

Non-Homogenizable Classes of Finite Structures*

Albert Atserias¹ and Szymon Toruńczyk²

1 Universitat Politècnica de Catalunya, Spain

2 University of Warsaw, Poland

Abstract

Homogenization is a powerful way of taming a class of finite structures with several interesting applications in different areas, from Ramsey theory in combinatorics to constraint satisfaction problems (CSPs) in computer science, through (finite) model theory. A few sufficient conditions for a class of finite structures to allow homogenization are known, and here we provide a necessary condition. This lets us show that certain natural classes are not homogenizable: 1) the class of locally consistent systems of linear equations over the two-element field or any finite Abelian group, and 2) the class of finite structures that forbid homomorphisms from a specific MSO-definable class of structures of treewidth two. In combination with known results, the first example shows that, up to pp-interpretability, the CSPs that are solvable by local consistency methods are distinguished from the rest by the fact that their classes of locally consistent instances are homogenizable. The second example shows that, for MSO-definable classes of forbidden patterns, treewidth one versus two is the dividing line to homogenizability.

1998 ACM Subject Classification G.2 Discrete mathematics

Keywords and phrases Fraïssé class, amalgamation class, reduct, Constraint Satisfaction Problem, bounded width

Digital Object Identifier 10.4230/LIPIcs.CSL.2016.16

1 Introduction

A relational structure with a countable domain is called homogeneous if it is highly symmetric in the precise technical sense that any isomorphism between any two of its finite induced substructures extends to an automorphism of the whole structure. In many areas of combinatorics, logic, discrete geometry, and computer science, homogeneous structures abound, often in the form of nicely behaved limit objects for classes of finite structures. Typical examples include the Rado graph \mathcal{R} , which can be seen as the limit of the class of all finite graphs; the linear order of the rational numbers \mathcal{Q} , seen as the limit of all finite linear orders; or the countable Urysohn space \mathcal{U} , the limit of all rational metric spaces. The literature on the subject is very extensive; we refer the reader to [16] for a recent survey.

Homogeneous structures arise as limits of well-behaved classes of finite structures, in a way made precise by Fraïssé's theorem, which describes them combinatorially in a finitary manner. It states that a homogeneous structure is characterized, up to isomorphism, by its age, i.e., the class of its finite induced substructures. Moreover, classes of finite structures arising as ages of homogeneous structures are precisely Fraïssé classes, i.e., classes closed under taking induced substructures and under amalgamation – a form of glueing pairs of structures along a common induced substructure (see [13] and Section 2 for precise definitions).

* Partly funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement ERC-2014-CoG 648276 AUTAR).



Thanks to Fraïssé’s theorem, combinatorial arguments involving finite structures can often be replaced by, or aided by, arguments involving highly symmetric, infinite structures. In combinatorics, for example, homogeneous structures appear unavoidably in structural Ramsey theory [17]. At the intersection between combinatorics and computer science, homogeneous structures appear in the theory of logical limit laws for various models of random graphs [15]. In computer science proper, homogeneous structures appear in the theory of constraint satisfaction problems [5], automata theory [6], and verification [7].

One of the advantages of working with homogeneous structures, rather than classes of finite structures, is that their automorphism groups are very rich. For example, over a finite relational signature, the homogeneity of the structure immediately implies that, up to automorphism, it has finitely many elements, pairs of elements, triples, etc. In model theoretic terms, this means that the structure is ω -categorical by the classical Ryll-Nardzewski theorem, and its first-order theory admits elimination of quantifiers. In turn, since in any such structure there are only finitely many first-order definable relations of each arity, homogeneous structures over finite relational signatures are, in a strong technical way, close to being finite.

Thus, with Fraïssé’s theorem in hand and the many applications of homogeneous structures in mind, it becomes quite important a task to identify more Fraïssé classes. More generally, one would like to identify classes of finite structures that are perhaps not Fraïssé classes themselves, but appear as reducts of some Fraïssé class over a richer yet finite signature. Such classes of finite structures are called *homogenizable* [9]. The point in case is that the lifted Fraïssé class can be thought of as taming its reduct by providing a homogeneous structure that plays the role of limit object for it. Many of the application examples mentioned above do actually go through lifted Fraïssé classes and their corresponding homogeneous Fraïssé limits. See [14] and the references therein for a discussion on this.

A noticeable amount of work has gone into providing sufficient conditions for a class of finite structures to be homogenizable. Instances include the model-theoretic methods of Covington [9], and the combinatorial explicit constructions of Hubička and Nešetřil [14]. Here we provide a combinatorial necessary condition for homogenizability (Theorem 3.2 in Section 3). This allows us to prove that certain natural classes of finite structures previously considered in the literature are not homogenizable.

Our first example of a non-homogenizable class comes from the theory of constraint satisfaction problems (CSPs). We show that the class of locally consistent systems of linear equations over the two-element field is not homogenizable. More generally, the result holds for systems of equations over any finite Abelian group. This answers a question first raised by the first author of this paper in [2]. Precisely, by a locally consistent system of equations we mean one whose satisfiability cannot be refuted by the (j, k) -consistency algorithm for small j and k , which is a well-studied heuristic algorithm for solving CSPs. Moreover, in combination with the resolution of the Bounded Width Conjecture by Barto and Kozik [4], this shows that the constraint languages whose classes of locally consistent instances are homogenizable are, up to pp-interpretability, precisely those that are solvable by local consistency methods. All this is worked out in Section 4.

In Section 5 we give a second example of a non-homogenizable class that, in this case, is motivated by the works of Hubička and Nešetřil [14], and Erdős, Tardif, and Tardos [10]. It was shown in [14] that every class of finite structures that is of the form $\text{Forb}_h(\mathcal{F})$, where \mathcal{F} is a *regular* class of connected finite structures, is homogenizable. In words, $\text{Forb}_h(\mathcal{F})$ is the class of finite structures that do not admit homomorphisms from any structure in \mathcal{F} . The notion of regularity considered in [14] is closely related to the notion of regularity in automata theory, and agrees with it on coloured paths and trees. However, our second example shows

that even if \mathcal{F} is MSO-definable and has maximum treewidth two, the class $\text{Forb}_h(\mathcal{F})$ need not be homogenizable. This shows that for MSO-definable classes, treewidth one versus two of the forbidden structures in \mathcal{F} is the dividing line to homogenizability.

2 Preliminaries

Signatures, structures, reducts, and expansions

A relational signature Σ is a set of relation symbols R_1, R_2, \dots , each with an associated natural number called its arity. In this paper, we consider only finite relational signatures. A Σ -structure $\mathbb{A} = (A; R_1^{\mathbb{A}}, R_2^{\mathbb{A}}, \dots)$ is composed of a set A , called its domain, and a relation $R^{\mathbb{A}} \subseteq A^k$ on A for each R in Σ , where k is the arity of R . We say that $R^{\mathbb{A}}$ is the interpretation of R in \mathbb{A} . A Σ -structure is sometimes referred to as a structure over the signature Σ . If Σ^+ is a signature that contains Σ and \mathbb{A}^+ is a Σ^+ -structure, then the Σ -reduct of \mathbb{A}^+ is the structure \mathbb{A} obtained from \mathbb{A}^+ by forgetting all relations from $\Sigma^+ - \Sigma$. In this case, we also say that \mathbb{A}^+ is an expansion of \mathbb{A} . Expansions and reducts are also called lifts and shadows, respectively.

Substructures, homomorphisms, and embeddings

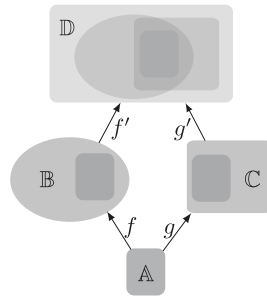
If \mathbb{A} is a Σ -structure and X is a subset of its domain A , we write $\mathbb{A}[X]$ for the substructure of \mathbb{A} induced by X , that is, the Σ -structure with domain X in which each relation symbol R in Σ is interpreted by $R^{\mathbb{A}} \cap X^k$, where k is the arity of R .

Let \mathbb{A} and \mathbb{B} be structures over the same relational signature Σ . Let A and B denote their domains. A homomorphism from \mathbb{A} to \mathbb{B} is a mapping $f : A \rightarrow B$ for which the inclusion $f(R^{\mathbb{A}}) \subseteq R^{\mathbb{B}}$ holds for every R in Σ . The homomorphism is strong if in addition the inclusion $f(A^k - R^{\mathbb{A}}) \subseteq B^k - R^{\mathbb{B}}$ holds for every R in Σ , where k is the arity of R . A monomorphism from \mathbb{A} to \mathbb{B} is an injective homomorphism. Whenever A is a subset of B and the inclusion $A \rightarrow B$ is a monomorphism, we say that \mathbb{A} is a substructure of \mathbb{B} . An embedding from \mathbb{A} to \mathbb{B} is an injective strong homomorphism. Whenever A is a subset of B and the inclusion $A \rightarrow B$ is an embedding, we say that \mathbb{A} is an induced substructure of \mathbb{B} . An isomorphism from \mathbb{A} to \mathbb{B} is a surjective embedding. If there is an isomorphism from \mathbb{A} to \mathbb{B} we say that the two structures are isomorphic. If $f : A \rightarrow B$ is a partial mapping with domain $X \subseteq A$ and image $Y \subseteq B$, we say that f is a partial homomorphism from \mathbb{A} to \mathbb{B} if it is a homomorphism from $\mathbb{A}[X]$ to $\mathbb{B}[Y]$. We write $\left(\begin{smallmatrix} \mathbb{B} \\ \mathbb{A} \end{smallmatrix}\right)$ to denote the set of all embeddings from \mathbb{A} to \mathbb{B} .

Amalgamation

If \mathbb{B} and \mathbb{C} are Σ -structures with domains B and C , we write $\mathbb{B} \cup \mathbb{C}$ for their union, i.e. the Σ -structure with domain $B \cup C$ and relations $R^{\mathbb{B} \cup \mathbb{C}} = R^{\mathbb{B}} \cup R^{\mathbb{C}}$ for every R in Σ . Let f and g be embeddings from the same structure \mathbb{A} into structures \mathbb{B} and \mathbb{C} , respectively. The structure \mathbb{D} is an amalgam of \mathbb{B} and \mathbb{C} through f and g if there exist embeddings f' and g' from \mathbb{B} to \mathbb{D} and \mathbb{C} to \mathbb{D} , respectively, such that the diagram in Figure 1 commutes, i.e., $f' \circ f = g' \circ g$.

We say that \mathbb{D} is a strong amalgam if $f'(B) \cap g'(C) = (f' \circ f)(A) = (g' \circ g)(A)$, where A , B and C denote the domains of \mathbb{A} , \mathbb{B} and \mathbb{C} , respectively. We say that \mathbb{D} is a free amalgam if it is strong and, additionally, $\mathbb{D} = \mathbb{D}[f'(B)] \cup \mathbb{D}[g'(C)]$. We also say that \mathbb{D} is the union of \mathbb{B} and \mathbb{C} amalgamated along \mathbb{A} through f and g via f' and g' . Note that the free amalgam of \mathbb{B} and \mathbb{C} through f and g is uniquely defined up to isomorphism, and is isomorphic to the disjoint union of \mathbb{B} and \mathbb{C} , quotiented by the equivalence relation identifying $f(x)$ with $g(x)$,



■ **Figure 1** Amalgamation of \mathbb{B} and \mathbb{C} through f and g . All mappings are embeddings.

for $x \in \mathbb{A}$. We denote this free amalgam $f \cup_{\mathbb{A}} g$. When f and g are implicit, we denote it $\mathbb{B} \cup_{\mathbb{A}} \mathbb{C}$. We also say that \mathbb{B} and \mathbb{C} are glued along \mathbb{A} .

Classes of structures

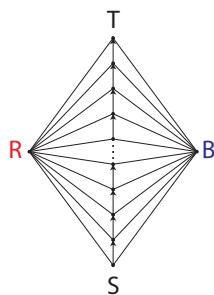
All our structures will have finite or countably infinite domain. Moreover we assume that all structures have a domain that is a subset of a common background countable set, say \mathbb{N} . For a fixed signature Σ , a class of structures is a set of structures that is closed under isomorphisms, i.e. if \mathbb{A} and \mathbb{B} are isomorphic structures and \mathbb{A} belongs to the class, then \mathbb{B} also belongs to the class. A class of structures \mathcal{C} is closed under amalgamation if for every two embeddings f and g from the same structure \mathbb{A} in \mathcal{C} into structures \mathbb{B} and \mathbb{C} in \mathcal{C} , there exists in \mathcal{C} an amalgam of \mathbb{B} and \mathbb{C} through f and g . A class of finite structures is an amalgamation class, also called a Fraïssé class, if it is closed under taking induced substructures and amalgamation. For example, the class of all finite graphs is an amalgamation class – in fact, it is closed under free amalgamation – so is the class of all finite digraphs. The class of all finite linear orders is also an amalgamation class, although it is not closed under free amalgamation. Fraïssé’s theorem states that a class is Fraïssé if and only if it is the class of finite induced substructures of a homogeneous structure.

For two signatures Σ and Σ^+ with the second containing the first, if \mathcal{C} and \mathcal{C}^+ are classes of Σ -structures and Σ^+ -structures, respectively, then we say that \mathcal{C} is the Σ -reduct of \mathcal{C}^+ if \mathcal{C} is the class of Σ -reducts of the structures in \mathcal{C}^+ .

Homogenizable classes

We say that a class of Σ -structures is *homogenizable* if there is a signature Σ^+ extending Σ , and an amalgamation class \mathcal{C}^+ over Σ^+ , such that \mathcal{C} is the Σ -reduct of \mathcal{C}^+ . For a class of Σ -structures \mathcal{F} , let $\text{Forb}_h(\mathcal{F})$ denote the class of all finite Σ -structures \mathbb{A} such that for no \mathbb{F} in \mathcal{F} there is a homomorphism from \mathbb{F} to \mathbb{A} . Hubička and Nešetřil define a notion of regularity, which we call HN-regularity (we omit its technical definition), and prove in Theorem 3.1 from [14] that if \mathcal{F} is a HN-regular class of finite connected structures, then $\text{Forb}_h(\mathcal{F})$ is homogenizable. In particular, if \mathcal{F} is finite, then $\text{Forb}_h(\mathcal{F})$ is homogenizable.

► **Example 2.1.** Let Σ be the signature that consists of one binary predicate \vec{E} and two unary predicates S and T . Let \mathbb{P}_n denote a simple directed \vec{E} -path of length n from a unique S -colored node to a unique T -colored node. The class $\mathcal{F} = \{\mathbb{P}_n : n \geq 0\}$ is HN-regular, and therefore, by [14], the class $\text{Forb}_h(\mathcal{F})$ is homogenizable. It consists of digraphs whose nodes are possibly labeled with S or T , and there is no directed path from an S -labeled node to a T -labeled node. We show that $\text{Forb}_h(\mathcal{F})$ is homogenizable by a direct construction. Let Σ^+



■ **Figure 2** Forbidden structure \mathbb{F}_n .

be the extension of Σ by two unary predicates I and O . Let \mathcal{C}^+ consist of all Σ^+ -structures \mathbb{A}^+ such that the domain of \mathbb{A}^+ is partitioned into $I^{\mathbb{A}^+}$ and $O^{\mathbb{A}^+}$, and that $S^{\mathbb{A}^+} \subseteq I^{\mathbb{A}^+}$, $T^{\mathbb{A}^+} \subseteq O^{\mathbb{A}^+}$, and there are no \vec{E} -edges starting in $I^{\mathbb{A}^+}$ and ending in $O^{\mathbb{A}^+}$. Then \mathcal{C}^+ is an amalgamation class, as it is closed under free amalgamation. The class $\text{Forb}_h(\mathcal{F})$ is the Σ -reduct of \mathcal{C}^+ : a structure \mathbb{A} in $\text{Forb}_h(\mathcal{F})$ extends to a structure \mathbb{A}^+ in \mathcal{C}^+ , in which $I^{\mathbb{A}^+}$ is the set of vertices reachable from $S^{\mathbb{A}}$ by a directed \vec{E} -path, and $O^{\mathbb{A}^+}$ is its complement. ◀

3 Necessary condition for homogenizability

Fix a finite relational signature Σ . In this section all structures are over this signature, or over a signature Σ^+ that extends Σ . Before we state the necessary condition for homogenizability we need some notation and terminology.

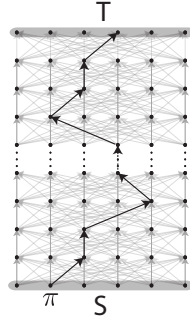
Let \mathcal{C} be a class of finite structures. If \mathbb{A} , \mathbb{L} and \mathbb{R} are structures in \mathcal{C} , and $L : \mathbb{A} \rightarrow \mathbb{L}$ and $R : \mathbb{A} \rightarrow \mathbb{R}$ are embeddings such that no amalgam of \mathbb{L} and \mathbb{R} through L and R is in \mathcal{C} , then we say that $L : \mathbb{A} \rightarrow \mathbb{L}$, $R : \mathbb{A} \rightarrow \mathbb{R}$ is a *diagram that witnesses failure of amalgamation* of \mathcal{C} . We illustrate the definitions with a running example.

► **Example 3.1** (Running example). Let \mathbb{F}_n denote the structure depicted in Figure 2, with n vertices in the middle column.

The signature Σ of this structure consists of one binary predicate E , the undirected edges, one binary predicate \vec{E} , the vertical directed edges, and four unary predicates R (for *red*), B (for *blue*), S (for *source*), and T (for *target*), each appearing in the structure exactly once. Observe that the colours S and T ensure that \mathcal{F} is an antichain in the homomorphism pre-order, i.e. there are no homomorphisms from \mathbb{F}_n to \mathbb{F}_m if $n \neq m$. Let $\mathcal{C} = \text{Forb}_h(\mathcal{F})$. In the running example, we will demonstrate that the class \mathcal{C} is not homogenizable.

Choose a large natural number n . Let \mathbb{L} denote the left part of the structure \mathbb{F}_n obtained by removing the blue vertex (labeled B). Symmetrically, let \mathbb{R} denote the right part of \mathbb{F}_n obtained by removing the red vertex (labeled R). Let \mathbb{A} denote the intersection of \mathbb{L} and \mathbb{R} , i.e., the \vec{E} -path with n vertices starting at the S -labeled vertex and ending at the T -labeled vertex. Let $L : \mathbb{A} \rightarrow \mathbb{L}$ and $R : \mathbb{A} \rightarrow \mathbb{R}$ be the inclusion mappings. Then any amalgamation of L and R necessarily is a homomorphic image of \mathbb{F}_n . Hence $L : \mathbb{A} \rightarrow \mathbb{L}$, $R : \mathbb{A} \rightarrow \mathbb{R}$ is a diagram that witnesses failure of amalgamation of \mathcal{C} . ◀

Let $L : \mathbb{A} \rightarrow \mathbb{L}$, $R : \mathbb{A} \rightarrow \mathbb{R}$ be a diagram that witnesses failure of amalgamation of \mathcal{C} . For a structure \mathbb{J} and a partial mapping $C : \binom{\mathbb{J}}{\mathbb{A}} \rightarrow \{L, R\}$, let \mathbb{J}^C be the structure that is obtained by gluing to \mathbb{J} , for each π in $\text{Dom}(C)$, a fresh copy of either \mathbb{L} or \mathbb{R} depending on whether $C(\pi) = L$ or $C(\pi) = R$. More formally, \mathbb{J}^C is defined by induction on the cardinality



■ **Figure 3** The structure $\mathbb{A} \otimes m$, with an embedding $\pi \in \mathcal{E}_{\mathbb{A},m}$.

of the domain of C : if $\text{Dom}(C) = \emptyset$, then $\mathbb{J}^C = \mathbb{J}$; otherwise, if $C = C' \cup \{\pi \mapsto \sigma\}$, where $\pi \in \binom{\mathbb{J}}{\mathbb{A}}$ and $\sigma \in \{L, R\}$, then define $\mathbb{J}^C = \pi' \cup_{\mathbb{A}} \sigma$, where $\pi' : \mathbb{A} \rightarrow \mathbb{J}^{C'}$ is $\pi : \mathbb{A} \rightarrow \mathbb{J}$ composed with the identity embedding from \mathbb{J} to $\mathbb{J}^{C'}$.

For a natural number m and a Σ -structure \mathbb{A} with domain A , let $\mathbb{A} \otimes m$ denote the structure with domain $A \times [m]$ in which the interpretation of a relation R in Σ of arity k is the set of all tuples $((a_1, i_1), (a_2, i_2), \dots, (a_k, i_k))$ where $(a_1, \dots, a_k) \in R^{\mathbb{A}}$ and $i_1, \dots, i_k \in [m]$. Observe that every function $f : A \rightarrow [m]$ induces an embedding $\pi_f : \mathbb{A} \rightarrow \mathbb{A} \otimes m$, defined by $\pi_f(a) = (a, f(a))$. Let $\mathcal{E}_{\mathbb{A},m}$ denote the set of all embeddings of the form π_f for $f : A \rightarrow [m]$. In particular, $\mathcal{E}_{\mathbb{A},m}$ is a subset of $\binom{\mathbb{A} \otimes m}{\mathbb{A}}$ containing exactly $m^{|A|}$ embeddings.

A diagram $L : \mathbb{A} \rightarrow \mathbb{L}, R : \mathbb{A} \rightarrow \mathbb{R}$ is *confusing* for \mathcal{C} if the following conditions hold:

1. it witnesses failure of amalgamation of \mathcal{C} , and
2. for every natural number m , if $\mathbb{J} = \mathbb{A} \otimes m$, then for every coloring $C : \mathcal{E}_{\mathbb{A},m} \rightarrow \{L, R\}$ the structure \mathbb{J}^C belongs to the class \mathcal{C} .

Its *order* is the cardinality of the domain of \mathbb{A} .

► **Theorem 3.2.** *If \mathcal{C} is a homogenizable class of finite structures, then there exists a natural number r such that every confusing diagram for \mathcal{C} has order at most r .*

This theorem is the main technical result of this paper. Before we prove it, we illustrate it by applying it to our running example.

► **Example 3.3.** Fix natural numbers m and n . Let $L : \mathbb{A} \rightarrow \mathbb{L}$ and $R : \mathbb{A} \rightarrow \mathbb{R}$ be defined as in Example 3.1. The structure $\mathbb{J} = \mathbb{A} \otimes m$ is depicted in Figure 3.

Its domain is $[n] \times [m]$, and every element $(i, j) \in [n] \times [m]$ with $i \leq n - 1$ is connected by an \vec{E} -edge to every element $(i + 1, k) \in [n] \times [m]$. The embeddings $\mathcal{E}_{\mathbb{A},m}$ correspond to functions $f : [n] \rightarrow [m]$. If $C : \mathcal{E}_{\mathbb{A},m} \rightarrow \{L, R\}$ is a coloring, then \mathbb{J}^C is obtained by considering all functions $f : [n] \rightarrow [m]$, and connecting every vertex along the path $\{(i, f(i)) : 1 \leq i \leq n\}$ to a fresh vertex which is red if $C(f) = L$, and blue if $C(f) = R$. Observe that no structure \mathbb{F} in \mathcal{F} maps homomorphically to \mathbb{J}^C . Therefore, \mathbb{J}^C belongs to $\mathcal{C} = \text{Forb}_h(\mathcal{F})$. Since m is arbitrary, this shows that the diagram $L : \mathbb{A} \rightarrow \mathbb{L}, R : \mathbb{A} \rightarrow \mathbb{R}$ is confusing for \mathcal{C} . Since its order is $|A| = n$, and n is arbitrary, Theorem 3.2 implies that \mathcal{C} is not a reduct of any amalgamation class. ◀

Theorem 3.2 follows easily from Lemma 3.4 stated below.

Let $L : \mathbb{A} \rightarrow \mathbb{L}, R : \mathbb{A} \rightarrow \mathbb{R}$ witness failure of amalgamation of \mathcal{C} . An (L, R) -*confusion* for \mathcal{C} is a structure \mathbb{J} in \mathcal{C} , together with a set $\mathcal{E} \subseteq \binom{\mathbb{J}}{\mathbb{A}}$, such that \mathbb{J}^C is in \mathcal{C} for every coloring $C : \mathcal{E} \rightarrow \{L, R\}$. For $\mathcal{E} \subseteq \binom{\mathbb{J}}{\mathbb{A}}$ and a natural number r bounded by the cardinality of the

domain of \mathbb{A} , let \mathcal{E}_r denote the set of all restrictions $\pi|_X$ of π in \mathcal{E} , where X ranges over all r -element subsets of the domain of \mathbb{A} .

► **Lemma 3.4.** *Let r and t be natural numbers, and let \mathcal{C} be a class of Σ -structures. There exist numbers p and q (depending on r and t only) such that the following condition implies that \mathcal{C} is not a reduct of any amalgamation class over a signature with at most t predicates of arity at most r :*

there is a diagram $L : \mathbb{A} \rightarrow \mathbb{L}$, $R : \mathbb{A} \rightarrow \mathbb{R}$ that witnesses failure of amalgamation of \mathcal{C} and of order at least r , and there is an (L, R) -confusion $(\mathbb{J}, \mathcal{E})$ for \mathcal{C} satisfying

$$|\mathcal{E}| > p \cdot |\mathcal{E}_r| + q \binom{|\mathbb{A}|}{r}. \quad (1)$$

First we show how Theorem 3.2 follows from Lemma 3.4.

Proof of Theorem 3.2. Suppose that \mathcal{C} has confusing diagrams of arbitrarily large order. For every two fixed natural numbers r and t , we apply Lemma 3.4 to conclude that \mathcal{C} is not a reduct of an amalgamation class over a signature with t symbols of arity at most r . Let p and q be as in the statement of the lemma. Consider a confusing diagram $L : \mathbb{A} \rightarrow \mathbb{L}$, $R : \mathbb{A} \rightarrow \mathbb{R}$ and let n be its order. Fix a natural number m , and let $\mathbb{J} = \mathbb{A} \otimes m$ and $\mathcal{E} = \mathcal{E}_{\mathbb{A}, m}$. Then $(\mathbb{J}, \mathcal{E})$ is an (L, R) -confusion for \mathcal{C} , by the definition of confusing diagram, and $|\mathcal{E}| = m^n$ and $|\mathcal{E}_r| = m^r$. Since the order n of the diagram can be chosen arbitrarily large, we can assume $n > r$. Taking m large enough, so that $p \cdot m^{n-1} > q \binom{n}{r}$ and $m > 2p$, we get:

$$p \cdot |\mathcal{E}_r| + q \binom{|\mathbb{A}|}{r} = p \cdot m^r + q \binom{n}{r} \leq p \cdot m^{n-1} + p \cdot m^{n-1} < m^n = |\mathcal{E}|,$$

which gives condition (1) in Lemma 3.4. Since t and r were arbitrary, this proves that \mathcal{C} is not the reduct of an amalgamation class. ◀

It remains to prove the lemma.

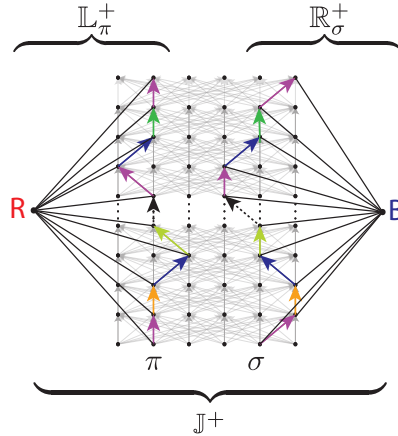
Proof of Lemma 3.4. Fix natural numbers r and t . In anticipation of the proof, let q be the maximum number of atomic types of r -tuples over any signature with at most t predicates of arity at most r , and let $p = \lceil \log_2(q) \rceil$. Suppose that \mathcal{C} is a class of Σ -structures as in the lemma, with a diagram $L : \mathbb{A} \rightarrow \mathbb{L}$, $R : \mathbb{A} \rightarrow \mathbb{R}$ that witnesses its failure of amalgamation, and an (L, R) -confusion $(\mathbb{J}, \mathcal{E})$ satisfying condition (1) from Lemma 3.4.

Let \mathbb{B}^+ be a Σ^+ -structure with domain B and let $f : A \rightarrow B$ be a function from some set A to B . Define the *pullback*¹ $f^*(\mathbb{B}^+)$ as the Σ^+ -structure with universe A , where the interpretation of a relation symbol R of Σ^+ of arity k is $f^{-1}(R^{\mathbb{B}^+})$, i.e., the inverse image of the interpretation of R in \mathbb{B}^+ under the mapping $f : A^k \rightarrow B^k$. By definition, $f^*(\mathbb{B}^+)$ is the unique Σ^+ -structure on A for which f is a strong homomorphism.

By definition of the structure \mathbb{J}^C , there is a distinguished embedding of \mathbb{J} into \mathbb{J}^C . Therefore, by composition, any embedding $\pi : \mathbb{A} \rightarrow \mathbb{J}$ in \mathcal{E} defines an embedding of \mathbb{A} into \mathbb{J}^C , denoted $\hat{\pi} : \mathbb{A} \rightarrow \mathbb{J}^C$. Note that for any expansion \mathbb{J}^+ of \mathbb{J}^C , the pullback $\hat{\pi}^*(\mathbb{J}^+)$ is an expansion of \mathbb{A} , which is isomorphic (via $\hat{\pi}$) to an induced substructure of \mathbb{J}^+ .

► **Claim 1.** *There is a coloring $C : \mathcal{E} \rightarrow \{L, R\}$ such that, for every expansion \mathbb{J}^+ of \mathbb{J}^C over the signature Σ^+ , there are two embeddings π and σ in \mathcal{E} such that the pullbacks $\hat{\pi}^*(\mathbb{J}^+)$ and $\hat{\sigma}^*(\mathbb{J}^+)$ are equal, but $C(\pi) \neq C(\sigma)$.*

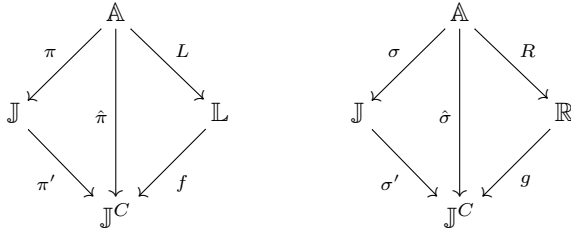
¹ the word *pullback* has two distinct uses in mathematics – the categorical one, as the dual of a pushout, and another one, used e.g. in differential geometry, as the dual of a pushforward. Our usage is in analogy to the latter.



■ **Figure 4** Two embeddings $\pi, \sigma \in \mathcal{E}$ with $C(\pi) = L$ and $C(\sigma) = R$, in the context of the running example. The colored arrows depict various predicates of the stipulated signature Σ^+ extending Σ (in general, they don't need to be binary). The fact that the sequences of colors along π and along σ are the same corresponds to the assumption that the pullbacks $\hat{\pi}^*(\mathbb{J}^+)$ and $\hat{\sigma}^*(\mathbb{J}^+)$ are equal. Therefore, the marked substructures \mathbb{L}_π^+ and \mathbb{R}_σ^+ of \mathbb{J}^+ (which correspond to \mathbb{L}^+ and \mathbb{R}^+ in the proof via the mappings f and g) have an isomorphic substructure (isomorphic to \mathbb{A}^+ in the proof). An amalgamation in \mathcal{C}^+ of \mathbb{L}_π^+ and \mathbb{R}_σ^+ along this substructure would yield as a Σ -reduct an amalgamation in \mathcal{C} of \mathbb{L} and \mathbb{R} along \mathbb{A} , a contradiction.

We show how the claim yields the lemma. Figure 4 illustrates the proof.

Assume that \mathcal{C} is the class of Σ -reducts of a class of Σ^+ -structure \mathcal{C}^+ . To reach a contradiction, suppose that \mathcal{C}^+ is closed under amalgamation. Let \mathcal{C} be as in the claim. Since \mathbb{J}^C belongs to \mathcal{C} by the definition of confusion, there exists an expansion \mathbb{J}^+ of \mathbb{J}^C in \mathcal{C}^+ . Let π and σ be as in the conclusion of the claim, and suppose without loss of generality that $C(\pi) = L$ and $C(\sigma) = R$. By the definition of \mathbb{J}^C , the embeddings $\pi : \mathbb{A} \rightarrow \mathbb{J}$ and $L : \mathbb{A} \rightarrow \mathbb{L}$ induce embeddings $\hat{\pi}, \pi', f$, such that the diagram to the left below commutes:



Let $\mathbb{L}^+ = f^*(\mathbb{J}^+)$ be the pullback structure; this structure is an expansion of \mathbb{L} . Moreover, \mathbb{L}^+ belongs to the class \mathcal{C}^+ , since it is a pullback under an injective mapping, and hence \mathbb{L}^+ is isomorphic to an induced substructure \mathbb{L}_π^+ of \mathbb{J}^+ , which is in \mathcal{C}^+ .

Similarly, the embeddings $\sigma : \mathbb{A} \rightarrow \mathbb{J}$ and $R : \mathbb{A} \rightarrow \mathbb{R}$ induce embeddings $\hat{\sigma}, \sigma', g : \mathbb{R} \rightarrow \mathbb{J}^C$, such that the diagram to the right above commutes. Let $\mathbb{R}^+ = g^*(\mathbb{J}^+)$ be the pullback structure, which is an expansion of \mathbb{R} , isomorphic to an induced substructure \mathbb{R}_σ^+ of \mathbb{J}^+ , hence belongs to the class \mathcal{C}^+ .

Let \mathbb{A}^+ be the pullback $\hat{\pi}^*(\mathbb{J}^+)$, which, by the claim, is the same as the pullback $\hat{\sigma}^*(\mathbb{J}^+)$. Note that by commutativity of the diagram to the left above, the pullback $\hat{\pi}^*(\mathbb{J}^+)$ is the same as the pullback $L^*(\mathbb{L}^+)$. Similarly, $\hat{\sigma}^*(\mathbb{J}^+)$ is the same as $R^*(\mathbb{R}^+)$. In other words, L is an embedding of \mathbb{A}^+ into \mathbb{L}^+ , and R is an embedding of \mathbb{A}^+ into \mathbb{R}^+ . Since \mathcal{C}^+ is closed under

amalgamation, there exists an amalgamation of the diagram $L : \mathbb{A}^+ \rightarrow \mathbb{L}^+$ and $R : \mathbb{A}^+ \rightarrow \mathbb{R}^+$, which consists of a structure \mathbb{U}^+ in \mathcal{C}^+ and two embeddings $L' : \mathbb{L}^+ \rightarrow \mathbb{U}^+$, $R' : \mathbb{R}^+ \rightarrow \mathbb{U}^+$. Taking Σ -reducts, we obtain an amalgamation in \mathcal{C} of $L : \mathbb{A} \rightarrow \mathbb{L}$ and $R : \mathbb{A} \rightarrow \mathbb{R}$. But L, R were suppose to witness failure of amalgamation in \mathcal{C} – a contradiction proving that \mathcal{C}^+ cannot be closed under amalgamation.

Next we show how to prove Claim 1 and hence Lemma 3.4. Call any embedding in \mathcal{E} a *spot*, and any restriction of a spot to an r -element subset of the domain of \mathbb{A} a *partial spot*. For each coloring C of the spots, and each two spots π and σ , define $\pi \approx_C \sigma$ if and only if $C(\pi) = C(\sigma)$. For each coloring D of the partial spots, and each two spots π and σ , define $\pi \sim_D \sigma$ if and only if $D(\pi|_X) = D(\sigma|_X)$ for every r -element subset X of the domain of \mathbb{A} . Both are equivalence relations on spots.

► **Claim 2.** *There is a coloring C of the spots using two colors, such that for all colorings D of the partial spots using q colors, there is a pair of spots π and σ such that $\pi \sim_D \sigma$ but $\pi \not\approx_C \sigma$.*

We omit the proof of Claim 2, which is by a counting argument relying on the inequality (1).

Finally we use Claim 2 to prove Claim 1. Let C be the coloring of Claim 2 with the two colors interpreted as the embeddings $L : \mathbb{A} \rightarrow \mathbb{L}$ and $R : \mathbb{A} \rightarrow \mathbb{R}$. For each expansion \mathbb{J}^+ of \mathbb{J}^C , let D be the coloring of partial spots defined as follows. Fix an arbitrary linear ordering of the domain of \mathbb{A} , and for each spot π and each r -element subset $X = \{a_1 < \dots < a_r\}$ of the domain of \mathbb{A} , let $D(\pi|_X)$ be the atomic type of the tuple $(\pi(a_1), \dots, \pi(a_r))$ in \mathbb{J}^+ . This is a coloring of partial spots using at most q colors. By Claim 2, there is a pair of spots π and σ such that $\pi \sim_D \sigma$ but $\pi \not\approx_C \sigma$. From $\pi \sim_D \sigma$ and the fact that r is at least as large as the maximum arity of any new predicate in Σ^+ , it follows that the pullbacks $\hat{\pi}^*(\mathbb{J}^+)$ and $\hat{\sigma}^*(\mathbb{J}^+)$ are equal. On the other hand, from $\pi \not\approx_C \sigma$ we get $C(\pi) \neq C(\sigma)$ by definition. This proves Claim 1 and Lemma 3.4. ◀

4 Classes of consistent structures

In this section we work out the first of our two examples of non-homogenizable classes. We start by defining some basic notions from the theory of constraint satisfaction problems as described, for example, in Chapter 6 of the monograph [12]. Recall that, for a structure \mathbb{T} , we write $\text{CSP}(\mathbb{T})$ for the class of all finite structures \mathbb{I} over the same signature as \mathbb{T} for which there is a homomorphism from \mathbb{I} to \mathbb{T} . The \mathbb{I} 's are called instances, the \mathbb{T} 's are called templates.

4.1 Local consistency

Let Σ be a relational signature, let \mathbb{A} and \mathbb{B} be Σ -structures, and let k and l be integers such that $1 \leq k \leq l$. A (k, l) -consistent family on \mathbb{A} and \mathbb{B} is a non-empty family \mathcal{F} of partial homomorphisms from \mathbb{A} to \mathbb{B} , such that the following three conditions hold for each f in \mathcal{F} :

1. $|\text{Dom}(f)| \leq l$,
2. if X is a subset of $\text{Dom}(f)$, then $f|_X$ is in \mathcal{F} ,
3. if $|\text{Dom}(f)| \leq k$ and X is a subset of A such that $\text{Dom}(f) \subseteq X$ and $|X| \leq l$, then there exists g in \mathcal{F} such that $\text{Dom}(g) = X$ and $f \subseteq g$.

If there is a (k, l) -consistent family on \mathbb{A} and \mathbb{B} , then we say that \mathbb{A} is (k, l) -consistent with respect to \mathbb{B} . Note for later use that the class of structures that are (k, l) -consistent with respect to \mathbb{B} is closed under inverse homomorphisms: if there is a homomorphism from \mathbb{A}' to \mathbb{A} , and \mathbb{A} is (k, l) -consistent with respect to \mathbb{B} , then \mathbb{A}' is also (k, l) -consistent with respect

to \mathbb{B} . To see this, it suffices to compose the homomorphism from \mathbb{A}' to \mathbb{A} with each partial homomorphism in the (k, l) -consistent family for \mathbb{A} to get a (k, l) -consistent family for \mathbb{A}' .

We describe the special case of $(2, 3)$ -consistency in terms of a pebble game. The game is played between spoiler and duplicator, each having three pebbles, numbered 1, 2 and 3. Spoiler can place his pebbles on the nodes of \mathbb{A} , while duplicator can place his pebbles on the nodes of \mathbb{B} . They can also keep the pebbles in their pockets, in which they have all pebbles at the beginning of the game. The game proceeds in rounds as follows. In each round, spoiler places some of the pebbles from his pocket on the nodes of \mathbb{A} and duplicator replies by placing his corresponding pebbles on the nodes of \mathbb{B} . If the partial mapping defined by the pebble placement is not a partial homomorphism from \mathbb{A} to \mathbb{B} , then duplicator loses. Otherwise, spoiler puts back some of the pebbles into his pocket, and duplicator removes the corresponding pebbles, and the game continues to the next round. It is not hard to see that \mathbb{A} is $(2, 3)$ -consistent with respect to \mathbb{B} if and only if duplicator can avoid losing forever.

4.2 Systems of linear equations over \mathbb{F}_2

We define a finite template \mathbb{T}_2 that can be used to represent the solvability of systems of linear equations over the 2-element field. Let us note that our definition of the template \mathbb{T}_2 will not be the standard one as it can be found, for example, in the original Feder-Vardi paper [11]. The main difference is that we want to have a signature of smallest possible arity, in this case two. We achieve this by letting \mathbb{T}_2 be the natural encoding of the standard template as its *incidence* structure. Concretely, \mathbb{T}_2 is defined as follows. Its domain is $D \cup R$, where

$$D = \{0, 1\},$$

$$R = \{(x, y, z) \in D^3 : x + y + z = 0 \pmod{2}\}.$$

The elements of D are called values, and those of R are called triples. The signature Σ includes three partial functions π_1 , π_2 , and π_3 that map triples in R to values in D , and four unary relations *value*, *triple*, C_0 and C_1 . Formally, in order to have a relational structure, \mathbb{T}_2 has binary relations that correspond to the graphs of the partial functions π_1 , π_2 and π_3 . The interpretations of the symbols in \mathbb{T}_2 are as follows:

1. π_1 , π_2 and π_3 map (x, y, z) in R to x , y and z , respectively,
2. *value* holds of all elements in D ,
3. *triple* holds of all elements in R ,
4. C_0 holds of 0 in D , and
5. C_1 holds of 1 in D .

The purpose of *triple* is to encode equations of the type $x + y + z = 0 \pmod{2}$, and the purposes of C_0 and C_1 are to encode equations of the type $x = 0$ and $x = 1$, respectively. Note that even though the language does not allow writing more complicated equations, such as $x + y + z = 1 \pmod{2}$ or $w + x + y + z = 0 \pmod{2}$, such equations can be simulated in the language of \mathbb{T}_2 with the help of auxiliary variables.

► **Theorem 4.1.** *The class of all finite structures that are $(2, 3)$ -consistent with respect to \mathbb{T}_2 is not homogenizable.*

Proof. In the following, we fix the template $\mathbb{T} = \mathbb{T}_2$, and when we refer to consistency, we mean $(2, 3)$ -consistency with respect to \mathbb{T} . Finite structures on the signature of \mathbb{T} are called instances. Homomorphisms $f : \mathbb{I} \rightarrow \mathbb{T}$ from an instance \mathbb{I} to \mathbb{T} are called solutions. By \mathcal{C} denote the class of consistent instances. Observe that, as noted earlier, the class of consistent instances is closed under inverse homomorphisms.

The plan is to apply Theorem 3.2 to \mathcal{C} , and for that we need to find a confusing diagram $L : \mathbb{A} \rightarrow \mathbb{L}, R : \mathbb{A} \rightarrow \mathbb{R}$ with arbitrarily large \mathbb{A} .

Let $n \geq 8$ be an exact power of two. Let t be a rooted, ordered complete binary tree with n leaves at depth $\log_2(n)$; in particular, no node at depth 2 is a leaf, and no node at depth $\log_2(n) - 1$ is a root. Let \mathbb{I} be the instance obtained from t , with elements of two types: *nodes*, which correspond to the nodes of t , and *triples*, which correspond to triples (v, v_0, v_1) , where v is an internal node in t , and v_0 and v_1 are its left and right sons, respectively. Nodes are labeled by the unary predicate *value* and triples are labeled by the unary predicate *triple*. We say that the triple (v, v_0, v_1) is the triple *below* node v , and is *adjacent* to, or *contains* v , v_0 , and v_1 . For each such triple, we declare:

$$\text{father}(v, v_0, v_1) = \pi_1(v, v_0, v_1) = v,$$

$$\text{left}(v, v_0, v_1) = \pi_2(v, v_0, v_1) = v_0,$$

$$\text{right}(v, v_0, v_1) = \pi_3(v, v_0, v_1) = v_1.$$

We call a structure of this kind simply a tree. Since we will work with Σ -structures that are made of trees, for the sake of intuition from now on we use the names *father*, *left*, and *right* in place of π_1 , π_2 , and π_3 . If i is a value in D , then the i -marking of \mathbb{I} is the Σ -structure $m_i(\mathbb{I})$ obtained from \mathbb{I} by marking the root by the predicate C_i . Observe that in any solution $v : m_i(\mathbb{I}) \rightarrow \mathbb{T}$ of $m_i(\mathbb{I})$, the sum of the values of the leaves is equal to i modulo 2. Conversely, any mapping from the leaves of \mathbb{I} to \mathbb{T} such that the sum of the values of the leaves is equal to i modulo 2 extends uniquely to a solution $v : m_i(\mathbb{I}) \rightarrow \mathbb{T}$.

The structures \mathbb{L} and \mathbb{R} are the markings $m_0(\mathbb{I})$ and $m_1(\mathbb{I})$ of the tree \mathbb{I} , respectively. The structure \mathbb{A} is the substructure of \mathbb{I} induced by the leaves of the tree. Note that \mathbb{A} consists of n isolated points, labeled by the unary relation *value*. The unary relations *triple*, C_0 and C_1 , as well as the binary relations π_1 , π_2 , and π_3 , are empty in \mathbb{A} . Note that \mathbb{L} and \mathbb{R} share \mathbb{A} as an induced substructure. Let $L : \mathbb{A} \rightarrow \mathbb{L}$ and $R : \mathbb{A} \rightarrow \mathbb{R}$ be the corresponding embeddings.

► **Lemma 4.2.** *The free amalgam of \mathbb{L} and \mathbb{R} through L and R is inconsistent.*

Proof. When spoiler has only two pebbles on the board, we allow him to perform a move we call a *slide*, in which he moves one pebble from a node v to a triple adjacent to it, or from a triple to a node belonging to this triple. Duplicator has to respond accordingly: if spoiler slides a pebble from a node v to a triple t containing v on the i -th coordinate, then duplicator must move his corresponding pebble from a value v in D to a triple in R containing v on the i -th coordinate. Symmetrically, in the case when spoiler slides his pebble from a triple to the node in the i -th coordinate, duplicator must move his corresponding pebble from the corresponding triple to the value on its i -th coordinate. The slide moves can be simulated in the original game, using a third pebble.

Denote the two (overlapping) trees \mathbb{I}_L and \mathbb{I}_R , respectively; they have common leaves in the free amalgam $\mathbb{L} \cup_{\mathbb{A}} \mathbb{R}$. Here is the strategy for spoiler; it consists of several steps. In the beginning of the k -th step, spoiler has two pebbles placed on corresponding nodes a and b of \mathbb{I}_L and \mathbb{I}_R , at depth $k - 1$ of the tree. In particular, in the beginning of the first step, two pebbles are placed on the roots of \mathbb{I}_L and \mathbb{I}_R , respectively. For a node v on which spoiler has his pebble, denote by $r(v)$ the value of the corresponding pebble of duplicator. The invariant is that $r(a) \neq r(b)$. This invariant is clearly satisfied in the beginning of the first step, since \mathbb{I}_L has its root labeled with C_0 and \mathbb{I}_R has its root labeled with C_1 .

In the k -th step, spoiler slides his pebble from node a to the triple a' below a in \mathbb{I}_L , and then slides his pebble from node b to the triple b' below b in \mathbb{I}_R . Duplicator's responses have to satisfy $r(\text{father}(a')) = r(a)$ and $r(\text{father}(b')) = r(b)$. In particular,

16:12 Non-Homogenizable Classes of Finite Structures

$r(\text{father}(a')) \neq r(\text{father}(b'))$, by the invariant. It follows that $\text{left}(r(a')) + \text{right}(r(a')) \neq \text{left}(r(b')) + \text{right}(r(b'))$, so either $\text{left}(r(a')) \neq \text{left}(r(b'))$ or $\text{right}(r(a')) \neq \text{right}(r(b'))$ (or both). Since the cases are symmetric, suppose without loss of generality that the first case occurs. Then spoiler slides the pebble from a' to $\text{left}(a')$ and then slides the pebble from b' to $\text{left}(b')$, and continues the game from these two nodes playing the role of a and b . The invariant is satisfied.

Since in each step the depth of a increases by 1, at some point, a must be a leaf of \mathbb{I}_L , and b is the corresponding leaf in \mathbb{I}_R . But then a and b are the same element in $\mathbb{L} \cup_{\mathbb{A}} \mathbb{R}$, and by the invariant $r(a) \neq r(b)$. In other words, spoiler has two pebbles placed at the same node of $\mathbb{L} \cup_{\mathbb{A}} \mathbb{R}$, but the corresponding pebbles of duplicator are not placed on the same element of \mathbb{T} . So duplicator loses. ◀

► **Lemma 4.3.** *Every amalgam of \mathbb{L} and \mathbb{R} through L and R is inconsistent.*

Proof. This follows at once from the previous lemma and the fact that \mathcal{C} is closed under inverse homomorphisms. Indeed, the free amalgam $\mathbb{L} \cup_{\mathbb{A}} \mathbb{R}$ through L and R maps homomorphically to any amalgam of \mathbb{L} and \mathbb{R} through L and R . ◀

Let m be a natural number, and let $\mathbb{J} = \mathbb{A} \otimes m$, and let $\mathcal{E} = \mathcal{E}_{\mathbb{A},m}$.

The proof of the following lemma is more involved, but is not very insightful, and is omitted due to lack of space.

► **Lemma 4.4.** *For every coloring $C : \mathcal{E} \rightarrow \{L, R\}$, the structure \mathbb{J}^C is consistent.*

Lemma 4.3 and Lemma 4.4 show that the diagram $L : \mathbb{A} \rightarrow \mathbb{L}, R : \mathbb{A} \rightarrow \mathbb{R}$ is confusing for the class of consistent structures. Since \mathbb{A} can be taken arbitrarily large, Theorem 4.1 follows immediately from Theorem 3.2. ◀

4.3 Other finite Abelian groups

The template \mathbb{T}_2 for systems of equations over the 2-element field can be generalized to all finite Abelian groups. Let G be a finite Abelian group; we write $+$ for the group operation and 0 for its neutral element. Let \mathbb{T}_G be the structure with domain $D \cup R$, where

$$D = G,$$

$$R = \{(x, y, z) \in D^3 : x + y + z = 0\}.$$

The elements of D are called values, and those of R are called triples. As in \mathbb{T}_2 , the signature of \mathbb{T}_G has three binary relations π_1, π_2 , and π_3 , two unary relations *value* and *triple*, and one unary relation C_a for each value a in D . The interpretations of all relation symbols are as in \mathbb{T} ; in particular, the unary relation symbol C_a is interpreted by the singleton set $\{a\}$. It is straightforward to check that \mathbb{T}_G can be used to encode arbitrary systems of equations over G . As in the 2-element field case, equations more complex than the basic $x + y + z = 0$ or $x = a$ can be encoded with the help of auxiliary variables.

► **Theorem 4.5.** *If G is a finite Abelian group with at least two elements, then the class of all finite structures that are (2,3)-consistent with respect to \mathbb{T}_G is not homogenizable.*

Proof. The proof of Theorem 4.1 does not rely in any way on the fact that the group is addition mod 2, except for it being Abelian and having at least two different values in it. ◀

It is known that, for any non-trivial finite Abelian group, the constraint satisfaction problem of the template \mathbb{T}_G has *unbounded width*, i.e. for every two natural numbers k and l there exist instances \mathbb{I} that do not have homomorphisms to \mathbb{T}_G , but are nonetheless (k, l) -consistent with respect to \mathbb{T}_G . We also say that \mathbb{T}_G does not have (k, l) -width for any k and l . This was proved by Feder and Vardi [11] for the standard template for linear equations mod 2, and later alternative proofs generalize quite well to the template \mathbb{T}_G (see, for instance, [1]). Moreover, the solution to the Bounded-Width Conjecture of Barto and Kozik [4] implies that all cases of templates of unbounded width are explained by the unbounded width of some \mathbb{T}_G . Technically:

► **Theorem 4.6** ([4], see also Theorem 4.1 in [3]). *Let \mathbb{T} be a core finite relational structure. If \mathbb{T} does not have bounded width, then \mathbb{T} pp-interprets \mathbb{T}_G for some non-trivial Abelian group G . Moreover, if the signature of \mathbb{T} has maximum arity at most two, then the conclusion holds even if \mathbb{T} does not have $(2, 3)$ -width.*

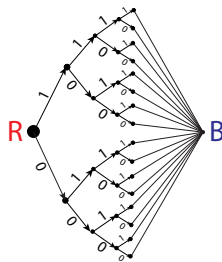
Thus, the templates \mathbb{T}_G are in a strict formal sense the canonical templates of unbounded width. Theorem 4.5 states that, for all such templates, their class of locally consistent instances is non-homogenizable. Interestingly, the converse to this is also true in a strong sense: for all templates that *do have* bounded width, their class of locally consistent instances *is* homogenizable. This follows quite directly from the fact that, for every finite template \mathbb{T} , the class of instances \mathbb{I} that have a homomorphism to \mathbb{T} is homogenized by expanding them by all their homomorphisms to \mathbb{T} . When these two observations are put together, we get that, up to the relation of pp-interpretability between templates, which is known to preserve the property of having bounded width, the templates that have bounded width are distinguished from those that do not by the fact that their classes of locally consistent instances are homogenizable. It seems plausible that our Theorem 4.5 could be adapted to show that *all* templates of unbounded width give themselves a non-homogenizable class of locally-consistent instances, without the need to resort to pp-interpretability, but this remains open.

5 Classes defined by forbidden homomorphisms

The positive result of Hubička and Nešetřil [14] shows that if \mathcal{F} is an HN-regular class of finite connected structures, then the class $\text{Forb}_h(\mathcal{F})$ is a reduct of an amalgamation class. HN-regularity is a notion reminiscent of the notion of regularity of word languages or of tree languages. Indeed, in the case of structures of treewidth one, HN-regularity and regularity in the sense of tree automata both correspond to MSO-definability. In this section we give an example of a non-homogenizable class of finite structures that is of the form $\text{Forb}_h(\mathcal{G})$, where \mathcal{G} is an MSO-definable class of connected finite structures of treewidth two. It follows from the result of Hubička and Nešetřil that this is optimal. Recall that the treewidth of a finite structure is defined as the treewidth of its Gaifman graph, and pathwidth is a restriction of treewidth (for definitions see, for example, [8]).

5.1 Pathwidth three

Consider the class \mathcal{F} from the running example in Section 3, shown not to be a reduct of any amalgamation class in Example 3.3. The class \mathcal{F} can be defined by an MSO sentence, which expresses that there are exactly four colored points, which are colored R , B , S , and T , respectively, and the rest of points form a directed simple \vec{E} -path from S to T with all vertices along the path connected by an undirected E -edge to both R and B . Moreover, each



■ **Figure 5** Forbidden structure \mathbb{G} , with 16 leaves.

structure \mathbb{F} in \mathcal{F} is connected and has pathwidth three: take a standard path-decomposition of the \vec{E} -path with bags of size 2 and add both red and blue vertices to each bag. This gives a path-decomposition with bags of size 4, so its pathwidth is 3 (thanks to the -1 in the definition of treewidth/pathwidth).

Now we show how to modify the class \mathcal{F} to obtain a class of structures of treewidth two.

5.2 Treewidth two

Consider a rooted, directed binary tree, in which every node is either a leaf, or an inner node with two sons, in which case it has a directed \vec{E}_0 -edge to its left son and a directed \vec{E}_1 -edge to its right son. Color its root red, by labeling it with the unary predicate R , and create an extra blue vertex (with unary predicate B), connected to all the leaves of the tree by an undirected E -edge. An example of such a structure, obtained from a full binary tree of depth 4, is depicted in Figure 5. Let \mathcal{G} denote the class of all structures obtained in this way.

The signature Σ of these structures consists of three binary predicates E, \vec{E}_0, \vec{E}_1 , and two unary predicates R and B , each appearing in the structure exactly once as indicated. It is straightforward to check that \mathcal{G} is MSO-definable on the class of all finite structures. Moreover, each structure \mathbb{G} in \mathcal{G} is connected and has treewidth two: just take a tree-decomposition of the binary tree and add the blue point to all its bags.

► **Claim 3.** *The class \mathcal{G} forms an antichain in the homomorphism preorder.*

Proof. Suppose that $h : \mathbb{G}_1 \rightarrow \mathbb{G}_2$ is a homomorphism of two structures in \mathcal{G} . Then h must map the root of \mathbb{G}_1 to the root of \mathbb{G}_2 (since only the root is colored red), and must map the leaves of \mathbb{G}_1 to the leaves of \mathbb{G}_2 (since only the leaves are adjacent to a blue node). Finally, a vertex v in \mathbb{G}_1 reached from the root by a path with labels $i_1 i_2 \dots i_k \in \{0, 1\}^*$ must be mapped to the unique vertex w of \mathbb{G}_2 reached from the root by the path obtained by reading the same labels. Hence, the mapping f is injective. Since no inner node of the tree can be mapped to a leaf, f must also be surjective. It follows that h is an isomorphism. ◀

► **Proposition 5.1.** *The class $\text{Forb}_h(\mathcal{G})$ is not homogenizable.*

Proof. We apply Theorem 3.2. To this end, choose an arbitrary structure $\mathbb{G} \in \mathcal{G}$, and consider the diagram $L : \mathbb{A} \rightarrow \mathbb{L}, R : \mathbb{A} \rightarrow \mathbb{R}$ defined as follows. \mathbb{L} is the left part of the structure \mathbb{G} , obtained by removing the blue vertex (labeled B), \mathbb{R} is the right part of the structure \mathbb{G} , obtained by keeping the blue vertex and the nodes adjacent to it, \mathbb{A} is the intersection of \mathbb{L} and \mathbb{R} , i.e., the substructure of \mathbb{G} induced by the leaves of the underlying binary tree. Let $L : \mathbb{A} \rightarrow \mathbb{L}$ and $R : \mathbb{A} \rightarrow \mathbb{R}$ be the two inclusions. It is clear that every amalgamation of L, R must contain a homomorphic image of \mathbb{G} , so L, R witnesses failure

of amalgamation of $\text{Forb}_h(\mathcal{G})$. Let m be an arbitrary number, $\mathbb{J} = \mathbb{A} \otimes m$, $\mathcal{E} = \mathcal{E}_{\mathbb{A},m}$, and $C : \mathcal{E} \rightarrow \{L, R\}$ be any coloring.

► **Claim 4.** *The structure \mathbb{J}^C does not contain a homomorphic image of any structure in \mathcal{G} .*

The proof uses similar ideas as in the proof of Claim 3, and is omitted.

Hence, the diagram L, R is confusing for \mathcal{C} . Since \mathbb{G} can be chosen so that \mathbb{A} is arbitrarily large, the conclusion follows from Theorem 3.2. ◀

5.3 Optimality

We argued already that the classes \mathcal{F} from Example 3.1 and \mathcal{G} from Section 5.2 are MSO-definable. Therefore, the set of colored paths that represent the path-decompositions of the structures in \mathcal{F} is regular in the automata-theoretic sense, and the set of colored trees that represent the tree-decompositions of the structures in \mathcal{G} is regular in the tree-automata-theoretic sense (see [8]). It is interesting to check why \mathcal{F} and \mathcal{G} are *not* regular classes of structures in the sense of Definition 2.3 of Hubička-Nešetřil [14]. By Example 3.3 and Proposition 5.1 we know that \mathcal{F} and \mathcal{G} cannot be HN-regular as otherwise $\text{Forb}_h(\mathcal{F})$ and $\text{Forb}_h(\mathcal{G})$ would be homogenizable by Theorem 3.1 in [14].

In order to check that a class is not HN-regular it suffices to identify minimal g-separating g-cuts of unbounded sizes in its structures. For \mathcal{F} , note that the set of all vertices in the \vec{E} -path is a minimal g-separating g-cut in \mathbb{F}_k , and its size is k and hence unbounded. For \mathcal{G} , the set of all leaves in the binary tree in any structure \mathbb{G} in \mathcal{G} is a minimal g-separating g-cut, and its size is also unbounded since all trees are represented in \mathcal{G} .

We note that every MSO-definable class of finite connected structures of treewidth one *is* HN-regular. This follows from the fact noted earlier that, for colored trees, HN-regularity, tree-automata regularity, and MSO-definability are equivalent. In particular, by Theorem 3.1 in [14], every class of the form $\text{Forb}_h(\mathcal{F})$, where \mathcal{F} is an MSO-definable class of connected finite structures of treewidth at most one, is homogenizable.

Conclusion

We study homogenizability – a combinatorial notion useful in computer science (see e.g. [5],[7]). Our main contribution is a necessary condition for homogenizability of a class of finite structures. We apply it to prove nonhomogenizability of a class related to constraint satisfaction problems, consisting of locally consistent structures with respect to the template encoding linear equations over a finite abelian group, and an MSO-definable class of structures of treewidth two, which is tight by the positive result of [14].

Acknowledgment. We thank the anonymous reviewers for the useful remarks.

References

- 1 A. Atserias, A. A. Bulatov, and A. Dawar. Affine systems of equations and counting infinitary logic. *Theor. Comput. Sci.*, 410(18):1666–1683, 2009.
- 2 A. Atserias and M. Weyer. Decidable relationships between consistency notions for constraint satisfaction problems. In *Proceedings of the 23rd CSL International Conference and 18th EACSL Annual Conference on Computer Science Logic*, CSL'09/EACSL'09, pages 102–116, Berlin, Heidelberg, 2009. Springer-Verlag.
- 3 L. Barto. The constraint satisfaction problem and universal algebra. *The Bulletin of Symbolic Logic*, 21:319–337, 9 2015.

- 4 L. Barto and M. Kozik. Constraint satisfaction problems solvable by local consistency methods. *J. ACM*, 61(1):3:1–3:19, January 2014.
- 5 M. Bodirsky. *Complexity of Constraints: An Overview of Current Research Themes*, chapter Constraint Satisfaction Problems with Infinite Templates, pages 196–228. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.
- 6 M. Bojańczyk, B. Klin, and S. Lasota. Automata theory in nominal sets. *Logical Methods in Computer Science*, 10(3:4):paper 4, 2014.
- 7 M. Bojańczyk, L. Segoufin, and S. Toruńczyk. Verification of database-driven systems via amalgamation. In Richard Hull and Wenfei Fan, editors, *Proceedings of the 32nd ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, PODS 2013, New York, NY, USA – June 22–27, 2013*, pages 63–74. ACM, 2013. doi:10.1145/2463664.2465228.
- 8 B. Courcelle and J. Engelfriet. *Graph Structure and Monadic Second-Order Logic – A Language-Theoretic Approach*, volume 138 of *Encyclopedia of mathematics and its applications*. Cambridge University Press, 2012.
- 9 J. Covington. Homogenizable relational structures. *Illinois J. Math.*, 34(4):731–743, 12 1990. URL: <http://projecteuclid.org/euclid.ijm/1255988065>.
- 10 P. L. Erdős, C. Tardif, and G. Tardos. On infinite-finite duality pairs of directed graphs. *Order*, 30(3):807–819, 2013.
- 11 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.
- 12 Erich Grädel, Phokion G. Kolaitis, Leonid Libkin, Maarten Marx, Joel Spencer, Moshe Y. Vardi, Yde Venema, and Scott Weinstein. *Finite Model Theory and Its Applications*. Texts in Theoretical Computer Science. An EATCS Series. Springer, 2007.
- 13 W. Hodges. *A Shorter Model Theory*. Cambridge University Press, New York, NY, USA, 1997.
- 14 J. Hubička and J. Nešetřil. Universal structures with forbidden homomorphisms. In *Logic Without Borders: Essays on Set Theory, Model Theory, Philosophical Logic and Philosophy of Mathematics*, pages 241–264. De Gruyter, 2015.
- 15 Ph. G. Kolaitis, H. J. Proemel, and B. L. Rothschild. K_{l+1} -Free Graphs: Asymptotic Structure and a 0–1 Law. *Transactions of the American Mathematical Society*, 303(2):637–671, 1987.
- 16 D. Macpherson. A survey of homogeneous structures. *Discrete Mathematics*, 311(15):1599–1634, 2011. Infinite Graphs: Introductions, Connections, Surveys.
- 17 J. Nešetřil. Ramsey theory. In R. L. Graham, M. Grötschel, and L. Lovász, editors, *Handbook of Combinatorics (Vol. 2)*, pages 1331–1403. MIT Press, Cambridge, MA, USA, 1995.