Extension Preservation on Dense Graph Classes

Ioannis Eleftheriadis
Department of Computer Science and Technology, University of Cambridge, UK

Abstract

Preservation theorems provide a direct correspondence between the syntactic structure of first-order sentences and the closure properties of their respective classes of models. A line of work has explored preservation theorems relativised to combinatorially tame classes of sparse structures [Atserias et al., JACM 2006; Atserias et al., SiCOMP 2008; Dawar, JCSS 2010; Dawar and Eleftheriadis, MFCS 2024]. In this article we initiate the study of preservation theorems for dense classes of graphs. In contrast to the sparse setting, we show that extension preservation fails on most natural dense classes of low complexity. Nonetheless, we isolate a technical condition which is sufficient for extension preservation to hold, providing a dense analogue to a result of [Atserias et al., SiCOMP 2008].

2012 ACM Subject Classification Theory of computation → Finite Model Theory

Keywords and phrases Extension preservation, finite model theory, dense graphs, cliquewidth

Digital Object Identifier 10.4230/LIPIcs.CSL.2025.7

Funding The author is supported by a George and Marie Vergottis Scholarship awarded through Cambridge Trust, an Onassis Foundation Scholarship, and a Robert Sansom Studentship.

1 Introduction

The early days of finite model theory were considerably guided by attempts aiming to relativise theorems and techniques of classical model theory to the finite realm. While many of these were trivially shown to admit no meaningful relativisation, others were extended in a way that broadened their applicability and rendered them extremely useful tools in the study of finite models. Preservation theorems were at the heart of this approach. Most notably, the Łoś-Tarski preservation theorem which asserts that a first-order formula is preserved by extensions between all structures if and only if it is equivalent to an existential formula, was shown to fail in the finite from early on [28, 22]. On the contrary, the homomorphism preservation theorem asserting that a formula is preserved by homomorphisms if and only if it is existential-positive, was open for several years until it was surprisingly shown to extend to finite structures [27], leading to applications in constraint satisfaction problems and database theory.

Still, considering all finite structures allows for combinatorial complexity, giving rise to wildness from a model-theoretic perspective, and intractability from a computational perspective. Indeed, problems which are hard in general become tractable when restring to classes of finite structures which are, broadly-speaking, tame [12]. In the context of preservation theorems, restricting on a subclass weakens both the hypothesis and the conclusion, therefore leading to an entirely new question. A study of preservation properties for such restricted classes of finite structures was initiated in [4] and [3] for homomorphism and extension preservation respectively. This investigation led to the introduction of different notions of wideness, which allow for arguments based on the *locality* of first-order logic. However, as it was recently realised [14], these arguments require slightly restrictive closure assumptions which are not always naturally present. In particular, it was shown that homomorphism preservation holds over any hereditary *quasi-wide* class which is closed under *amalgamation over bottlenecks*.

Quasi-wideness is a Ramsey-theoretic condition which informally says that in any large enough structure in the class one can remove a bounded number of elements, called bottleneck points, so that there remains a large set of pairwise far-away elements. Here, the number of bottleneck points is allowed to dependent on the choice of distance. Hereditary quasi-wide classes were later identified with *nowhere dense* classes [24]. Over the years, a successful program was developed aiming to understand the combinatorial and model-theoretic features of nowhere dense classes, and exploit them for algorithmic purposes [25]. The culmination of this was the seminal result that first-order model checking is fixed-parameter tractable on nowhere dense classes [21].

In recent years, the focus has shifted towards extending this well understood theory to more general, possibly dense, well-behaved classes, which fall out of the classification provided by the sparsity program. In these efforts, the model-theoretic notions of *monadic stability* and *monadic dependence* have played central roles. Monadic stability, initially introduced by Baldwin and Shelah [5] in the context of classification of complete first-order theories, prohibits arbitrarily large definable orders in monadic expansions. In the language of first-order transductions, a class is monadically stable whenever it does not transduce the class of finite linear orders. More generally, a class is monadically dependent if it does not transduce the class of all graphs. In the context of monotone classes of graphs, Adler and Adler [1] first observed that the above notions coincide with nowhere density, a result which was also extended to arbitrary relational structures [7]. The generalisation of sparsity theory to dense classes eventually led to the result that first-order model checking is fixed-parameter tractable on all monadically stable graph classes [15], which in particular include transductions of nowhere dense classes. It is conjectured that the above result extends to all monadically dependent classes, while a converse was recently established for hereditary graph classes (under standard complexity-theoretic assumptions) [18].

The purpose of the present article is to initiate the investigation of preservation theorems on tame dense graph classes. Much like nowhere dense classes are equivalently characterised by quasi-wideness, monadically stable and monadically dependent graph classes also admit analogous wideness-type characterisations. In the case of monadic stability, the relevant condition is known as *flip-flatness* [17]; this may be viewed as a direct analogue of quasi-wideness which replaces the vertex deletion operation by flips, i.e. edge-complementations between subsets of the vertex set. For monadically dependent classes the relevant condition, known as *flip-breakability* [18], allows to find two large sets such that elements in one are far away from elements in the other, again after performing a bounded number of flips. However, unlike quasi-wideness which was introduced in the context of preservation and then shown to coincide with nowhere density, these conditions were introduced purely for the purpose of providing combinatorial characterisations of monadic stability and monadic dependence respectively. The immediate question thus becomes whether these conditions, or variants thereof, can be used to obtain preservation in restricted tame dense classes, in analogy to the use of wideness in [3, 4, 13, 14].

The first observation is that the arguments for homomorphism preservation are not directly adaptable in this context due to the nature of flips. Indeed, while the vertex-deletion operation respects the existence of a homomorphism between two structures, the flip operation is not at all rigid with respect to homomorphisms precisely because the latter do not reflect relations, e.g. the graph $K_1 + K_1$ homomorphically maps to K_2, but $\overline{(K_1 + K_1)} = K_2$ does not map to $\overline{K_2} = K_1 + K_1$. This issue evidently disappears if one considers embeddings. As it was observed in [14, Corollary 2.3], the extension preservation property implies the homomorphism preservation property in hereditary classes of finite structures, so considering extension preservation is more general for our purposes.

However, this generality comes at a cost. Indeed, the argument for extension preservation from [3] requires that the number of vertex-deletions is independent of the choice of radius, a condition known as *almost-wideness*. This is a more restrictive assumption which therefore applies to fewer sparse classes. It is not known whether extension preservation is obtainable for quasi-wide classes. At the same time, unlike [14, Theorem 4.2] whose proof is essentially a direct application of Gaifman's locality theorem based on an argument of Ajtai and Gurevich [2], the proof of extension preservation [3, Theorem 4.3] is admittedly much more cumbersome. One explanation for this is that the homomorphism preservation argument relies on the fact that the disjoint union operation endows the category of graphs and homomorphisms with *coproducts*, i.e. for any graphs A, B, C if there are homomorphisms $f : A \to C$ and $g : B \to C$ then there is a homomorphism $f + g : A + B \to C$ whose pre-compositions with the respective inclusion homomorphisms $\iota_A : A \to A + B$ and $\iota_B : B \to A+B$ are equal to f and g respectively. On the other hand, no construction satisfies the above property in the category of graphs with embeddings; in fact coproducts do not even exist in the category of graphs with strong homomorphisms (see [23, Corollary 4.3.15]).

Our first contribution is negative, showing that extension preservation can fail on tame dense classes of low complexity. In particular, we show that extension preservation fails on the class of all graphs of (linear) cliquewidth at most k, for all $k \geq 4$. This answers negatively a question of [14]. This is contrary to the sparse picture, where it was shown that extension preservation holds in the class of graphs of treewidth at most k, for every $k \in \mathbb{N}$ [3, Theorem 5.2]. Interestingly, extension preservation holds for the class of all graphs of cliquewidth 2 as this class coincides with the class of cographs which is known to be well-quasi-ordered [11]. Our construction is based on the encoding of linear orders via the neighbourhoods of certain vertices. Orders are also central to the original counterexample for the failure of extension preservation in the finite due to Tait [28]. There, the orders are crucially presented over a signature with two relation symbols and one constant, which does not allow for a direct translation to undirected graphs. Sadly, the fact that orders appear to provide counterexamples rules out the possibility of using an argument based on flip-breakability to establish preservation.

The second contribution of the article is positive. In particular, we provide a dense analogue to [3, Theorem 4.3]. For this, we introduce *strongly flip-flat* classes, i.e. those flip-flat classes such that the number of flips is independent of the choice of radius. Moreover, we formulate the dense analogue of the amalgamation construction, which we call the *flip-sum*, whose existence in the class is necessary for the argument to be carried out. The main theorem (Theorem 18 below) may thus be formulated as saying that extension preservation holds over any hereditary strongly flip-flat class which is closed under flip-sums over bottleneck partitions.

2 Preliminaries

We assume familiarity with the standard notions of finite model theory and structural graph theory, and refer to [19] and [25] for reference. In this article, graphs shall always refer to simple undirected graphs i.e. structures over the signature $\tau_E = \{E\}$ where E is interpreted as a symmetric and anti-reflexive binary relation. For a graph G we write $V(G)$ for its domain (or vertex set), and $E(G)$ for its edge set. In general, for a τ-structure A and a relation symbol $R \in \tau$ of arity $r \in \mathbb{N}$ we write $R^A \subseteq A^r$ for the interpretation of R in A. We shall abuse notation and not distinguish between structures and their respective domains.

Given two structures A, B in the same relational signature τ, a homomorphism $f : A \to B$ is a map that preserves all relations, i.e. for all $R \in \tau$ and tuples $\bar{a}$ from A we have $\bar{a} \in R^A \implies f(\bar{a}) \in R^B$. A *strong* homomorphism is a homomorphism $f : A \to B$ that additionally reflects all relations, i.e. $f(\bar{a}) \in R^B \implies \bar{a} \in R^A$. An injective strong homomorphism is called an *embedding* or *extension*.

A τ-structure B is said to be a *weak substructure* of a τ-structure A if $B \subseteq A$ and the inclusion map $\iota : B \hookrightarrow A$ is a homomorphism. Likewise, B is an *induced substructure* of A if the inclusion map is an embedding; we write $B \leq A$ for this. Given a structure A and a subset $S \subseteq A$ we write $A[S]$ for the unique induced substructure of A with domain S. An induced substructure B of A is said to be *proper* if $B \subsetneq A$; we write $B \lneq A$ for this. We say that a class of structures in the same signature is *hereditary* if it is closed under induced substructures. Moreover a class is called *addable* if it is closed under taking disjoint unions, which we denote by $A + B$.

By the *Gaifman graph* of a structure A we mean the undirected graph $\mathsf{Gaif}(A)$ with vertex set A such that two elements are adjacent if, and only if, they appear together in some tuple of a relation of A. Given a structure A, $r \in \mathbb{N}$, and $a \in A$, we write $N_r^A(a)$ for the *r-neighbourhood of a in A*, that is, the set of elements of A whose distance from a in $\mathsf{Gaif}(A)$ is at most r. We shall often abuse notation and write $N_r^A(a)$ for the induced substructure $A[N_r^A(a)]$ of A. For a set $C \subseteq A$ we define $N_r^A(C) := \bigcup_{a \in C} N_r^A(a)$. A set $S \subseteq A$ is said to be *r-independent* if $b \notin N_r^A(a)$ for any $a, b \in S$.

For $r \in \mathbb{N}$, let $\text{dist}(x, y) \leq r$ be the first-order formula expressing that the distance between x and y in the Gaifman graph is at most r, and $\text{dist}(x, y) > r$ its negation. Clearly, the quantifier rank of $\text{dist}(x, y) \leq r$ is at most r. A *basic local sentence* is a sentence

$$\exists x_1, \ldots, x_n \Big(\bigwedge_{i \neq j} \text{dist}(x_i, x_j) > 2r \wedge \bigwedge_{i \in [n]} \psi^{N_r(x_i)}(x_i) \Big),$$

where $\psi^{N_r(x_i)}(x_i)$ denotes the relativisation of ψ to the r-neighbourhood of x_i, i.e. the formula obtained from ψ by replacing every quantifier $\exists x\, \theta$ with $\exists x(\text{dist}(x_i, x) \leq r \wedge \theta)$, and likewise every quantifier $\forall x\, \theta$ with $\forall x(\text{dist}(x_i, x) \leq r \to \theta)$. We call r the *locality radius*, n the *width*, and ψ the *local condition* of ϕ. Recall the Gaifman locality theorem [19, Theorem 2.5.1].

▶ **Theorem 1** (Gaifman Locality). *Every first-order sentence of quantifier rank q is equivalent to a Boolean combination of basic local sentences of locality radius 7^q.*

A class $\mathcal{C}$ of structures is said to be *quasi-wide* if for every $r \in \mathbb{N}$ there exist $k_r \in \mathbb{N}$ and $f_r : \mathbb{N} \to \mathbb{N}$ such that for all $m \in \mathbb{N}$ and all $A \in \mathcal{C}$ of size at least $f_r(m)$ there exists $S \subseteq A$ such that $A \setminus S$ contains an r-independent set of size m. Moreover, if $k_r := k \in \mathbb{N}$ is independent of r, then $\mathcal{C}$ is said to be *almost-wide*. Finally, we say that a class $\mathcal{C}$ is *uniformly* quasi-wide (uniformly almost-wide respectively) if the hereditary closure of $\mathcal{C}$ is quasi-wide (almost-wide respectively).

For a graph G and a pair of disjoint vertex subsets U and V, the subgraph *semi-induced* by U and V is the bipartite graph with sides U and V that contains all edges of G with one endpoint in U and second in V. By the *half-graph of order n* we mean the bipartite graph with vertices $\{u_i, v_i : i \in [n]\}$ and edges $\{(u_i, v_j) : i \leq j\}$.

For first-order formulas $\delta(x)$ and $\phi(x, y)$ the interpretation $I_{\delta,\phi}$ is defined to be the operation that maps a graph G to the graph $H := I_{\delta,\phi}(G)$ with vertex set $V(H) := \{v \in V(G) \colon G \models \delta(v)\}$ and edge set

$$E(H) := \left\{ (u, v) \in V(H)^2 \colon u \neq v \wedge G \models (\phi(u, v) \vee \phi(v, u)) \right\}.$$

For a graph class $\mathcal{C}$, we write $I_{\delta,\phi}(\mathcal{C}) := \{I_{\delta,\phi}(G) : G \in \mathcal{C}\}$. We say that a class $\mathcal{C}$ *is an interpretation of a class* $\mathcal{D}$, or that $\mathcal{D}$ *interprets* $\mathcal{C}$, if there is some $I_{\delta,\phi}$ such that $\mathcal{C} \subseteq I_{\delta,\phi}(\mathcal{D})$. We say that $\mathcal{C}$ *is a transduction of a class* $\mathcal{D}$, or $\mathcal{D}$ *transduces* $\mathcal{C}$, if there are $k \in \mathbb{N}$ and unary predicates $P_1, \dots, P_k$ and formulas $\delta(x)$ and $\phi(x, y)$ over the signature $\tau_E \cup \{P_1, \dots, P_k\}$ such that $\mathcal{C} \subseteq I_{\delta,\phi}(\mathcal{D}^k)$, where $\mathcal{D}^k$ is the class of all $\tau_E \cup \{P_1, \dots, P_k\}$-structures whose τ_E-reducts are in $\mathcal{D}$. A graph class $\mathcal{C}$ is *monadically dependent* if $\mathcal{C}$ does not transduce the class of all graphs. $\mathcal{C}$ is moreover *monadically stable* if $\mathcal{C}$ does not transduce the class of all half-graphs.

We say that a formula ϕ is preserved by extensions over a class of structures $\mathcal{C}$ if for all $A, B \in \mathcal{C}$ such that there is a embedding from B to A, $B \models \phi$ implies that $A \models \phi$. We say that a class of structures $\mathcal{C}$ has the *extension preservation property* if for every formula ϕ preserved by extensions over $\mathcal{C}$ there is an existential formula ψ such that $M \models \phi \iff M \models \psi$ for all $M \in \mathcal{C}$. We analogously define the *homomorphism preservation property*, replacing "embeddings" with "homomorphisms" and "existential" with "existential positive" in the above.

Given a formula ϕ and a class of structures $\mathcal{C}$, we say that $M \in \mathcal{C}$ is a *minimal induced model* of ϕ in $\mathcal{C}$ if $M \models \phi$ and for any proper induced substructure N of M with $N \in \mathcal{C}$ we have $N \not\models \phi$. The relationship between minimal models and extensions preservation is highlighted by the following folklore lemma. We provide a proof for completeness.

▶ **Lemma 2.** *Let $\mathcal{C}$ be a hereditary class of finite structures. Then a sentence preserved by extensions in $\mathcal{C}$ is equivalent to an existential sentence over $\mathcal{C}$ if and only if it has finitely many minimal induced models in $\mathcal{C}$.*

Proof. Suppose that ϕ has finitely many minimal induced models in $\mathcal{C}$, say $M_1, \dots, M_n$. For each $i \in [n]$, let ψ_i be the primitive sentence inducing a copy of M_i and write $\psi := \bigvee_{i\in[n]} \psi_i$; evidently ψ is existential. We argue that ϕ is equivalent to ψ over $\mathcal{C}$. Indeed, if $A \in \mathcal{C}$ models ϕ then A contains a minimal induced model B of ϕ as an induced substructure. By hereditariness $B \in \mathcal{C}$ and so B is isomorphic to some M_i. Since there is clearly an embedding $B \to A$ it follows that $A \models \psi$. On the other hand if $A \models \psi$, then $A \models \psi_i$ for some $i \in [n]$ and so some M_i embeds into A. Since $M_i \models \phi$ and ϕ is preserved by extensions this implies that $A \models \phi$ as required.

Conversely, assume that ϕ is equivalent to an existential sentence over $\mathcal{C}$. In particular, ϕ is equivalent to some disjunction $\bigvee_{i\in[n]} \psi_i$ where each ψ_i is primitive. It follows that each ψ_i is the formula inducing one of finitely many structures $M^i_1, \dots, M^i_{k_i}$. Now, if A is a minimal induced model of ϕ then in particular $A \models \psi_i$ for some $i \in [n]$, i.e. there is some $j \in [k_i]$ and an embedding $h : M^i_j \to A$. If h is not surjective, then $A[h[M^i_j]]$ is a proper induced substructure of A, which is in $\mathcal{C}$ by hereditariness, and models ϕ; this contradicts the minimality of A. Hence, the size of every minimal induced model of ϕ in $\mathcal{C}$ is bounded by $\max_{i\in[n]} \max_{j\in[k_i]} |M^i_j|$. It follows that ϕ can have only finitely many minimal induced models in $\mathcal{C}$. ◀

3 Failure of preservation on graphs of cliquewidth 4

One consequence of Lemma 2 is that extension preservation holds over any class $\mathcal{C}$ that is *well-quasi-ordered* by the induced substructure relation, i.e. classes for which there exists no infinite collection of members which pairwise do not embed into one another. In particular, this applies to the class of *cographs* [11], which are precisely the graphs of cliquewidth 2 (see [9] for background on cliquewidth). Hence, one may reasonably inquire whether extension preservation is generally true for the class $\mathcal{CW}_k$ of all graphs of cliquewidth at most k. This would in particular reflect an analogous phenomenon that is true in the sparse setting, that is, that extension preservation holds over the class $\mathcal{TW}_k$ of all graphs of treewidth at most k [3, Theorem 5.2].

Classes of bounded cliquewidth are not monadically stable, as even the class of cographs contains arbitrarily large semi-induced half-graphs, but they are monadically dependent. In fact, their structural properties imply tame behaviour going much beyond the context of first-order logic (see [10] for a survey). Still, as it turns out, extension preservation fails even at the level of cliquewidth 4. To show this, we produce a formula ϕ preserved by extensions over the class of all finite undirected graphs, which admits infinitely many minimal models of cliquewidth 4. In particular, Lemma 2 implies that extension preservation fails on any class that includes these minimal models. Our idea is based on encoding two interweaving linear orders on the two parts of a semi-induced graph. Two vertices on the same part are comparable in this ordering whenever their neighbourhoods in the other part are set-wise comparable. This effectively forces a semi-induced half-graph.

Our formula is in the form of an implication, proceeded by a primitive part which induces a gadget corresponding to the beginning and end of the two linear orders. The antecedent of the implication first makes sure that the above relation is a pre-ordering on each side of the semi-induced graph, while it imposes that the vertices of the gadget corresponding to the minimal and maximal elements are indeed minimal and maximal in this pre-ordering. Moreover, it essentially ensures that successors, i.e. vertices of the same part whose neighbourhoods over the other part differ by a single element, are adjacent on one part and non-adjacent on the other. The consequent then imposes that any vertex on the first side has an adjacent successor, while every vertex on the other side has a non-adjacent successor. Because each one of the two pre-orders precisely compares neighbourhoods over the other part, this forces the pre-orders to be anti-symmetric, and thus the two parts to have the same number of vertices.

Finally, two additional vertices are also added on one side of the semi-induced bipartite graph, which are part of the gadget and serve no role in this ordering. These make sure that our intended minimal models form an anti-chain in the embedding relation, as they crucially result in the existence of a unique embedding of the gadget into the models (Lemma 5 below).

We now turn to formal definitions. Let $I(v_1, v_2, v_3, v_4, v_5, v_6, u_1, u_2, u_3, u_4, u_5, u_6, a, b)$ be the formula that induces the graph of Figure 1 below. In the following, we treat the free variables of I as constants for simplicity. The notation $\forall(x \in U)$ will denote the relativisation of the universal quantifier to the neighbours of v_1 that are not v_2, i.e. $\forall(x \in U)\ \psi(x)$ is shorthand for $\forall x(E(x, v_1) \wedge x \neq v_2 \rightarrow \psi(x))$. Likewise, the notation $\forall(x \in V)$ denotes the relativisation of the universal quantifier to the non-neighbours of v_1 that are not a or b, i.e. $\forall(x \in V)\ \psi(x)$ is shorthand for $\forall x(\neg E(x, v_1) \wedge x \neq a \wedge x \neq b \rightarrow \psi(x))$. Existential quantifiers relativised to U and V are defined analogously. Consider the auxiliary formulas:

$$x \leq_V y := \forall(z \in U)[E(z, x) \rightarrow E(z, y)];$$

$$x <_V y := x \leq_V y \wedge \neg(y \leq_V x);$$

$$\chi_1 := \forall(x \in V)\forall(y \in V)[x \leq_V y \vee y \leq_V x];$$

$$\chi_2 := \forall(x \in U)[E(x, v_6) \rightarrow x = u_6];$$

$$\chi_3 := \forall(x \in V)\forall(y \in V)[x <_V y \wedge E(x, y) \rightarrow \exists!(z \in U)(E(y, z) \wedge \neg E(x, z))].$$

In analogy, we define:

$$x \leq_U y := \forall(z \in V)[E(z, x) \rightarrow E(z, y)];$$

$x <_U y := x \leq_U y \wedge \neg(y \leq_U x)$;

$\xi_1 := \forall(x \in U)\forall(y \in U)[x \leq_U y \vee y \leq_U x]$;

$\xi_2 := \forall(x \in V)[E(x, u_1) \to x = v_1]$;

$\xi_{2^*} := \forall(x \in V)\ E(x, u_6)$;

$\xi_3 := \forall(x \in U)\forall(y \in U)[x <_U y \wedge \neg E(x, y) \to \exists!(z \in V)(E(y, z) \wedge \neg E(x, z))]$.

We then define:

$\phi_1 := \chi_1 \wedge \chi_2 \wedge \chi_3$;

$\phi_2 := \forall(x \in V)[x \neq v_1 \to \exists(y \in V)(E(x, y)\ \wedge\ x <_V y)]$;

$\psi_1 := \xi_1 \wedge \xi_2 \wedge \xi_{2^*} \wedge \xi_3$;

$\psi_2 := \forall(x \in U)[x \neq u_6 \to \exists(y \in U)(\neg E(x, y) \wedge x <_U y)]$.

Putting the above together we finally define:

$$\phi := \exists \bar{v}, \bar{u}, a, b(I(\bar{v}, \bar{u}, a, b) \wedge [\phi_1(\bar{v}, \bar{u}, a, b) \wedge \psi_1(\bar{v}, \bar{u}, a, b) \to \phi_2(\bar{v}, \bar{u}, a, b) \wedge \psi_2(\bar{v}, \bar{u}, a, b)])$$

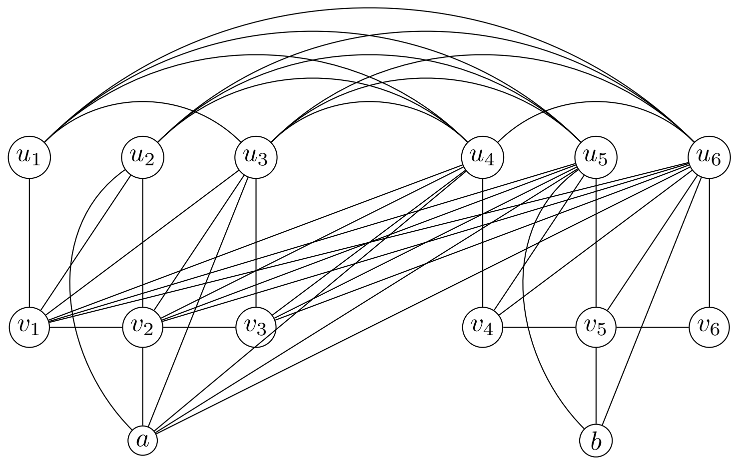

Figure 1 The gadget induced by $I(\bar{v}, \bar{u}, a, b)$.

▶ **Proposition 3.** *The formula ϕ is preserved by extensions over the class of all finite graphs.*

Proof. Let G, H be two graphs such that G embeds into H, and $G \models \phi$. Without loss of generality we assume that $V(G) \subseteq V(H)$ and that the identity map is an embedding. We shall argue that $H \models \phi$.

Since $G \models \phi$, we may fix (tuples of) vertices $\bar{v}, \bar{u}, a, b \in V(G)$ such that $G \models I(\bar{v}, \bar{u}, a, b)$. Evidently, H also models $I(\bar{v}, \bar{u}, a, b)$. If $H \not\models (\phi_1(\bar{v}, \bar{u}, a, b) \wedge \psi_1(\bar{v}, \bar{u}, a, b))$ then $H \models \phi$; we may therefore assume that $H \models (\phi_1(\bar{v}, \bar{u}, a, b) \wedge \psi_1(\bar{v}, \bar{u}, a, b))$. Let $U := N_H(v_1) \setminus \{v_2\} \subseteq V(H)$ be the neighbours of v_1 in H that are not v_2, and $V := V(H) \setminus (U \cup \{a, b\}) \subseteq V(H)$ be the non-neighbours of v_1 that are not a or b (in particular $v_1 \in V$). We call the vertices $x \in V$ *V-elements*. For each V-element x we write $U_x := \{y \in U : H \models E(x, y)\}$ for the *U-neighbourhood of x*. We similarly define *U-elements* and *V-neighbourhoods*. From each of the conjuncts χ_i of ϕ_1 we deduce that in H:

χ_1: V-elements have pairwise comparable U-neighbourhoods;
χ_2: the only member of U_{v_6} is u_6;
χ_3: if two adjacent V-elements x, y satisfy $U_x \subsetneq U_y$ then $|U_y| = |U_x| + 1$.

We shall argue that items (1)-(3) are still true within G, replacing V with $V' := V \cap V(G)$ and U with $U' := U \cap V(G)$, and so $G \models \phi_1(\bar{v}, \bar{u}, a, b)$. We write $U'_x := U_x \cap V(G)$ for the relativised U'-neighbourhoods. Clearly, items (1) and (2) are still true in G. For item (3), suppose that $x, y \in V'$ are two adjacent V'-elements, such that $V'_x \subsetneq V'_y$. Since V-elements in H have pairwise comparable U-neighbourhoods, we deduce that $V_x \subsetneq V_y$ and therefore that $|V_y| = |V_x| + 1$ as H satisfies χ_3. In particular, it follows that $|V'_y| = |V'_x| + 1$ as required, and so G models $\chi_3(\bar{v}, \bar{u}, a, b)$ and consequently $\phi_1(\bar{v}, \bar{u}, a, b)$.

Similarly, from each of the conjuncts ξ_i of ψ_1 we deduce that in H:
ξ_1: U-elements have pairwise comparable V-neighbourhoods;
ξ_2: the only element of V_{u_1} is v_1;
ξ_{2^*}: V_{u_6} is equal to V;
ξ_3: if two non-adjacent U-elements satisfy $V_x \subsetneq V_y$ then $|V_y| = |V_x| + 1$.

Arguing as before, we obtain that $G \models \psi_1(\bar{v}, \bar{u}, a, b)$. Since $G \models \phi$ and $G \models (\phi_1(\bar{v}, \bar{u}, a, b) \wedge \psi_1(\bar{v}, \bar{u}, a, b))$ we deduce that $G \models (\phi_2(\bar{v}, \bar{u}, a, b) \wedge \psi_2(\bar{v}, \bar{u}, a, b))$, i.e. the following are true in G:
ϕ_2: every V'-element that is not v_1 is adjacent to a V'-element of strictly greater U'-neighbourhood;
ψ_2: every U'-element that is not u_6 is non-adjacent to a U'-element of strictly greater V'-neighbourhood.

We proceed to show that the above implies that $V = V'$ and $U' = U$, and hence $G = H$. In particular, this implies that $H \models \phi$ as claimed.

Since G is finite and satisfies ϕ_2 we obtain some $n \in \mathbb{N}$ and a sequence of distinct elements $\alpha_1 := v_6, \alpha_2, \ldots, \alpha_n := v_1$ of V' such that $U'_{\alpha_i} \subsetneq U'_{\alpha_{i+1}}$ and $G \models E(\alpha_i, \alpha_{i+1})$ for all $i \in [n-1]$. In particular, $U_{\alpha_i} \subsetneq U_{\alpha_{i+1}}$ and $H \models E(\alpha_i, \alpha_{i+1})$ for all $i \in [n]$. As H satisfies χ_3 we obtain that $|U_{\alpha_{i+1}}| = |U_{\alpha_i}| + 1$ for all i. Moreover, since H satisfies χ_2 and every element of U is adjacent to v_1, we obtain that $U_{\alpha_1} = \{u_6\}$ and $U_{\alpha_n} = U$. In particular, we deduce that $n = |U| \leq |V'|$. Symmetrically, we obtain some $k \in \mathbb{N}$ and a sequence of elements $\beta_1 := u_1, \beta_2, \ldots, \beta_k := u_6$ of U' such that $V'_{\beta_i} \subsetneq V'_{\beta_{i+1}}$ and $G \models \neg E(\beta_i, \beta_{i+1})$ for all $i \in [k-1]$. Hence, $V_{\beta_i} \subsetneq V_{\beta_{i+1}}$ and $H \models \neg E(\beta_i, \beta_{i+1})$ for all $i \in [k-1]$. Once again, since H satisfies ξ_2, ξ_{2^*}, and ξ_3 we obtain that $V_{\beta_1} = \{v_1\}$, $V_{\beta_n} = V$, and $|V_{\beta_{i+1}}| = |V_{\beta_i}| + 1$. It thus follows that $k = |V| \leq |U'|$. Putting the above together we have that

$$|U| \leq |V'| \leq |V| \leq |U'| \leq |U|.$$

Consequently, $n = k$ while $V = V' = \{\alpha_1, \ldots, \alpha_n\}$ and $U = U' = \{\beta_1, \ldots, \beta_n\}$ as needed. ◀

We now define the intended minimal induced models of our formula ϕ.

▶ **Definition 4.** *For $n \geq 7$ we define the graph $\mathcal{H}_n$ with vertex and edge set*

$$V(\mathcal{H}_n) := \{v_1, \ldots, v_n\} \cup \{u_1, \ldots, u_n\} \cup \{a\} \cup \{b\};$$

$$E(\mathcal{H}_n) := \{(v_i, u_j) : i \leq j\} \cup \{(v_i, v_j) : j = i + 1\} \cup \{(u_i, u_j) : j \neq i + 1\}$$
$$\cup\ \{(a, u_i) : i \geq 2\} \cup \{(b, u_i) : i \geq n - 1\} \cup \{(a, v_2), (b, v_{n-1})\},$$

respectively. We also write $\mathcal{I}_n$ for the subgraph of $\mathcal{H}_n$ induced on the set

$$V(\mathcal{I}_n) := \{v_1, v_2, v_3, v_{n-2}, v_{n-1}, v_n, u_1, u_2, u_3, u_{n-2}, u_{n-1}, u_n, a, b\} \subseteq V(\mathcal{H}_n).$$

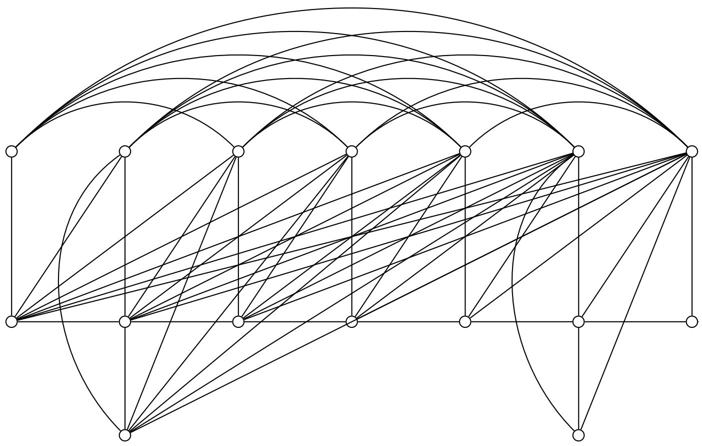

Figure 2 The graph $\mathcal{H}_7$.

We aim to establish that the graphs $\mathcal{H}_n$ are all minimal induced models of ϕ. Towards this, we first argue that the only embedding of $\mathcal{I}_n$ in $\mathcal{H}_n$ is the inclusion map. While this lemma is not conceptually difficult, it requires analysing and ruling out different cases corresponding to potential images of the gadget. Its proof may be found in Section A.

▶ **Lemma 5.** *Let $n \geq 7$ and $f : \mathcal{I}_n \to \mathcal{H}_n$ be an embedding. Then f is the inclusion map.*

▶ **Proposition 6.** *For each $n \geq 7$ the graphs $\mathcal{H}_n$ are minimal induced models of ϕ.*

Proof. We fix some $n \geq 7$. We first argue that $\mathcal{H}_n \models \phi$ for every $n \geq 7$. Indeed, we clearly have that

$$\mathcal{H}_n \models I(v_1, v_2, v_3, v_{n-2}, v_{n-1}, v_n, u_1, u_2, u_3, u_{n-2}, u_{n-1}, u_n, a, b).$$

Moreover, the set $U := \{u_1, \dots, u_n\} \subseteq V(\mathcal{H}_n)$ is precisely the set of neighbours of v_1 which are not v_2, while the set $V := \{v_1, \dots, v_n\} \subseteq V(\mathcal{H}_n)$ is precisely the set of non-neighbours of v_1 which are not a or b. Evidently, we then have that for every vertex $v_i \in V \setminus \{v_1\}$ the vertex $v_{i-1} \in V$ is adjacent to v_i and its neighbourhood over U strictly contains that of v_i. Consequently $\mathcal{H}_n \models \phi_2(\bar{v}, \bar{u}, a, b)$. Likewise, for every vertex $u_i \in U \setminus \{u_n\}$ the vertex $u_{i+1} \in U$ is non-adjacent to u_i and its neighbourhood over V strictly contains that of u_i. It follows that $\mathcal{H}_n \models \psi_2(\bar{v}, \bar{u}, a, b)$, and so $\mathcal{H}_n \models \phi$ as required.

Now, suppose that H is a proper induced subgraph of $\mathcal{H}_n$, and assume for a contradiction that $H \models \phi$, i.e. there are vertices $x_1, \dots, x_6, y_1, \dots, y_6, \alpha, \beta$ of H

$$H \models (I(\bar{x}, \bar{y}, \alpha, \beta) \wedge [\phi_1(\bar{x}, \bar{y}, \alpha, \beta) \wedge \psi_1(\bar{x}, \bar{y}, \alpha, \beta) \to \phi_2(\bar{x}, \bar{y}, \alpha, \beta) \wedge \psi_2(\bar{x}, \bar{y}, \alpha, \beta)]).$$

Since these vertices induce a copy of $\mathcal{I}_n$, it follows by Lemma 5 that

$$(x_1, x_2, x_3, x_4, x_5, x_6, y_1, y_2, y_3, y_4, y_5, y_6, \alpha, \beta) = \\ (v_1, v_2, v_3, v_{n-2}, v_{n-1}, v_n, u_1, u_2, u_3, u_{n-2}, u_{n-1}, u_n, a, b),$$

and so $\mathcal{I}_n \leq H \lneq \mathcal{H}_n$. Moreover, letting $U' := U \cap V(H)$ and $V' := V \cap V(H)$ we see that

- the elements in V' have pairwise comparable neighbourhoods over U', and the elements of U' have pairwise comparable neighbourhoods over V';
- the only neighbour of v_n in U' is u_n, and the only neighbour of u_1 in V' is v_1;

- u_n is adjacent to every element of V;
- if $x, y \in V'$ are adjacent and the U'-neighbours of y are strictly more than the U'-neighbours of x then there is some $i \in [n-1]$ such that $y = v_i$ and $x = v_{i+1}$, and there is a unique vertex in U' that is adjacent to y and not adjacent to x, namely u_i;
- if $x, y \in U'$ are adjacent and the V'-neighbours of y are strictly more than the V'-neighbours of x then there is some $i \in [n-1]$ such that $y = v_{i+1}$ and $y = v_i$, and there is a unique vertex in V' that is adjacent to y and not adjacent to x, namely v_i.

It follows that $H \models (\phi_1(\bar{v}, \bar{u}, a, b) \wedge \psi_1(\bar{v}, \bar{u}, a, b))$. Since $H \models \phi$ this implies that $H \models (\phi_2(\bar{v}, \bar{u}, a, b) \wedge \psi_2(\bar{v}, \bar{u}, a, b))$. However, since $V(H) \subsetneq V(\mathcal{H}_n)$, there is some $i \in [4, n-3]$ such that $v_i \notin V(H)$ or $u_i \notin V(H)$. Assume the former, and let $i \in [4, n-3]$ be maximal such that $v_i \notin V(H)$. It follows that there is no vertex in $x \in V'$ that is adjacent to v_{i+1} and its neighbourhood over U' is strictly greater than that of v_{i+1}, contradicting that $H \models \psi_1(\bar{v}, \bar{u}, a, b)$. By a symmetric argument we obtain a contradiction if $u_i \notin V(H)$, and thus follows that $H \not\models \phi$. ◀

▶ **Theorem 7.** *Extension preservation fails on any hereditary graph class containing the graphs $\mathcal{H}_n$ for arbitrarily large $n \in \mathbb{N}$.*

Proof. Let $\mathcal{C}$ be a class of graphs containing the graphs $\mathcal{H}_n$ for arbitrarily large n. Since the formula ϕ is preserved under extensions over the class of all finite graphs, it is in particular preserved under extensions over $\mathcal{C}$. Since $\mathcal{C}$ is hereditary and ϕ has infinitely many minimal induced models in $\mathcal{C}$, namely the graphs $\mathcal{H}_n$, it follows by Lemma 2 that ϕ is not equivalent to an existential formula over $\mathcal{C}$. ◀

Finally, we observe that the graphs $\mathcal{H}_n$ have bounded cliquewidth, which is easily seen to be at most 4. For this, we crucially use the fact that successive pairs are adjacent on one side and non-adjacent on the other. One could simplify the construction, e.g. by using adjacency to denote succession on both sides, but this would slightly increase the cliquewidth.

▶ **Observation 8.** *The graphs $\mathcal{H}_n$ have (linear) cliquewidth 4.*

▶ **Corollary 9.** *Extension preservation fails on $\mathcal{CW}_k$ for every $k \geq 4$.*

As witnessed by the above, orders appear to provide strong counterexamples to extension preservation. In the next section we explore preservation in certain monadically stable classes, where no such issues are expected to arise.

4 Extension preservation on strongly flip-flat classes

Local information on dense graphs can be as complicated as global information, as for instance is the case with cliques. This fact seemingly renders locality useless in the context of dense graph classes. Nonetheless, our understanding of tame classes indicates that it is still possible to recover meaningful local information, after possibly “sparsifying” our graphs in a controlled manner. The flip operation, which is central to the emerging theory of dense graph classes, plays precisely this role. We introduce it in the following definition.

▶ **Definition 10.** *Let G be a graph and $k \in \mathbb{N}$. A k-partition P of G is a partition of the vertex set into k labelled parts $P_1, \dots, P_k$, i.e. $V(G) = \bigcup_{i \in [k]} P_i$ and $P_i \cap P_j = \emptyset$ for $i \neq j$. By a k-flip F we denote a symmetric subset of $[k]^2$, i.e. a set of tuples $F = \{(i, j) : i, j \in [k]\}$ such that $(i, j) \in F \iff (j, i) \in F$. Given a k-partition P of G and a k-flip F we define the graph $G \triangle_F P$ on the same vertex set as G and on the edge set*

$$E(G \triangle_F P) := E(G) \triangle \{(u, v) : u \neq v, u \in P_i, v \in P_j, \text{ and } (i, j) \in F\}.$$

where $\triangle$ denotes the symmetric difference operation.

We note that the notation for flips existing in the literature uses the notation $\oplus$ rather than $\triangle$ (e.g. in [17]); here we have opted for the latter as the symbol $\oplus$ was used in [3] and [14] to denote the amalgamation operation. Moreover, instead of partitioning our graph, we may define k-flips by applying a sequence of at most k atomic operations, each one switching the edges and non-edges between two arbitrary subsets A, B of our vertex set. Evidently, these definitions are equivalent up to blowing up the number of flips by a value that only depends on k, while we have opted for the partition definition here to simplify our construction in Definition 14 below.

▶ **Definition 11.** *We say that a hereditary class of graphs $\mathcal{C}$ is* flip-flat[1] *if for every $r \in \mathbb{N}$ there exist $k_r \in \mathbb{N}$ and a function $f_r : \mathbb{N} \to \mathbb{N}$ satisfying that for every $m \in \mathbb{N}$ and every $G \in \mathcal{C}$ of size at least $f_r(m)$ there is a k_r-partition P of G, a k_r-flip $F \subseteq [k_r]^2$, and a set $A \subseteq V(G)$ of size at least m which is r-independent in $G \triangle_F P$. If in the above $k_r := k \in \mathbb{N}$ does not depend on r, then we say that $\mathcal{C}$ is* strongly flip-flat.

It was established in [17, Theorem 1.3] that a hereditary class of graphs is flip-flat if, and only if, it is monadically stable. In particular, every transduction of a quasi-wide class is flip-flat. The qualitative difference between strong flip-flatness and flip-flatness is precisely the same as that of almost-wideness and quasi-wideness. We make this idea precise in the following straightforward proposition, which establishes that every transduction of a uniformly almost-wide class is strongly flip-flat. For this, we use the following lemma from [29, Lemma H.3], which follows easily from Gaifman's locality theorem.

▶ **Lemma 12** (Flip transfer lemma, [29])**.** *There exists a (computable) function $\Xi : \mathbb{N}^3 \to \mathbb{N}$ satisfying the following. Fix $k, c, q \geq 1$ and $\mathcal{T}_{\delta,\phi}$ a transduction involving c colours and formulas of quantifier rank at most q. Let G, H be graphs such that $H \in T_{\delta,\phi}(G)$. Then for every k-partition P of G and k-flip F there exists a $\Xi(k, c, q)$-partition P_H of H and a $\Xi(k, c, q)$-flip F_H such that for all $u, v \in V(H)$:*

$$\mathrm{dist}_{G\triangle_F P}(u, v) \leq 2^q \cdot \mathrm{dist}_{H\triangle_{F_H} P_H}(u, v).$$

▶ **Proposition 13.** *Every transduction of a uniformly almost-wide graph class is strongly flip-flat.*

Proof. Let $\mathcal{C}$ be a uniformly almost-wide graph class and fix $k_{\mathcal{C}} \in \mathbb{N}$ witnessing this, so that for every $r, m \in \mathbb{N}$ there is $f_r(m) \in \mathbb{N}$ satisfying that every G of size at least $f_r(m)$ in the hereditary closure of $\mathcal{C}$ contains an r-independent set of size m after removing at most $k_{\mathcal{C}}$ elements. Let $\mathcal{D}$ a class such that there is a transduction $T_{\delta,\phi}$ satisfying $\mathcal{D} \subseteq T_{\delta,\phi}(\mathcal{C})$. Let $c \in \mathbb{N}$ be the number of unary predicates used by T, and q be the maximum of the quantifier ranks of δ and ϕ. We argue that $\mathcal{D}$ is strongly flip-flat with $k := \Xi(2^{k_{\mathcal{C}}}, c, q)$.

Indeed, fix $r, m \in \mathbb{N}$ and a graph $H \in \mathcal{D}$ of size at least $f_{2^q \cdot r}(m)$. It follows that there exists some $G \in \mathcal{C}$ such that $H \in T_{\delta,\phi}(G)$, and since $f_{2^q \cdot r}(m) \leq |V(H)|$, we obtain by uniform almost-wideness that $G[V(H)]$ contains a $(2^q \cdot r)$-independent set of size m after removing a set of size at most $k_{\mathcal{C}}$. In particular, there is a $2^{k_{\mathcal{C}}}$-partition P of G and a $2^{k_{\mathcal{C}}}$-flip F such that $(G\triangle_F P)[V(H)]$ contains an $(2^q \cdot r)$-independent subset of size m; call this set A. Consequently, Lemma 12 implies that there is a k-partition P of H and a k-flip F such that

[1] The original definition of flip-flatness in [17] is the uniform variant of the definition we have provided here. A simple analysis of the obstructions to monadic stability from [16], reveals that these definitions are equivalent for hereditary classes of graphs.

for all $a, b \in A \subseteq V(H)$

$$r = \frac{2^q \cdot r}{2^q} \leq \frac{\mathrm{dist}_{G \triangle_F P}(a,b)}{2^q} \leq \mathrm{dist}_{H \triangle_{F_H} P_H}(a,b),$$

i.e. A is an r-independent set of size m in $H \triangle_{F_H} P_H$. It follows that $\mathcal{D}$ is strongly flip-flat. ◀

In particular, transductions of bounded degree classes, classes of bounded shrub-depth [20], and transductions of proper minor-closed classes [4, Theorem 5.3] are all strongly flip-flat. However, obtaining preservation via locality and wideness in the style of [3, 4, 13, 14] additionally requires subtle closure assumptions. The proofs of the above articles are essentially structured into two parts. The first part argues via locality that for every formula ϕ preserved by extensions (or homomorphisms in the case of [4, 13, 14]) over a class $\mathcal{C}$ closed under substructures and disjoint unions there exist $r, m \in \mathbb{N}$ such that no minimal induced model of ϕ in $\mathcal{C}$ can contain an r-independent set of size m. In the second part, wideness is used to bound the size of minimal models of ϕ, as large enough models would have to contain r-independent sets of size m, under the proviso that a bounded number of bottleneck points have been removed. To account for the removal of these points, we have to work with an adjusted formula ϕ' in an expanded vocabulary, together with suitably adjusted structures ([2] called these *plebian companions*) on which we apply the argument of the first part. Working with ϕ', however, has translated the requirement of closure under disjoint unions to closure under a more involved operation which depends on the choice of bottlenecks. Consequently, preservation can fail on natural tame classes which do not satisfy this closure condition, e.g. for planar graphs [14, Theorem 5.8].

In the context of vertex deletions, the corresponding operation was *amalgamation over bottlenecks* [14, Theorem 4.2]. Here, we must formulate a different operation to account for the fact that flips are required to witness wideness. This is precisely the construction below.

▶ **Definition 14.** *Given $k \in \mathbb{N}$, a graph G, a k-partition P of G, and a k-flip $F \subseteq [k^2]$ we write $G \star_{(F,P)} G$ for the graph whose vertex set $V(G \star_{(F,P)} G) := V(G + G)$ is the same as the disjoint union of two copies of G, and whose edge set is*

$$E(G \star_{(F,P)} G) := E(G+G) \cup \{(u,v) : u, v \text{ are in distinct copies of } G, u \in P_i, v \in P_j, (i,j) \in F\}$$

We call this the flip-sum of G over (F, P).

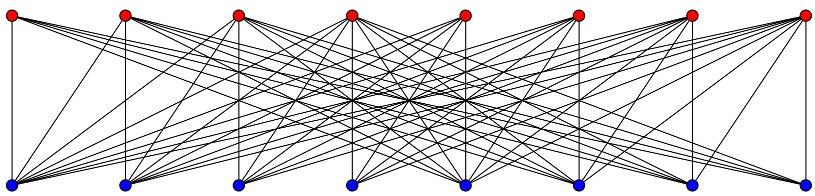

Figure 3 The graph $H \star_{(F,P)} H$, where H is the half-graph of order 4, P is the partition into red (top) and blue (bottom) vertices and $F = \{(1,2),(2,1)\}$.

We now introduce the relevant translation for the formulas.

▶ **Definition 15.** *Given a k-flip $F \subseteq [k]^2$, consider the formula*

$$E_F(x,y) := E(x,y) \triangle_{(i,j)\in F} (P_i(x) \wedge P_j(y)).$$

over the signature $\tau_E^k := \tau_E \cup \{P_1, \dots, P_k\}$, where $\triangle_{(i,j)\in F}$ denotes the consecutive application of the XOR *operator over all tuples $(i,j) \in F$. Given a τ_E-formula ϕ, we define the τ_E^k-formula*

ϕ^k obtained from ϕ by replacing every atom $E(x,y)$ with the formula $E_F(x,y)$. Moreover, for every graph G and k-partition P we write $G_{(F,P)}$ for the $\{P_1,\dots,P_k\}$-expansion of $G\triangle_F P$ where each predicate is interpreted by the respective part of P. It is then clear from the definitions and the fact that the flip operation is involutive that

$$G \models \phi \iff G_{(F,P)} \models \phi^k.$$

Our goal in Theorem 18 is to start with a strongly flip-flat class and a formula ϕ and apply the argument of [3, Theorem 4.3] to the formula ϕ^k and the structures $G_{(F,P)}$. However, as previously explained, ϕ^k is not necessarily preserved under embeddings over $\mathcal{C}$. We can nonetheless use the following easy lemma in case that the class is closed under the desired flip-sums, which will be sufficient for our purposes.

▶ **Lemma 16.** *Let $\mathcal{C}$ be a hereditary class of graphs and ϕ a formula preserved under extensions over $\mathcal{C}$. Fix a graph $G \in \mathcal{C}$, a k-partition P of G, and a k-flip $F \subseteq [k]^2$. If $G \star_{(F,P)} G \in \mathcal{C}$ then*

$$G_{(F,P)} \models \phi^k \implies G_{(F,P)} + G_{(F,P)}[S] \models \phi^k$$

for any $S \subseteq V(G)$.

Proof. Fix $\mathcal{C}, \phi, G, P, F$ as in the statement above, and let $S \subseteq V(G)$. Write G^* for the subgraph of $G \star_{(F,P)} G$ induced on the vertex set of $G + G[S]$; it follows that $G^* \in \mathcal{C}$ by hereditariness. As $G_{(F,P)} \models \phi^k$ we obtain that $G \models \phi$, and since G^* contains an induced copy of G and ϕ is preserved by extensions over $\mathcal{C}$ it follows that $G^* \models \phi$. Let P^* be the natural k-partition of G^* inherited from G, i.e. for each $i \in [k]$ the i-th part P_i^* of P^* contains the union of the i-th parts of G and $G[S]$. It follows from the definitions that the structure $G^*_{(F,P^*)}$ is isomorphic to $G_{(F,P)} + G_{(F,P)}[S]$. Finally, since $G^* \models \phi$ we obtain that $G^*_{(F,P^*)} \models \phi^k$ and so $G_{(F,P)} + G_{(F,P)}[S] \models \phi^k$ as claimed. ◀

We shall also make use of the following observation, which simply says that the induced substructures of $G_{(F,P)}$ are the same as expansions of flips of induced substructures of G.

▶ **Observation 17.** *Let G be a graph, P a k-partition of G, and F a k-flip. Then for every $S \subseteq V(G)$ the structure $G_{(F,P)}[S]$ is equal to $G[S]_{(F,P_S)}$, where P_S is the k-partition of $G[S]$ obtained by restricting each part of P on S.*

We are now ready to state the main theorem of this section.

▶ **Theorem 18.** *Fix a hereditary class of graphs $\mathcal{C}$. Suppose that there is some $k \in \mathbb{N}$ such that for all $r \in \mathbb{N}$ there is a function $f_r : \mathbb{N} \to \mathbb{N}$ satisfying that for every $m \in \mathbb{N}$ and every $G \in \mathcal{C}$ of size at least $f(m)$ there is a k-partition P of $V(G)$, some k-flip F, and $A \subseteq V(G)$ such that*

1. *$|A| \geq m$;*
2. *A is r-independent in $G\triangle_F P$;*
3. *$G \star_{(F,P)} G \in \mathcal{C}$.*

Then extension preservation holds over $\mathcal{C}$.

The proof of Theorem 18 is an adaptation of the proof of [3, Theorem 4.3], which established that extension preservation holds over any class closed under weak substructures and disjoint unions which is *wide*, i.e. for every $r \in \mathbb{N}$ there exists $f_r : \mathbb{N} \to \mathbb{N}$ such that for every $m \in \mathbb{N}$ every structure with at least $f_r(m)$-many elements contains an r-independent set of size m. Here, we replace wideness by strong flip-flatness by working with the formula ϕ^k of Definition 15. Moreover, as previously explained, addability is replaced by assumption 3 above. Preservation is then ensured by Lemma 22. Finally, going from closure under weak substructures to closure under induced substructures follows by analysing [3, Theorem 4.3]. The detailed proof can be found in Section B.

▶ **Example 19.** For $d \in \mathbb{N}$, write $\mathcal{D}_d$ be the class of all graphs G such that the maximum degree of G is at most d, or the maximum degree of $\overline{G}$, i.e. the complement graph of G, is at most d. Let $f_r(m) = (m-1)(d+1)^r + 1$ and consider a graph $G \in \mathcal{D}_d$ of size at least $f_r(m)$. If G has maximum degree d, then G must contain an r-independent set of size m. Consequently, letting $P = \{V(G)\}$ and $F = \emptyset$, we see that $G \star_{(F,P)} G$ is simply the disjoint union of two copies of G, which still has maximum degree d and is therefore in $\mathcal{D}_d$. On the other hand if $\overline{G}$ has maximum degree d, then for $P = \{V(G)\}$ and $F = \{(1,1)\}$, we see that $\bar{G} = G \triangle_F P$ has an r-independent set of size m. Since $G \star_{(F,P)} G$ is the complement of the disjoint union of two copies of $\overline{G}$ and so $\overline{G \star_{(F,P)} G}$ has maximum degree d, it follows that $G \star_{(F,P)} G \in \mathcal{D}_d$. Consequently, extension preservation holds over $\mathcal{D}_d$ by Theorem 18.

As mentioned above, Lemma 2 implies that any well-quasi-ordered class has the extension preservation property. In particular, this applies to classes of bounded *shrubdepth* [20, Corollary 3.9]. Still, in the following example we indirectly show that the class of all graphs of $\mathcal{SC}$*-depth* at most k has extension preservation by showing that it satisfies the requirements of Theorem 18, as an illustration that, although closure under flip-sums is a technical condition, it can be present in interesting tame dense classes.

▶ **Definition 20** ([20], Definition 3.5)**.** *We inductively define the class $\mathcal{SC}(k)$ as:*
- *$\mathcal{SC}(0) = \{K_1\}$;*
- *If $G_1, \dots, G_n \in \mathcal{SC}(k), H := G_1 + \dots + G_n$ and $X \subseteq V(H)$, then $\overline{H}^X := H \triangle_F P \in \mathcal{SC}(k+1)$ for $P_1 := X, P_2 := V(H) \setminus X$ and $F = \{(1,1)\}$, i.e. $\overline{H}^X$ is the graph obtained from H by flipping the edges within X.*

▶ **Example 21.** Fix $k \in \mathbb{N}$ and let $G \in \mathcal{SC}(k)$. Consider an $\mathcal{SC}$*-decomposition tree of* G, i.e. a labelled tree $\mathcal{T}$ of height $k+1$ whose leaves are labelled by the vertices of G, every non-leaf node is labelled by the graph $\overline{(G_1 + \dots G_n)}^X$ where G_i are the labels of its children, X is a subset of $\bigcup_{i \in [n]} V(G_i)$, and the root ρ is labelled by G. Let $f(m) = m^{k+1}$, and suppose that $|G| > f(m)$. Since $\mathcal{T}$ has height $k+1$ and its leaves correspond to the vertices of G, there must exist some vertex t of $\mathcal{T}$ with at least m children. Let $t_1 := \rho, t_2, \dots, t_\ell := t$ be the unique path from the root of $\mathcal{T}$ to t, and for each $i \in [\ell]$ let $X_i \subseteq V(G)$ be the set coming from the label of t_i. Letting P the partition of $V(G)$ into 2^ℓ parts depending on the membership of a vertex within each of $X_1, \dots, X_\ell$ and $F \subseteq [2^\ell]^2$ be the flip that corresponds to complementing each of $X_1, \dots, X_\ell$, it follows that $G \triangle_F P$ contains at least m distinct connected components. Let $\mathcal{T}'$ be the tree obtained from $\mathcal{T}$ by the following operation. We first create a copy of each subtree of $\mathcal{T}$ rooted at a child of t_ℓ and connect them to t_ℓ. The labels are naturally carried from each original subtree to the copy. If the label of t_ℓ in $\mathcal{T}$ was $\overline{(G_1 + \dots + G_m)}^X$, then its label in $\mathcal{T}'$ is $\overline{(G_1 + G_1' + \dots + G_m + G_m')}^{X \cup X'}$ where each G_i' corresponds to the copy of G_i, and X' corresponds to the set of copies of the vertices in X. From there, we perform the same operation for $i = \ell - 1, \dots, 1$, this time copying only the children of t_i that are not t_{i+1}. This completes the construction of $\mathcal{T}'$. It is easy to then see than the root of $\mathcal{T}'$ corresponds to the graph $G \star_{(F,P)} G$, thus witnessing that $G \star_{(F,P)} G \in \mathcal{SC}(k)$. It follows that the class $\mathcal{SC}(k)$ satisfies the requirements of Theorem 18, and thus extension preservation holds over this class.

5 Conclusion

We conclude with some questions and remarks. Firstly, it would be of independent interest to provide a characterisation of strongly flip-flat classes, akin to the characterisation of almost-wide classes via shallow minors given in [24, Theorem 3.21]. This could either be a characterisation via excluded induced subgraphs occurring in flips, in analogy to the one of monadic stability provided in [15], or in terms of *shallow vertex minors*, in analogy to the one in [8].

Moreover, it is unclear whether one can produce a formula preserved by extensions with minimal induced models of cliquewidth 3. The issue with interweaving definable orders is that one simultaneously requires for two sets to semi-induce a half-graph while (non-)adjacency is used to mark successors; this requires to keep track of at least four colours classes in a clique decomposition. We therefore leave the question of whether extension preservation holds over graphs of cliquewidth 3 open. It is also easily seen that the structures $\mathcal{H}_n$ have *twin-width* 2 (see [6] for definitions). The status of extension preservation on the class of all graphs of twin-width 1 is also unknown.

The role of orders was crucial in our construction in Section 3. In the context of undirected graphs, orders are instantiated through half-graphs. It natural to then inquire if, for every fixed $k, \ell \in \mathbb{N}$, the class of all graphs of cliquewidth at most k which omit semi-induced half-graphs of size larger than ℓ has the extension preservation property. Every such class is known to be equal to a transduction of a class of bounded treewidth by [26], and so by Proposition 13, it is strongly flip-flat. It would therefore be interesting to provide a direct combinatorial argument witnessing this, so as to be able to verify if such classes satisfy the closure requirements of Theorem 18.

References

1. Hans Adler and Isolde Adler. Interpreting nowhere dense graph classes as a classical notion of model theory. *European Journal of Combinatorics*, 36:322–330, 2014. doi:10.1016/j.ejc.2013.06.048.
2. Miklos Ajtai and Yuri Gurevich. Datalog vs first-order logic. *Journal of Computer and System Sciences*, 49(3):562–588, 1994. 30th IEEE Conference on Foundations of Computer Science. doi:10.1016/S0022-0000(05)80071-6.
3. Albert Atserias, Anuj Dawar, and Martin Grohe. Preservation under extensions on well-behaved finite structures. *SIAM Journal on Computing*, 38(4):1364–1381, 2008. doi:10.1137/060658709.
4. Albert Atserias, Anuj Dawar, and Phokion G Kolaitis. On preservation under homomorphisms and unions of conjunctive queries. *Journal of the ACM (JACM)*, 53(2):208–237, 2006. doi:10.1145/1131342.1131344.
5. John T Baldwin and Saharon Shelah. Second-order quantifiers and the complexity of theories. *Notre Dame Journal of Formal Logic*, 26(3):229–303, 1985. doi:10.1305/NDJFL/1093870870.
6. Édouard Bonnet, Eun Jung Kim, Stéphan Thomassé, and Rémi Watrigant. Twin-width I: tractable fo model checking. *ACM Journal of the ACM (JACM)*, 69(1):1–46, 2021. doi:10.1145/3486655.
7. Samuel Braunfeld, Anuj Dawar, Ioannis Eleftheriadis, and Aris Papadopoulos. Monadic nip in monotone classes of relational structures. In *50th International Colloquium on Automata, Languages, and Programming (ICALP 2023)*. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2023.
8. Hector Buffière, Eun Jung Kim, and Patrice Ossona de Mendez. Shallow vertex minors, stability, and dependence. *arXiv preprint arXiv:2405.00408*, 2024.
9. Bruno Courcelle and Stephan Olariu. Upper bounds to the clique width of graphs. *Discrete Applied Mathematics*, 101(1):77–114, 2000. doi:10.1016/S0166-218X(99)00184-5.

10 Konrad K. Dabrowski, Matthew Johnson, and Daniël Paulusma. *Clique-width for hereditary graph classes*, pages 1–56. London Mathematical Society Lecture Note Series. Cambridge University Press, 2019. doi:10.1017/9781108649094.002.

11 Peter Damaschke. Induced subgraphs and well-quasi-ordering. *Journal of Graph Theory*, 14(4):427–435, 1990. doi:10.1002/JGT.3190140406.

12 Anuj Dawar. Finite model theory on tame classes of structures. In *International Symposium on Mathematical Foundations of Computer Science*, pages 2–12. Springer, 2007. doi:10.1007/978-3-540-74456-6_2.

13 Anuj Dawar. Homomorphism preservation on quasi-wide classes. *Journal of Computer and System Sciences*, 76(5):324–332, 2010. doi:10.1016/J.JCSS.2009.10.005.

14 Anuj Dawar and Ioannis Eleftheriadis. Preservation theorems on sparse classes revisited. *arXiv preprint arXiv:2405.10887*, 2024. doi:10.48550/arXiv.2405.10887.

15 Jan Dreier, Ioannis Eleftheriadis, Nikolas Mählmann, Rose McCarty, Michał Pilipczuk, and Szymon Toruńczyk. First-order model checking on monadically stable graph classes. *arXiv preprint arXiv:2311.18740*, 2023.

16 Jan Dreier, Nikolas Mählmann, and Sebastian Siebertz. First-order model checking on structurally sparse graph classes. In *STOC 2023*, pages 567–580. ACM, 2023. doi:10.1145/3564246.3585186.

17 Jan Dreier, Nikolas Mählmann, Sebastian Siebertz, and Szymon Toruńczyk. Indiscernibles and Flatness in Monadically Stable and Monadically NIP Classes. In Kousha Etessami, Uriel Feige, and Gabriele Puppis, editors, *50th International Colloquium on Automata, Languages, and Programming (ICALP 2023)*, volume 261 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 125:1–125:18, Dagstuhl, Germany, 2023. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.ICALP.2023.125.

18 Jan Dreier, Nikolas Mählmann, and Szymon Toruńczyk. Flip-breakability: A combinatorial dichotomy for monadically dependent graph classes. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 1550–1560, 2024. doi:10.1145/3618260.3649739.

19 H-D. Ebbinghaus and J. Flum. *Finite Model Theory.* Springer, 2nd edition, 1999.

20 Robert Ganian, Petr Hliněný, Jaroslav Nešetřil, Jan Obdržálek, and Patrice Ossona De Mendez. Shrub-depth: Capturing height of dense graphs. *Logical Methods in Computer Science*, 15, 2019.

21 Martin Grohe, Stephan Kreutzer, and Sebastian Siebertz. Deciding first-order properties of nowhere dense graphs. *J. ACM*, 64(3), June 2017. doi:10.1145/3051095.

22 Yuri Gurevich. Toward logic tailored for computational complexity. *Computation and Proof Theory*, pages 175–216, 1984.

23 Ulrich Knauer and Kolja Knauer. *Algebraic Graph Theory: Morphisms, Monoids and Matrices.* De Gruyter, Berlin, Boston, 2019. doi:doi:10.1515/9783110617368.

24 Jaroslav Nešetřil and Patrice Ossona De Mendez. First order properties on nowhere dense structures. *The Journal of Symbolic Logic*, 75(3):868–887, 2010. URL: http://www.jstor.org/stable/20799288, doi:10.2178/JSL/1278682204.

25 Jaroslav Nešetřil and Patrice Ossona De Mendez. *Sparsity: graphs, structures, and algorithms*, volume 28. Springer Science & Business Media, 2012.

26 Jaroslav Nešetřil, Patrice Ossona de Mendez, Michał Pilipczuk, Roman Rabinovich, and Sebastian Siebertz. Rankwidth meets stability. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 2014–2033. SIAM, 2021.

27 Benjamin Rossman. Homomorphism preservation theorems. *Journal of the ACM (JACM)*, 55(3):1–53, 2008. doi:10.1145/1379759.1379763.

28 W. W. Tait. A counterexample to a conjecture of Scott and Suppes. *Journal of Symbolic Logic*, 24(1):15–16, 1959. doi:10.2307/2964569.

29 Szymon Toruńczyk. Flip-width: Cops and robber on dense graphs. In *2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 663–700. IEEE, 2023.

A The proof of Lemma 5

Here we provide a proof of Lemma 5, which we now restate.

▶ **Lemma 5.** *Let $n \geq 7$ and $f : \mathcal{I}_n \to \mathcal{H}_n$ be an embedding. Then f is the inclusion map.*

This is achieved in two steps. First, we consider the subgraph $\mathcal{I}'$ of $\mathcal{H}_n$ induced on $\{v_1, v_2, v_3, u_1, u_2, u_3, a\}$. Evidently, the map $g_n : \mathcal{I}' \to \mathcal{H}_n$ sending

$$(v_1, v_2, v_3, u_1, u_2, u_3, a) \mapsto (v_{n-2}, v_{n-1}, v_n, u_{n-2}, u_{n-1}, u_n, b)$$

is an embedding. We argue that this is the only non-trivial embedding of $\mathcal{I}'$ in $\mathcal{H}_n$.

▶ **Lemma 22.** *Let $n \geq 7$ and $f : \mathcal{I}' \to \mathcal{H}_n$ be an embedding. Then f is either the inclusion map or equal to g_n.*

Proof. As before, we write $V := \{v_1, \dots, v_n\} \subseteq V(\mathcal{H}_n)$ and $U := \{u_1, \dots, u_n\} \subseteq V(\mathcal{H}_n)$. We shall consider the possible images of the vertex v_2. Suppose that $f(v_2) = u_i$ for some $i \in [n]$. Clearly, since the vertices v_1, v_3, a are pairwise non-adjacent, we cannot have $f[\{v_1, v_3, a\}] \subseteq U$. We hence distinguish cases.

1. Suppose that v_1, v_3, a are all mapped to vertices in V under f. Since these are non-adjacent, we must have $f[\{v_1, v_3, a\}] = \{v_m, v_r, v_\ell\}$ for some $m + 2 < r + 1 < \ell \leq i$. Now, consider $f(u_1)$; this must be some vertex in $\mathcal{H}_n$ which is adjacent to only one of v_m, v_r, v_ℓ and not adjacent to u_i. This necessarily implies that $f(u_1) = v_{i+1}$, $f(v_1) = v_\ell$ while $\ell = i$. Consider $f(u_2)$; this must be a vertex non-adjacent to v_{i+1}, and adjacent to u_i, v_i and exactly one of $\{v_m, v_r\}$. From this we deduce that $f(u_2) = v_{i-1}$, $f(a) = v_r$, and $r = i - 2$. Finally, the vertex $f(u_3)$ must be adjacent to $v_m, v_{i-2}, v_i, u_i, v_{i+1}$ and non-adjacent to v_{i-1}; obviously no such vertex exists in $\mathcal{H}_n$, and we thus obtain a contradiction.
2. Suppose that two of v_1, v_3, a are mapped to vertices in V and one is mapped to a vertex in U. In this case we must have that $f[\{v_1, v_3, a\}] = \{u_m, v_r, v_\ell\}$ for some $m + 1 < r + 1 < \ell \leq i$. Consider $f(u_1)$; this must be non-adjacent to v_i and adjacent to exactly one of u_m, v_r, v_ℓ. This further results in two distinct cases. If $f(u_1) = v_{i+1}$, then we have $f(v_1) = v_\ell$ and $\ell = i$, which leads to a contradiction with an analogous argument to the above. If $f(u_1) = u_{i-1}$, then necessarily $f(v_1) = v_r$ while $m = i - 2, r = i - 1, \ell = i$. Considering $f(u_2)$, we now see that this vertex must be non-adjacent to u_{i-1} and adjacent to u_i, v_{i-1} and exactly one of $\{u_{i-2}, v_i\}$; evidently there is no such vertex in $\mathcal{H}_n$ and we thus obtain a contradiction.
3. Suppose that exactly one of v_1, v_3, a is mapped to a vertex in V and two are mapped to vertices of U. This forces that $f[\{v_1, v_3, a\}] = \{u_{m-1}, u_m, v_\ell\}$ for some $m < \ell \leq i$ and $m + 1 < i$. Again, consider $f(u_1)$; this is non-adjacent to u_i and adjacent to exactly one of $\{u_m, u_{m+1}, v_\ell\}$. Once again this leads to two options. If $f(u_1) = v_{i+1}$, then we have $f(v_1) = v_\ell$ and $\ell = i$, which leads to a contradiction as in Case 1. On the other hand, if $f(u_1) = u_{i-1}$ then we necessarily obtain that $f(v_1) = u_{m-1}$ while $m = i - 2$ and $\ell = i$. The vertex $f(u_2) \in \mathcal{H}_n$ must then be non-adjacent to u_{i-1} and adjacent to u_{i-1}, u_i and exactly one of u_{i-2}, v_i; since there is no such vertex in $\mathcal{H}_n$ we once again obtain a contradiction.
4. Suppose that one of v_1, v_3, a is mapped to a or b under f. Since u_3 is adjacent to all of v_1, v_2, v_3, a it must necessarily be that $f(u_3) = u_m$ for some $m > i + 1$. The vertex $f(u_2)$ must then be non-adjacent to u_m, and adjacent to u_i and exactly two of $f(v_1), f(v_3), f(a)$. As no such vertex exists in this case, we obtain a contradiction.

Since the above cases lead to a contradiction, we see that $f(v_2) \notin U$. Since no v_i for $i \in [n] \setminus \{2, n-1\}$ has three neighbours which induce an independent set, this necessarily implies that $f(v_2)$ is equal to v_2 or v_{n-1}. Assume that $f(v_2) = v_2$. Again, since v_1, v_3, a share no edges, we must necessarily have $f[\{v_1, v_3, a\}] = \{v_1, v_3, a\}$. Since u_3 is adjacent to all of v_1, v_2, v_3, a we see that $f(u_3) = u_m$ for some $m \geq 3$. As u_2 is non-adjacent to u_3 and adjacent to v_2 to exactly two of v_1, v_3, a, we see that $f(u_3) = u_3$ and $f(u_2) = u_2$, which in turn ensure that f is the inclusion map. By similar reasoning, we deduce that if $f(v_2)$ is equal to v_{n-1} then $f = g_n$ as required. ◀

Proof of Lemma 5. Let $f : \mathcal{I}_n \to \mathcal{H}_n$ be an embedding. It follows by Lemma 22 that either f is the inclusion map, or it is the map given by swapping the two induced copies of $\mathcal{I}'$, i.e. the map

$$(v_1, v_2, v_3, u_1, u_2, u_3, a) \mapsto (v_{n-2}, v_{n-1}, v_n, u_{n-2}, u_{n-1}, u_n, b);$$

$$(v_{n-2}, v_{n-1}, v_n, u_{n-2}, u_{n-1}, u_n, b) \mapsto (v_1, v_2, v_3, u_1, u_2, u_3, a).$$

Since u_{n-2} is adjacent to v_2, the latter case would imply that u_1 is adjacent to v_{n-1}, which is a contradiction. Hence, f is the inclusion map as claimed. ◀

B The proof of Theorem 18

Before proceeding with Theorem 18 we introduce some relevant definitions. Fix a relational signature τ and $q, d \in \mathbb{N}$, and let A be a τ-structure. By the *(q, d)-type of some $a \in A$* we shall mean the set containing all the MSO formulas $\theta(x)$ of quantifier rank[2] at most q, up to logical equivalence, such that $N_d^A(a) \models \theta(a)$. When we speak of a (q, d)-type t over τ, without reference to a particular element in a structure, we shall mean a (q, d)-type of some element in some τ-structure. We say that an element $a \in A$ *realises a (q, d)-type t* whenever $N_d^A(a) \models \theta(a)$ for all $\theta(x) \in t$. Evidently, the number of (q, d)-types is bounded by some $p \in \mathbb{N}$ depending only on τ and q. Given a τ-structure A, a set $C \subseteq A$, and a (q, d)-type t, we say that t is *covered by C in A* if all $a \in A$ realising t satisfy $N_d^A(a) \subseteq C$. For $n \in \mathbb{N}$ we also say that t is *n-free over C in A* if there is a $2d$-independent set $S \subseteq A$ of size n such that each $a \in S$ realises t and $N_d^A(a) \cap C = \emptyset$.

▶ **Lemma 23.** *Fix a relational signature τ and $q, d \in \mathbb{N}$. Let p be the number of (q, d)-types over τ. Then for every τ-structure A and $n \in \mathbb{N}$, there exists a radius $e \leq 2dp$ and a set $D \subseteq A$ of at most $(n-1)p$ points such that each (q, d)-type is either covered by $N_e^A(D)$ or is n-free over over $N_e^A(D)$.*

Proof. Fix an enumeration $t_1, \dots, t_p$ of all (q, d)-types over τ. We shall define D and e inductively starting at $D_0 = \emptyset$ and $e_0 = 0$. Assuming D_i and e_i have been defined, we let $C = N_{e_i}^A(D_i)$. If all types are covered by C or are n-free over C then we are done; otherwise, we let $j \in [p]$ be minimal such that t_j is neither covered by C nor n-free over C. We then define a set $E \subseteq A$ inductively, starting with $E_0 := \emptyset$ and at step $\ell + 1$ adding to E_ℓ a realisation $a \in A \setminus N_{2d}^A(C \cup E_\ell)$ of t_j if there exists one; this iteration must stop within $n - 1$ steps, as otherwise t_j would be n-free over C. In particular, $|E| \leq n - 1$ and t_j is covered by $N_{e_i+2d}^A(D_i \cup E)$. We subsequently let $D_{i+1} = D_i \cup E$ and $e_{i+1} = e_i + 2d$. It follows that the construction must stop within at most p steps, since at each step we cover a previously uncovered type, which in addition, remains covered for the rest of the construction. Consequently, $|D| \leq (n-1)p$ and $e \leq 2dp$ as claimed. ◀

[2] Here both first-order and second-order quantifiers contribute to the quantifier rank.

▶ **Theorem 18.** *Fix a hereditary class of graphs $\mathcal{C}$. Suppose that there is some $k \in \mathbb{N}$ such that for all $r \in \mathbb{N}$ there is a function $f_r : \mathbb{N} \to \mathbb{N}$ satisfying that for every $m \in \mathbb{N}$ and every $G \in \mathcal{C}$ of size at least $f(m)$ there is a k-partition P of $V(G)$, some k-flip F, and $A \subseteq V(G)$ such that*

1. $|A| \geq m$*;*
2. A *is* r*-independent in* $G \triangle_F P$*;*
3. $G \star_{(F,P)} G \in \mathcal{C}$*.*

Then extension preservation holds over $\mathcal{C}$*.*

Proof. Fix $\mathcal{C}$ as above, and let ϕ be a formula preserved by extensions overs $\mathcal{C}$. We shall obtain a bound on the size of the minimal induced models of ϕ, by arguing that any large enough model of ϕ contains a proper induced substructure which also models ϕ. We can then conclude that ϕ is equivalent to an existential formula over $\mathcal{C}$ using Lemma 2.

Letting $k \in \mathbb{N}$ be as the in the statement of Theorem 18, we consider the formula ϕ^k from Definition 15. Using Gaifman's locality theorem we rewrite ϕ^k into a boolean combination of basic local sentences, i.e. we may assume that there is some $\ell \in \mathbb{N}$ and τ_E^k-sentences ψ_i for $i \in [\ell]$ such that

$$\phi^k = \bigvee_{i \in \ell} \psi_i \text{ and } \psi_i = \bigwedge_{j \in A_i} \chi_{ij} \wedge \bigwedge_{j \in B_i} \neg \chi_{ij},$$

where each χ_{ij} is a basic local sentence. We henceforth fix the following constants:

- ρ is the maximum over all the locality radii of the χ_{ij};
- s is the sum of all widths of the χ_{ij};
- γ is the maximum over all the quantifier ranks of the χ_{ij};
- $q := \gamma + 3\rho + 3$;
- $d := 2(\rho + 1)(\ell + 1)s + 6\rho + 2$;
- p is the number of (q, d)-types over the signature τ_E^k;
- $n := (\ell + 2)s$;
- $m := (n - 1)q + s + \ell s + 1$;
- $r := 4dp + 2\rho + 1$.

Our goal is to establish that any minimal induced model of ϕ in $\mathcal{C}$ must have size less than $f_r(m)$, where f is as in the statement of Theorem 18. So, assume that some $G \models \phi$ has size at least $f_r(m)$. It follows by assumption that there is a k-partition P and a k-flip F such that $G \triangle_F P$ contains an r-independent set of size m. We henceforth work with the structure $G^* := G_{(F,P)}$, i.e. the expansion of $G \triangle_F P$ with unary predicates corresponding to the parts of P. By definition, we have that $G^* \models \phi^k$.

By Lemma 23 we obtain a radius $e \leq 2dp$ and a set $D \subseteq V(G^*)$ of at most $(n-1)p$ vertices such that each (q, d)-type in G^* is either covered by $N_e^{G^*}(D)$ or is n-free over $N_e^{G^*}(D)$; we henceforth refer to types of the former kind as *rare*, and to types of the latter kind as *frequent*.

We proceed to inductively construct increasing sequences of sets $S_0 \subseteq S_1 \subseteq \cdots \subseteq V(G^*)$, $C_0 \subseteq C_1 \subseteq \cdots \subseteq V(G^*)$, and $I_0 \subseteq I_1 \subseteq \cdots \subseteq I$ which satisfy the following conditions for every i:

1. $S_i \subseteq N_\rho^{G^*}(C_i)$;
2. $|C_i| \leq is$;
3. $|I_i| = i$;
4. no disjoint extension of $G^*[S_i]$ satisfies $\bigvee_{j \in I_i} \psi_j$;
5. $N_e^{G^*}(D)$ and $N_d^{G^*}(C_i)$ are disjoint.

Clearly, this construction must terminate within ℓ steps. Indeed, assume for a contradiction that we have constructed S_ℓ, C_ℓ, and I_ℓ satisfying conditions 1-5 above. If so, then $I_\ell = I$ while $G^* + G^*[S_\ell]$ is a disjoint extension of $G^*[S_\ell]$ which satisfies $\phi^k = \bigvee_{i \in I} \psi_i$ by Lemma 16, therefore contradicting condition 4. At the end of the construction we will obtain some $N < \ell$ and some $S_N \subsetneq V(G^*)$ satisfying $G^*[S_N] \models \phi^k$. Combining Definition 15 with Observation 17, this will imply that $G[S_N] \models \phi$, and hence that G cannot be a minimal model of ϕ as required.

Initially, we set $S_0 = C_0 = I_0 = \emptyset$. Assume that S_i, C_i, and I_i have been defined. Write $H^* := G^* + G^*[S_i]$ for the disjoint union of G^* with its substructure induced on S_i. By our closure assumptions on $\mathcal{C}$ and Lemma 16 we deduce that $H^* \models \phi^k$. In particular, there exists some $i' \in I$ such that $H^* \models \psi_{i'}$, while $i' \notin I_i$ due to property 4. We let $I_{i+1} = I_i \cup \{i'\}$ and henceforth drop the reference to the index i' as it will remain fixed for the remaining of the argument, e.g. by writing ψ and χ_j instead of $\psi_{i'}$ and $\chi_{i'j}$ respectively.

As H^* satisfies $\psi = (\bigwedge_{j \in A} \chi_j \wedge \bigwedge_{j \in B} \neg\chi_j)$, it satisfies the basic local sentences χ_j with $j \in A$. For each $j \in A$, we may thus choose a minimal set $W_j \subseteq V(H^*)$ of witnesses for the outermost existential quantifiers of the basic local sentence χ_j, and let $W := \bigcup_{j \in A} W_j$ be their union. As s is the sum of the widths of all the χ's it follows that $|W| \leq s$. We partition W into those witnesses that appear in the disjoint copy of G^*, and those that appear in the disjoint copy of $G^*[S_i]$, and write W_G and W_H for these respective parts.

Now, suppose that some $v \in W_G$ satisfies $N^{G^*}_{\rho+1}(C_i) \cap N^{G^*}_{\rho}(v) \neq \emptyset$; we argue that we may replace v with some witness $v' \in V(G^*)$ such that $N^{G^*}_{\rho+1}(C_i) \cap N^{G^*}_{\rho}(v') = \emptyset$. Indeed, we first choose some $u \in C_i$ such that $N^{G^*}_{\rho+1}(u) \cap N^{G^*}_{\rho}(v) \neq \emptyset$. Consequently, we have that $N^{G^*}_{\rho}(v) \subseteq N^{G^*}_{3\rho+1}(u) \subseteq N^{G^*}_{d}(u)$. Property 5 then ensures that the (q,d)-type t (in G^*) of u is frequent, and so it has $n > (\ell+1)s \geq |W \cup C_i|$ realisations whose d-neighbourhoods are pairwise disjoint and disjoint from $N^{G^*}_{e}(D)$. We may thus pick a realisation $u' \in V(G^*)$ of t such that $N^{G^*}_{\rho+1}(W \cup C_i) \cap N^{G^*}_{3\rho+1}(u') = \emptyset$. Let τ be the (γ,ρ)-type of v, and consider the formula

$$\theta(x) := \exists y[\forall z(\mathrm{dist}(y,z) \leq \rho \rightarrow \mathrm{dist}(x,z) \leq 3\rho+1) \wedge \bigwedge_{\eta \in \tau} \eta^{N_r(y)}(y)].$$

Clearly, the quantifier rank of θ is bounded by $3\rho + 3 + \gamma \leq q$, while $N^{G^*}_{d}(u) \models \theta(u)$ with v serving as the existential witness. Consequently $\theta(x)$ is in t, and as u and u' have the same (q,d)-type, it follows that $N^{G^*}_{d}(u') \models \theta(u')$. It follows that there is $v' \in V(G^*)$ such that $N^{G^*}_{\rho}(v') \subseteq N^{G^*}_{3\rho+1}(u') \subseteq N^{G^*}_{d}(u')$, while v and v' have the same (γ,ρ)-type. In particular, their ρ-neighbourhoods satisfy the same FO-formulas of quantifier rank $\leq \gamma$. Finally, observe that $N^{G^*}_{\rho+1}(W \cup C_i) \cap N^{G^*}_{3\rho+1}(v') = \emptyset$ and so $N^{G^*}_{\rho+1}(C_i) \cap N^{G^*}_{\rho}(v') = \emptyset$; we may thus replace v by v' in W_G as a witness.

After replacing all such witnesses in $_G$, we can ensure that

$$|\{v \in W_G : N^{G^*}_{\rho+1}(C_i) \cap N^{G^*}_{\rho}(v) \neq \emptyset\}| = 0 \tag{$\star$}$$

Consider the induced substructure $U^* := G^*[N^{G^*}_{e}(D) \cup N^{G^*}_{\rho}(W_G) \cup S_i]$. We claim that U^* satisfies $\bigwedge_{j \in A} \chi_j$. Indeed, notice that $S_i \subseteq N^{G^*}_{\rho}(C_i)$, while $N^{G^*}_{\rho+1}(C_i)$ is disjoint from $N^{G^*}_{e}(D)$ by property 5 and disjoint from $N^{G^*}_{\rho}(W_G)$ by ($\star$). It follows that U^* is the disjoint union of $G^*[N^{G^*}_{e}(D) \cup N^{G^*}_{\rho}(W_G)]$ and $G^*[S_i]$; thus all the witnesses from W and their ρ-neighbourhoods can be found in U^*, implying that $U^* \models \chi_j$ for all $j \in A$ as these are basic local sentences.

Now, observe that U^* is a proper induced substructure of G^*. This is because

$$|D \cup W_G \cup C_i| \leq (n-1)p + s + \ell s < m;$$

$$N_e^{G^*}(D) \cup N_\rho^{G^*}(W_G) \cup S_i \subseteq N_{2dp+\rho}^{G^*}(D \cup W_G \cup C_i) \subseteq N_{\lfloor r/2 \rfloor}^{G^*}(D \cup W_G \cup C_i),$$

and so, unlike G^*, U^* does not contain an r-independent set of size m. Consequently, if $U^* \models \phi^k$ then we set $S_N := N_e^{G^*}(D) \cup N_\rho^{G^*}(W_G) \cup S_i$ and our construction terminates.

We hereafter assume that $U^* \not\models \phi^k$, and proceed with the definition of S_{i+1} and C_{i+1}. Since $U^* \models \bigwedge_{j \in A} \chi_j$ it must be that $U^* \not\models \bigwedge_{j \in B} \neg\chi_j$. We can therefore fix some $j \in B$ such that $U^* \models \chi_j$. Suppose that

$$\chi_j = \exists x_1, \ldots, \exists x_{s'}[\bigwedge_{a \neq b} \text{dist}(x_a, x_b) > 2\rho' \wedge \bigwedge_a \xi^{N_{\rho'}(x_a)}(x_a)]$$

for some $\rho' \leq \rho$, $s' \leq s$, and a formula ξ of quantifier rank $\gamma' \leq \gamma$. Fix a set $V = \{w_1, \ldots, w_{s'}\} \subseteq U^*$ of witnesses for the outermost existential quantifier of χ_j. Notice that if the (q,d)-type in G^* of every $w \in V$ was rare then $N_{\rho'}^{G^*}(V) \subseteq N_e^{G^*}(D) \subseteq V(G^*)$, implying that $G^* \models \chi_j$ and thus $H^* \models \chi_j$ as H^* is a disjoint extension of G^* and χ_j is a basic local sentence. We can thus fix some $w \in V$ whose (q,d)-type in G^*, say t_w, is frequent. As a result, there is a set $Z \subseteq V(G^*)$ of n realisations of t_w whose d-neighbourhoods are pairwise disjoint and disjoint from $N_e^{G^*}(D)$. Now, since $4\rho + 3 \leq d$, $n = (\ell+2)s$, and $|C_i| \leq \ell s$, there exists a subset $Z' \subseteq Z$ of at least s elements which additionally satisfies $N_{\rho+1}^{G^*}(C_i) \cap N_\rho^{G^*}(Z') = \emptyset$.

Consider $F := N_{\rho'}^{U^*}(w)$. Evidently, $U^*[F] = G^*[F]$ and so $G^*[F] \models \xi^{N_{\rho'}(x)}(w)$. For a set variable X consider the formula $\xi^{N_{\rho'}(x) \cap X}(x, X)$ obtained from ξ by simultaneously relativising the quantifiers of ξ to the r'-neighbourhoods of x and to the set X. Observe that the quantifier rank of $\xi^{N_{\rho'}(x) \cap X}(x, X)$ is at most $\gamma' + \rho' < q$, and moreover $G^* \models \xi^{N_{\rho'}(x) \cap X}(w, F)$. Since $\rho' < d$ it follows that the MSO formula $\exists X \xi^{N_{\rho'}(x) \cap X}(x, X)$ is in t_w. As every $\omega \in Z'$ has the same (q,d)-type in G^* as w, we may find sets $F_\omega \subseteq N_{\rho'}^{G^*}(\omega)$ for every $\omega \in Z'$ such that $G^* \models \xi^{N_{\rho'}(x) \cap X}(\omega, F_\omega)$. In particular, this implies that $G^*[F_\omega] \models \xi^{N_{\rho'}(x)}(\omega)$. We finally let:

$$C_{i+1} = C_i \cup Z'; \quad S_{i+1} = S_i \cup \bigcup_{\omega \in Z'} F_\omega.$$

We argue that these satisfy the properties 1-5. First, observe that $|C_{i+1}| = |C_i| + s \leq is + s = (i+1)s$. Moreover, as $F_\omega \subseteq N_{\rho'}^{G^*}(\omega)$, $\omega \in C_{i+1}$, and $\rho' \leq \rho$ we have $S_{i+1} \subseteq N_\rho^{G^*}(C_{i+1})$. By the fact that every $\omega \in Z'$ realises a frequent type we also have that $N_e^{G^*}(D) \cap N_d^{G^*}(C_{i+1}) = \emptyset$. It remains to argue that no disjoint extension of $G^*[S_{i+1}]$ satisfies $\bigvee_{j \in I_{i+1}} \psi_j$.

Towards this, we note that $G^*[S_{i+1}]$ is a disjoint extension of $G^*[S_i]$ by the fact that $S_i \subseteq N_\rho^{G^*}(C_i)$ and $N_{\rho+1}^{G^*}(C_i) \cap N_\rho^{G^*}(Z') = \emptyset$. Therefore, no disjoint extension of $G^*[S_{i+1}]$ satisfies ψ_j for $j \in I_i$. At the same time, every disjoint extension of $G^*[S_{i+1}]$ contains witnesses for the outermost existential quantifiers of $\chi_{i'j}$, namely the elements $\omega \in Z'$, which are pairwise at distance at least $2d > 2\rho'$ and satisfy $G^*[F_\omega] \models \xi^{N_{\rho'}(x)}(\omega)$ and $N_{\rho'}^{G^*[S_{i+1}]}(\omega) = F_\omega$, and thus $G^*[S_{i+1}] \models \xi^{N_{\rho'}(x)}(\omega)$. It follows that every disjoint extension of $G^*[S_{i+1}]$ satisfies $\chi_{i'j}$, and so it cannot satisfy $\psi_{i'}$ as needed. This complete our inductive construction of S_{i+1}, C_{i+1}, and I_{i+1}. ◀