi'(\mathfrak{A}') \in \text{expfm}(\Phi_2)$. It also follows from this definition that for all $\mathfrak{N} \in n\text{-colours}(\Phi_1)$ and every embedding $e \in \binom{\mathfrak{A}}{\mathfrak{N}}$ we have $e \in \binom{\xi'(\mathfrak{A}')}{\xi(\mathfrak{N})}$. By our choice of n , this implies that the τ -reduct of $\xi'(\mathfrak{A}')$ coincides with \mathfrak{A} . Hence $\xi'(\mathfrak{A}')$ witnesses that $\mathfrak{A} \in \text{fm}(\Phi_2)$. \blacktriangleleft

Proof of Theorem 2. By Corollary 9, we can restrict ourselves to connected GMSNP. Now the statement follows directly from Proposition 14 combined with Proposition 15 and Lemma 11. To see this, note that, through the application of Proposition 14, the arity and width increase polynomially, the height increases single-exponentially, and the length increases double-exponentially. According to Lemma 11, the complexity of testing the existence of a recolouring only depends single-exponentially on the height, so a single-exponential increase in height is not an issue. Moreover, the complexity only depends polynomially on the length, so a double-exponential increase in length is not an issue either. We conclude that the combined complexity is indeed 2NEXPTIME. \blacktriangleleft

4 Recolouring-ready GMSNP

We first introduce the notion of *recolouring-readiness*. Then, we give a high-level explanation of how to obtain recolouring-ready GMSNP sentences and elaborate on our claim that recolouring-readiness can be paired with a notion of a recolouring generalising the corresponding notion for MMSNP in a fashion that is less abstract than the notion introduced in Section 3.2 and closer to the idea of edge-recolourings as in equation (3) of Section 1.4.

We will use the following weaker version of homogeneity, generalising the notion of *1-homogeneity* of Bodirsky, Madelaine, and Mottet [12] from vertices to relational τ -tuples. A relational structure \mathfrak{B} whose signature contains τ is *τ -edge-homogeneous* if it satisfies the homogeneity condition restricted to tuples \bar{t}_1, \bar{t}_2 contained in $R^{\mathfrak{B}}$ for some $R \in \tau$.

► **Remark 16.** The τ -reduct of a τ -edge-homogeneous structure is not necessarily homogeneous. Consider the disjoint union of the *random graph* $(V; E)$ with a countably infinite set marked by a unary predicate R . For $\tau = \{E\}$, the τ -reduct of this structure is τ -edge-homogeneous but not homogeneous.

A connected GMSNP τ -sentence Φ is called *recolouring-ready* if there exists a τ -edge-homogeneous ω -categorical Ramsey structure $(\mathfrak{C}_\Phi, <)$ with the following properties:

- i. $\text{age}(\mathfrak{C}_\Phi, <)$ consists of all linear-order expansions of structures from $\text{age}(\mathfrak{C}_\Phi)$;
- ii. $\text{age}(\mathfrak{C}_\Phi)$ consists of the finite models of the first-order part of Φ ;
- iii. $\text{age}(\mathfrak{C}_\Phi^\tau) = \text{CSP}(\mathfrak{C}_\Phi^\tau) = \text{fm}(\Phi)$.

Theorem 17 below is a detailed version of Theorem 4 from the introduction.

► **Theorem 17.** *For every connected GMSNP τ -sentence Φ there exists a logically equivalent recolouring-ready connected GMSNP τ -sentence $\Omega(\Phi)$ such that:*

- 1. $\text{fm}(\Omega(\Phi)) = \text{fm}(\Phi)$;
- 2. $\text{wd}(\Omega(\Phi)) \in \mathcal{O}(\text{wd}(\Phi))$;
- 3. $\text{ar}(\Omega(\Phi)) \in \mathcal{O}(\text{ar}(\Phi))$;

4. $\text{ht}(\Omega(\Phi)) \in \mathcal{O}(2^{(\text{wd}(\Phi) + \text{ar}(\Phi))^{\text{ar}(\Phi)} \text{ht}(\Phi)});$
5. $\text{lh}(\Omega(\Phi)) \in \mathcal{O}(2^{2^{(\text{wd}(\Phi) + \text{ar}(\Phi))^{\text{ar}(\Phi)} \cdot \text{ht}(\Phi)}}).$

Moreover, $\Omega(\Phi)$ can be computed from Φ in nondeterministic triple-exponential time.

Proof sketch. We only describe how $\Omega(\Phi)$ is constructed and discuss its size. The overall strategy is similar to the proof of Proposition 14, but with several thorny technical details.

First, for every $X \in \sigma$, we add a new symbol \bar{X} of the same arity as X . Then, for every $X \in \sigma$, we replace in every clause of Φ every negated X -atom with a positive \bar{X} -atom. Next, we make every clause correctly labelled. Intuitively, for every clause $\phi(\bar{x})$, every tuple $\bar{y} \subseteq \bar{x}$ such that there is an atom $R(\bar{z})$ in $\phi(\bar{x})$ with $\bar{y} \subseteq \bar{z}$, and every $X \in \sigma$ whose arity matches the arity of \bar{y} , we replace $\phi(\bar{x})$ with two new clauses $(\phi \wedge X(\bar{y}))(\bar{x})$ and $(\phi \wedge \bar{X}(\bar{y}))(\bar{x})$. The process is repeated until the canonical databases of all clauses are correctly labelled in the sense of Proposition 14. This step results in a doubly-exponential increase of the number of clauses, but is necessary because the definition of \mathfrak{C}_Φ here is more involved than in Proposition 14. Denote the resulting sentence by $\bar{\Phi}$ and the family of canonical databases of the clauses of $\bar{\Phi}$ by $\bar{\mathcal{F}}$. Now, for every $\mathfrak{F} \in \bar{\mathcal{F}}$ and every piece (\mathfrak{P}, \bar{t}) of \mathfrak{F} , we extend the tuple \bar{t} with new variables in every possible way so that the arity of the resulting tuple $\bar{s} \supseteq \bar{t}$ matches the arity of some relation $R \in \tau$. We define σ_+ as the signature consisting of all such symbols $(\mathfrak{P} \cup \bar{s}, \bar{s})(\bar{s})$ associated with these pieces. Note that $|\sigma_+|$ is at most the number of $(\tau \cup \bar{\sigma})$ -structures on at most $\text{wd}(\Phi) + \text{ar}(\Phi)$ elements. So, $|\sigma_+|$ is in $\mathcal{O}(2^{(\text{wd}(\Phi) + \text{ar}(\Phi))^{\text{ar}(\Phi)} \text{ht}(\Phi)})$. Finally, we obtain the sentence $\bar{\Phi}_+$ from $\bar{\Phi}$ by adding new clauses defining the predicates with a symbol in σ_+ . We first add clauses stating that the canonical query of a piece $(\mathfrak{P} \cup \bar{s}, \bar{s})$ implies the predicate $(\mathfrak{P} \cup \bar{s}, \bar{s})$; that is, the implication (\Leftarrow) in equation (4) holds. As in item (ii) from the proof of Proposition 14, we do this in a way that accounts for the fact that already present σ_+ -predicates $(\mathfrak{P} \cup \bar{s}, \bar{s})(\bar{x})$ themselves encode formulas of the form $(\mathfrak{P} \cup \bar{s}, \bar{s})(\bar{x})^{-1\exists}$. Similarly, we reproduce item (iii) from the proof of Proposition 14; for convenience of the reader, we provide an explicit set of clauses: Let n be the size of the largest element of $\bar{\mathcal{F}}$. We add to the first-order part of $\bar{\Phi}$ every clause in the signature $\tau \cup \bar{\sigma}_+$ of at most n variables whose forbidden pattern is a conjunction of positive atoms with the following property: replacing every subconjunction $(\mathfrak{P} \cup \bar{s}, \bar{s})(\bar{u}) \wedge R(\bar{u})$ (where $R(\bar{u})$ holds in the piece) by $(\mathfrak{P} \cup \bar{s}, \bar{s})(\bar{u})^{-1} \wedge R(\bar{u})$ yields a formula whose canonical database admits a homomorphism from a member of $\bar{\mathcal{F}}$. By construction, the sentence $\bar{\Phi}_+$ is still in GMSNP; we set $\Omega(\Phi) := \bar{\Phi}_+$.

It can be verified that item (1) holds, i.e. that $fm(\bar{\Phi}_+) = fm(\Phi)$. Items (2) and (3) follow from the construction, item (4) follows from the upper bound on $|\sigma_+|$. The parameter $\text{lh}(\bar{\Phi}_+)$ is bounded from above by the number of $(\tau \cup \bar{\sigma}_+)$ -structures on at most $\text{wd}(\Phi) + \text{ar}(\Phi)$ elements, so it is in $\mathcal{O}(2^{(\text{wd}(\Phi) + \text{ar}(\Phi))^{\text{ar}(\Phi)} |\sigma_+|})$. Hence, item (5) follows as well. \blacktriangleleft

Let \mathfrak{A} be a structure over a relational signature containing τ . We say that \mathfrak{A} is τ -guarded if there exist $R \in \tau$ and $\bar{t} \in R^{\mathfrak{A}}$ such that $A \subseteq \bar{t}$. Now, let Φ_1 and Φ_2 be GMSNP τ -sentences; as in Section 3.2, we set $n := \max(\text{ar}(\Phi_1), \text{ar}(\Phi_2))$. A structure \mathfrak{A} with domain $[n]$ over a signature containing $<$ is *standardly ordered* if $<^{\mathfrak{A}}$ coincides with the natural ordering $1 < \dots < n$. A *GMSNP-recolouring* from Φ_1 to Φ_2 is a mapping ξ from standardly ordered τ -guarded structures in $n\text{-colours}(\Phi_1)^<$ to standardly ordered τ -guarded structures in $n\text{-colours}(\Phi_2)^<$ with the following two properties. First, the τ -reduct of any structure and its ξ -image are isomorphic. Secondly, for every $\mathfrak{A} \in \text{expfm}(\Phi_1)^<$, there is a structure $\xi'(\mathfrak{A}) \in \text{expfm}(\Phi_2)^<$ on the same domain such that $<^{\mathfrak{A}} = <^{\xi'(\mathfrak{A})}$ and, for all standardly ordered τ -guarded $\mathfrak{N} \in n\text{-colours}(\Phi_1)^<$ and for every embedding $e \in \binom{\mathfrak{A}}{\xi(\mathfrak{N})}$, we have $e \in \binom{\xi'(\mathfrak{A})}{\xi(\mathfrak{N})}$.

Modulo the fact that the ordering of the colours matters in this definition, it is what one would intuitively understand under a recolouring of relational τ -tuples. The fact that the containment between recolouring-ready GMSNP sentences can be reduced to the existence of such a recolouring can be proved almost exactly as in the proof of Proposition 15.

► **Proposition 18.** *For connected recolouring-ready GMSNP τ -sentences Φ_1 and Φ_2 , the following are equivalent:*

1. $fm(\Phi_1) \subseteq fm(\Phi_2)$;
2. *There exists a GMSNP-recolouring from Φ_1 to Φ_2 .*

Now we compare the above defined GMSNP-recolourings, which depend on an external linear order, with the (linear-order-free) edge-recolourings from equation (3) in Section 1.4. We claim that, for connected recolouring-ready GMSNP sentences Φ_1 and Φ_2 over signatures $\tau \cup \sigma_1$ and $\tau \cup \sigma_2$, respectively, the two notions only differ up to a minor syntactic modification Ω' : there exists a GMSNP-recolouring from Φ_1 to Φ_2 if and only if there exists an edge-recolouring from $\Omega'(\Phi_1)$ to $\Omega'(\Phi_2)$.

We describe Ω' . For every τ -guarded $\mathfrak{C} \in n\text{-colours}(\Phi_i)^{<}$ of size $k \leq n$, we introduce a new k -ary existential symbol $X_{\mathfrak{C}}$; the set of all such symbols is denoted by $\sigma_i^{n\text{-colours}}$. The *atomic diagram* of a structure \mathfrak{C} is the conjunction extending its canonical query by listing not only all the atoms that hold in \mathfrak{C} , but also their negations [23]. We replace the clauses of Φ_i by all possible clauses ψ over $\tau \cup \sigma_i^{n\text{-colours}}$ with at most $(\text{wd}(\Phi_i) + 2 \cdot \text{ar}(\Phi_i))$ -many variables satisfying the monotonicity and the guarding axioms such that replacing every atom $X_{\mathfrak{C}}(\bar{x})$ in the forbidden pattern of ψ by the atomic diagram of \mathfrak{C} yields a formula that is not satisfiable together with the first-order part of Φ_i . Note that a similar trick was used in the proof sketch of Proposition 14. Next, we introduce a new binary existential symbol $<$, add all clauses forbidding cycles in $<$ of size at most $\text{wd}(\Phi_i)$, and for every $\mathfrak{C} \in n\text{-colours}(\Phi_i)^{<}$ we add the clause $\forall x_1, \dots, x_k (X_{\mathfrak{C}}(x_1, \dots, x_k) \Rightarrow (\bigwedge_{i < j} x_i < x_j))$. The resulting GMSNP sentence is denoted by $\Omega'(\Phi_i)$.

Note that $\Omega'(\Phi_i)$ is logically equivalent to Φ_i : any $\mathfrak{A}_i \in \text{expfm}(\Phi_i)$ can be turned into a model \mathfrak{A}_i^- of the first-order part of $\Omega'(\Phi_i)$ without changing the τ -relations by first equipping it with a linear order $<$, adding the predicates $\sigma_i^{n\text{-colours}}$ according to their semantics, and then removing the σ_i -predicates; \mathfrak{A}_i^- satisfies the consistency clauses by definition. In the converse direction, we obtain a model \mathfrak{A}_i^+ of the first-order part of Φ_i from a model \mathfrak{A}_i of $\Omega'(\Phi_i)$ by the obvious replacement and by forgetting $<$, thanks to the consistency requirements. Note that it is also the consistency requirements that guarantee that the obvious replacement is not contradictory on overlapping τ -guarded n -colours; to ensure this, we allowed clauses ψ up to size $2 \cdot \text{ar}(\Phi_i)$.

Let ξ be a GMSNP-recolouring from Φ_1 to Φ_2 ; we claim it naturally induces an edge-recolouring from $\Omega'(\Phi_1)$ to $\Omega'(\Phi_2)$. Namely, for any $\mathfrak{A}_1 \in \text{expfm}(\Omega'(\Phi_1))$, we replace every predicate $X_{\mathfrak{C}}$ in $\sigma_1^{n\text{-colours}}$ which holds on some tuple in \mathfrak{A}_1 by the predicate encoding the image of \mathfrak{C} , ordered by the appearance of its elements in the predicate, under ξ . The resulting structure \mathfrak{A}_2 is an element of $\text{expfm}(\Omega'(\Phi_2))$: suppose otherwise. Then \mathfrak{A}_2^+ realizes a forbidden pattern of Φ_2 on some set $S = \{s_1, \dots, s_{\text{wd}(\Phi_2)}\}$; without loss of generality, since $\Omega'(\Phi_2)$ forbids cycles of size at most $\text{wd}(\Phi_2)$, we may assume that $s_i < s_j$ implies $i < j$ for all pairs (s_i, s_j) guarded in that pattern. Let \mathfrak{S} be the structure induced by S in \mathfrak{A}_1^+ , equipped with the order $s_1 < \dots < s_{\text{wd}(\Phi_2)}$; then $\xi'(\mathfrak{S})$, where ξ' is obtained from ξ by the definition of a GMSNP-recolouring, is a model of the first-order part of Φ_2 which agrees with \mathfrak{A}_2^+ on guarded tuples, a contradiction.

Conversely, given an edge-recolouring from $\Omega'(\Phi_1)$ to $\Omega'(\Phi_2)$, we trivially get that $fm(\Phi_1) \subseteq fm(\Phi_2)$, which means that there exists a GMSNP-recolouring from Φ_1 to Φ_2 due to Proposition 18 because Φ_1 and Φ_2 are recolouring-ready.

5 Conclusion

We proved the decidability of the containment problem for GMSNP, thereby settling an open question posed in [6, 15]. Our decision procedure runs in non-deterministic double-exponential time, which exactly matches the lower bound on the complexity obtained in [15].

In the proof of Theorem 2, our main result, we employ structural Ramsey theory to effectively reduce from the containment problem for GMSNP to the problem of testing the existence of a recolouring between SNP sentences. As mentioned in the introduction, the use of structural Ramsey theory is only one of the possible ways to outsource combinatorics. In fact, in all works on the containment problem for MMSNP except for [12] a different tool was used, commonly known as the *sparse incomparability lemma* [27, Theorem 1] (Feder and Vardi originally used a weaker, randomised version of this result, which they attributed to Erdős [19, Theorem 5]). It would be interesting to know if this alternative approach can also be used to analyse the complexity of containment for GMSNP, perhaps in combination with the reduction to *relativised emptiness* from [15].

Regarding our approach using structural Ramsey theory, one can clearly see that our methods have the potential to work in a broader setting than GMSNP. More specifically, by Proposition 15, the 2NEXPTIME upper bound on the complexity of containment holds for all pairs of SNP sentences whose first-order part defines a class of finite structures with the AP and the RP. The issue with this statement is that the said fragment of SNP does not have any known explicit description; it might be difficult to judge Proposition 15 as a stand-alone contribution since the complexity of recognizing which SNP sentences fall within its scope might be high, potentially even undecidable. For a discussion of the complexity of related meta problems, see [35] (cf. also [36]). We would be interested in locating virtually any fragment of existential second-order logic that properly generalises GMSNP, can be embedded into SNP with AP and RP similarly as GMSNP, and whose syntax is efficiently verifiable as in the case of GMSNP.

References

- 1 F. Baader. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press, 2003. doi:10.1017/CB09780511711787.
- 2 A. Barsukov, M. Pinsker, and J. Rydval. Containment for Guarded Monotone Strict NP. *CoRR*, abs/2310.01254, 2023. doi:10.48550/arxiv.2310.01254.
- 3 L. Barto, J. Bulín, A. A. Krokhnin, and J. Oprsal. Algebraic Approach to Promise Constraint Satisfaction. *J. ACM*, 68(4):28:1–28:66, 2021. doi:10.1145/3457606.
- 4 L. Barto, J. Opršal, and M. Pinsker. The wonderland of reflections. *Israel Journal of Mathematics*, 223(1):363–398, 2018. doi:10.1007/s11856-017-1621-9.
- 5 L. Barto and M. Pinsker. Topology Is Irrelevant (In a Dichotomy Conjecture for Infinite Domain Constraint Satisfaction Problems). *SIAM J. Comput.*, 49(2):365–393, 2020. doi:10.1137/18M1216213.
- 6 M. Bienvenu, B. ten Cate, C. Lutz, and F. Wolter. Ontology-Based Data Access: A Study through Disjunctive Datalog, CSP, and MMSNP. *ACM Trans. Database Syst.*, 39(4):33:1–33:44, 2014. doi:10.1145/2661643.
- 7 Z. Bitter and A. Mottet. Generalized Completion Problems with Forbidden Tournaments. In R. Kráľovic and A. Kucera, editors, *49th International Symposium on Mathematical Foundations of Computer Science, MFCS 2024, August 26-30, 2024, Bratislava, Slovakia*, volume 306 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 28:1–28:17, Dagstuhl, Germany, 2024. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.MFCS.2024.28.

- 8 M. Bodirsky. *Complexity of Infinite-Domain Constraint Satisfaction*. Lecture Notes in Logic. Cambridge University Press, 2021. Postprint available online: <https://wwwpub.zih.tu-dresden.de/~bodirsky/Book.pdf>. doi:10.1017/9781107337534.
- 9 M. Bodirsky and V. Dalmau. Datalog and Constraint Satisfaction with Infinite Templates. *J. Comput. Syst. Sci.*, 79(1):79–100, 2013. A preliminary version appeared in the proceedings of the Symposium on Theoretical Aspects of Computer Science (STACS’06). doi:10.1016/j.jcss.2012.05.012.
- 10 M. Bodirsky and S. Guzmán-Pro. Forbidden Tournaments and the Orientation Completion Problem. *SIAM J. Discret. Math.*, 39(1):170–205, 2025. doi:10.1137/23m1604849.
- 11 M. Bodirsky, S. Knäuer, and F. Starke. ASNP: A tame fragment of existential second-order logic. In *CiE 2020, Proceedings*, volume 12098 of *Lecture Notes in Computer Science*, pages 149–162. Springer, 2020. doi:10.1007/978-3-030-51466-2_13.
- 12 M. Bodirsky, F. R. Madelaine, and A. Mottet. A Proof of the Algebraic Tractability Conjecture for Monotone Monadic SNP. *SIAM J. Comput.*, 50(4):1359–1409, 2021. doi:10.1137/19M128466X.
- 13 M. Bodirsky and M. Pinsker. Canonical functions: a proof via topological dynamics. *Contributions Discret. Math.*, 16(2):36–45, 2021. doi:10.11575/cdm.v16i2.71724.
- 14 M. Bodirsky, M. Pinsker, and T. Tsankov. Decidability of Definability. *J. Symb. Log.*, 78(4):1036–1054, 2013. doi:10.2178/jsl.7804020.
- 15 P. Bourhis and C. Lutz. Containment in monadic disjunctive datalog, mmsnp, and expressive description logics. In C. Baral, J. P. Delgrande, and F. Wolter, editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the Fifteenth International Conference, KR 2016, Cape Town, South Africa, April 25–29, 2016, KR’16*, pages 207–216. AAAI Press, 2016. URL: <https://cdn.aaai.org/ocs/12847/12847-57546-1-PB.pdf>.
- 16 J. Brakensiek and V. Guruswami. An Algorithmic Blend of LPs and Ring Equations for Promise CSPs. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 436–455. SIAM, 2019. doi:10.1137/1.9781611975482.28.
- 17 A. A. Bulatov. A Dichotomy Theorem for Nonuniform CSPs. In C. Umans, editor, *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15–17, 2017*, pages 319–330. IEEE Computer Society, 2017. doi:10.1109/FOCS.2017.37.
- 18 A. K. Chandra and P. M. Merlin. Optimal implementation of conjunctive queries in relational data bases. In *Proceedings of the Ninth Annual ACM Symposium on Theory of Computing, STOC ’77*, pages 77–90, New York, NY, USA, 1977. Association for Computing Machinery. doi:10.1145/800105.803397.
- 19 T. Feder and M. Y. Vardi. The Computational Structure of Monotone Monadic SNP and Constraint Satisfaction: A Study through Datalog and Group Theory. *SIAM J. Comput.*, 28(1):57–104, 1998. doi:10.1137/S0097539794266766.
- 20 R. Feller and M. Pinsker. An algebraic proof of the graph orientation problem dichotomy for forbidden tournaments. Submitted, preprint: arXiv.2405.20263, 2024. doi:10.48550/arXiv.2405.20263.
- 21 M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. Mathematical Sciences Series. W. H. Freeman, 1979.
- 22 S. Guzmán-Pro. GMSNP and Finite Structures. Submitted, preprint: arXiv.2406.13529, 2024. doi:10.48550/arXiv.2406.13529.
- 23 W. Hodges. *A Shorter Model Theory*. Cambridge University Press, USA, 1997.
- 24 J. Hubička and J. Nešetřil. Universal Structures with Forbidden Homomorphisms. In Å. Hirvonen, J. Kontinen, R. Kossak, and A. Villaveces, editors, *Logic Without Borders - Essays on Set Theory, Model Theory, Philosophical Logic and Philosophy of Mathematics*, volume 5 of *Ontos Mathematical Logic*, pages 241–264. De Gruyter, 2015. doi:10.1515/9781614516873.241.

- 25 J. Hubička and J. Nešetřil. All those Ramsey classes (Ramsey classes with closures and forbidden homomorphisms). *Adv. Math.*, 356:106791, 2019. doi:10.1016/j.aim.2019.106791.
- 26 P. Kolaitis and M. Vardi. The decision problem for the probabilities of higher-order properties. In *Proceedings of the nineteenth annual ACM Symposium on Theory of Computing*, pages 425–435, 1987. doi:10.1145/28395.28441.
- 27 G. Kun. Constraints, MMSNP and expander relational structures. *Combinatorica*, 33(3):335–347, 2013. doi:10.1007/s00493-013-2405-4.
- 28 F. R. Madelaine. Universal Structures and the logic of Forbidden Patterns. *Log. Methods Comput. Sci.*, 5, 2009. doi:10.2168/LMCS-5(2:13)2009.
- 29 F. R. Madelaine. On the Containment of Forbidden Patterns Problems. In D. Cohen, editor, *Principles and Practice of Constraint Programming - CP 2010 - 16th International Conference, CP 2010, St. Andrews, Scotland, UK, September 6-10, 2010. Proceedings*, volume 6308 of *Lecture Notes in Computer Science*, pages 345–359. Springer, 2010. doi:10.1007/978-3-642-15396-9_29.
- 30 F. R. Madelaine and I. A. Stewart. Constraint Satisfaction, Logic and Forbidden Patterns. *SIAM J. Comput.*, 37(1):132–163, 2007. doi:10.1137/050634840.
- 31 A. Mottet, T. Nagy, M. Pinsker, and M. Wrona. Smooth Approximations and Relational Width Collapses. In N. Bansal, E. Merelli, and J. Worrell, editors, *48th International Colloquium on Automata, Languages, and Programming, ICALP 2021, July 12-16, 2021, Glasgow, Scotland (Virtual Conference)*, volume 198 of *LIPIcs*, pages 138:1–138:20. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2021. doi:10.4230/LIPIcs.ICALP.2021.138.
- 32 A. Mottet, T. Nagy, M. Pinsker, and M. Wrona. Collapsing the Bounded Width Hierarchy for Infinite-Domain Constraint Satisfaction Problems: When Symmetries Are Enough. *SIAM J. Comput.*, 53(6):1709–1745, 2024. doi:10.1137/22M1538934.
- 33 J. Nešetřil and V. Rödl. Ramsey classes of set systems. *J. Comb. Theory A*, 34(2):183–201, 1983. doi:10.1016/0097-3165(83)90055-9.
- 34 C. Papadimitriou and M. Yannakakis. Optimization, approximation, and complexity classes. In *Proceedings of the twentieth annual ACM Symposium on Theory of Computing*, pages 229–234, 1988. doi:10.1145/62212.62233.
- 35 J. Rydval. Homogeneity and Homogenizability: Hard Problems for the Logic SNP. In K. Bringmann, M. Grohe, G. Puppis, and O. Svensson, editors, *51st International Colloquium on Automata, Languages, and Programming (ICALP 2024)*, volume 297 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 150:1–150:20, Dagstuhl, Germany, 2024. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.ICALP.2024.150.
- 36 J. Rydval. Finitely Bounded Homogeneity Turned Inside-Out. To appear in *J. Math. Log.*, preprint: arXiv.2108.00452, 2025.
- 37 O. Shmueli. Equivalence of Datalog queries is undecidable. *J. Log. Program.*, 15(3):231–241, 1993. doi:10.1016/0743-1066(93)90040-N.
- 38 M. Wrochna and S. Živný. Improved hardness for H -colourings of G -colourable graphs. In S. Chawla, editor, *Proceedings of the 2020 ACM-SIAM Symposium on Discrete Algorithms, SODA 2020, Salt Lake City, UT, USA, January 5-8, 2020*, pages 1426–1435. SIAM, 2020. doi:10.1137/1.9781611975994.86.
- 39 D. Zhuk. A Proof of the CSP Dichotomy Conjecture. *J. ACM*, 67(5):30:1–30:78, 2020. doi:10.1145/3402029.