On Preservation Under Homomorphisms in the Finite

Phokion G. Kolaitis
IBM Almaden

joint work with

Albert Atserias and Anuj Dawar
UP Catalunya — Cambridge University

Preservation Theorems in Model Theory

Theme: *Semantic* properties of FO-formulas have *syntactic* counterparts

Łoś-Tarski Theorem

If a FO-sentence is preserved under extensions, then it is equivalent to an existential FO-sentence.

Lyndon's Positivity Theorem

If a FO-sentence $\varphi(P)$ is monotone in P, then it is equivalent to a positive-in-P FO-sentence $\psi(P)$.

Preservation Theorems in Model Theory

Preservation under Surjective Homomorphisms

If a FO-sentence is preserved under surjective homomorphisms, then it is equivalent to a positive FO-sentence.

Preservation under Homomorphisms

If a FO-sentence is preserved under homomorphisms, then it is equivalent to an existential positive FO-sentence.

Note: An essential use of infinite structures is made in the proofs of all preservation theorems:

- compactness theorem;
- saturated structures.

Finite Model Theory

Finite model theory is the study of logics on classes of finite structures.

Examples of Logics

- FO and fragments of FO, such as FO^k, $k \geq 2$.
- SO and fragments of SO, such as Σ^1_1.
- Fixed-point logics, such as Datalog and LFP.
- Finite-variable infinitary logics, $\mathrm{L}^k_{\infty\omega}$, $k \geq 2$.

Examples of Classes of Finite Structures

- Class of all finite graphs.
- Class of all planar graphs.
- Classes of graphs of bounded treewidth

Finite Model Theory

Question:

Do the classical preservation theorems hold in the finite?

Preservation under Extensions in the Finite

Conjecture: Scott and Suppes - 1958

The preservation-under-extensions theorem holds in the finite.

Scott and Suppes: *Foundational aspects of theories of measure*, JSL 23, 1958

Theorem: Tait - 1959

The preservation-under-extensions theorem fails in the finite.

Tait: *A counterexample to a conjecture of Scott and Suppes*, JSL 24, 1959

Rediscovered by Gurevich and Shelah - 1984

Preservation under Extensions in the Finite

Theorem: Tait - 1959

There is a FO-sentece φ that is preserved under extensions on finite structures, but is not equivalent to any existential FO-sentence on finite structures.

Proof: Vocabulary has binary relational symbols $<$ and S, constant symbols c and d.

The sentence φ is $(\psi_1 \wedge \psi_2 \wedge \psi_3 \rightarrow \psi_4)$, where

- ψ_1: $<$ is a linear order
- ψ_2: S compatible with the *successor* relation:

$$\forall x \forall y (S(x,y) \rightarrow (x < y) \wedge \neg \exists z (x < z < y))$$

- ψ_3: $\forall x (c \leq x \leq d)$
- ψ_4: $\forall x (x < d \rightarrow \exists y S(x,y))$ ∎

Lyndon's Positivity Theorem in the Finite

Theorem: Ajtai and Gurevich - 1987

There is a FO-sentence $\varphi(P)$ that is monotone in P on finite structures, but is not equivalent to any positive-in-P FO-sentence $\psi(P)$ on finite structures.

Note:

- Original proof is very complicated.
- Elementary, combinatorial proof was obtained by Stolboushkin - 1995

Homomorphisms & Preservation in the Finite

Definition: Relational structures
$\mathbf{A} = (A, R_1^{\mathbf{A}}, \ldots, R_m^{\mathbf{A}})$ and $\mathbf{B} = (B, R_1^{\mathbf{B}}, \ldots, R_m^{\mathbf{B}})$.

$h : \mathbf{A} \to \mathbf{B}$ is a *homomorphism* if

for every $i \leq m$ and every $(a_1, \ldots, a_n) \in A^n$,

$$R_i^{\mathbf{A}}(a_1, \ldots, a_n) \implies R_i^{\mathbf{B}}(h(a_1), \ldots, h(a_n)).$$

Definition: φ is *preserved under surjective homomorphisms on finite structures* means that for all finite structures $\mathbf{A}$ and $\mathbf{B}$, if $\mathbf{A} \models \varphi$ and $h : \mathbf{A} \to \mathbf{B}$ is a surjective homomorphism, then $\mathbf{B} \models \varphi$.

Definition: φ is *preserved under homomorphisms on finite structures* means that for all finite structures $\mathbf{A}$ and $\mathbf{B}$, if $\mathbf{A} \models \varphi$ and $h : \mathbf{A} \to \mathbf{B}$ is a homomorphism, then $\mathbf{B} \models \varphi$.

Pres. under Surj. Homomorphisms in the Finite

Theorem: Rosen – 1995

There is a FO-sentence that is preserved under surjective homomorphisms on finite structures, but is not equivalent to any positive FO-sentence on finite structures.

Note: The construction is modification of Stolboushkin's construction of a FO-sentence that is monotone in P on finite structures, but is not equivalent to any positive-in-P FO-sentence on finite structures.

Pres. under Homomorphisms in the Finite

Question: Let φ be a FO-sentence that is preserved under homomorphisms on finite structures. Is φ equivalent to an existential positive FO-sentence on finite structures?

Remarks:

- First raised in the 1980s. At least two different solutions were announced and later withdrawn.
- It is of interest in database theory: existential positive FO-queries are the most FAQs against database systems (and are expressible in SQL).
- An important partial result was proved in the late 1980s.

Datalog and Preserv. under Homomorphisms

Theorem: Ajtai and Gurevich - 1989

If φ is a FO-sentence is equivalent to a Datalog sentence on finite structures, then it is also equivalent to an existential positive FO-sentence on finite structures.

Note: Datalog programs are existential positive inductive definitions; hence, they are preserved under homomorphisms.

Corollary: If a Datalog program defines a first-order query on finite structures, then it is *uniformly bounded*.

Thus, Transitive Closure is not FO-definable on finite structures.

Preserv. under Homomorphisms in the Finite

Theorem: Rossman – 2005

The preservation-under-homomorphisms theorem **holds** in the finite:

If a FO-sentence is preserved under homomorphisms on finite structures, then it is equivalent to an existential-positive sentence on finite structures.

Note:

- To appear in LICS 2005.
- B. Rossman will be going to graduate school in the fall of 2005.
- Key ideas in the proof:
  - Analysis of existential positive types.
  - New notion: *existential positive saturation.*

Relativized Preservation Theorems

Fact:

As a general rule, preservation theorems do **not** relativize from a class of structures to its subclasses.

For example, the Łoś-Tarski Theorem **holds** for the class of all structures, but **fails** for the class of all finite structures.

Note:

Recall that finite model theory is the study of logics on classes of finite structures (not just the class of all finite structures).

Question:

Does the preservation-under-homomorphisms theorem hold for "interesting" classes of finite structures other than the class of all finite structures?

Preserv. under Homomorphisms in the Finite

Definition: $\mathcal{C}$ a class of finite structures.

• φ is *preserved under homomorphisms on* $\mathcal{C}$ means that for all structures $\mathbf{A}$ and $\mathbf{B}$ in $\mathcal{C}$, if $\mathbf{A} \models \varphi$ and $h : \mathbf{A} \to \mathbf{B}$ is a homomorphism, then $\mathbf{B} \models \varphi$.

We say that the *preservation-under-homomorphisms theorem holds for* $\mathcal{C}$ if every FO-sentence preserved under homomorphisms on $\mathcal{C}$ is equivalent to an existential positive FO-sentence on $\mathcal{C}$.

Note: Preservation theorems need not carry over from a class to a subclass.

Preserv. under Homomorphisms in the Finite

Main Results: (joint with A. Atserias & A. Dawar)

- The preservation-under-homomorphisms theorem holds for every class of finite structures of bounded treewidth.

More generally,

- The preservation-under-homomorphisms theorem holds for every class of finite structures whose Gaifman graphs exclude at least one minor.

- New proof of the Ajtai-Gurevich Theorem;

extension from Datalog to a certain infinitary logic.

Minimal Models

Fact: Preservation-under-homomorphisms is tightly connected to the *minimal models* of a FO-sentence.

Definition: Relational structures $\mathbf{A} = (A, R_1^{\mathbf{A}}, \ldots, R_m^{\mathbf{A}})$ and $\mathbf{B} = (B, R_1^{\mathbf{B}}, \ldots, R_m^{\mathbf{B}})$.

We say that $\mathbf{B}$ is a *submodel* of $\mathbf{A}$ if

1. $B \subseteq A$
2. For every $i \leq m$, we have that $R_i^{\mathbf{B}} \subseteq R_i^{\mathbf{A}}$.

Note: For graphs, "submodel" means "subgraph".

Definition: $\mathbf{A}$ a structure, φ a sentence.

We say that $\mathbf{A}$ is a *minimal model* of φ if

1. $\mathbf{A} \models \varphi$
2. If $\mathbf{B}$ is a proper submodel of $\mathbf{A}$, then $\mathbf{B} \not\models \varphi$

Minimal Models

Proposition: Let $\mathcal{C}$ be a class of finite structures closed under submodels, and let φ be a sentence in some logic that is preserved under homomorphisms on $\mathcal{C}$. The following are equivalent:

1. φ has finitely many minimal models in $\mathcal{C}$.
2. φ is equivalent to an existential positive first-order sentence on $\mathcal{C}$.

Proof:

$(1) \Longrightarrow (2)$: For each minimal model $\mathbf{B}$ of φ in $\mathcal{C}$, form the existential closure of the *positive diagram* of $\mathbf{B}$, and then take the disjunction of these positive existential first-order sentences. ∎

Homomorphisms, Minimal Models, and Datalog

Theorem: Ajtai and Gurevich - 1994

If a FO-sentence φ is equivalent to a Datalog sentence on finite structures, then it is also equivalent to an existential positive FO-sentence on finite structures.

Architecture of Proof:

Part I: Density of Minimal Models

If a FO-sentence φ is preserved under homomorphisms on finite structures, then all minimal models of φ satisfy a certain *density* condition D.

Part II: If all minimal models of a Datalog sentence satisfy the density condition D, then there are finitely many of them.

Gaifman Graphs and Density

Definition: Structure $\mathbf{A} = (A, R_1^{\mathbf{A}}, \ldots, R_m^{\mathbf{A}})$.

- The *Gaifman graph* of $\mathbf{A}$ is the (undirected) graph $\mathcal{G}(\mathbf{A}) = (A, E)$ such that

  $E(a, a') \iff a$, a' occur in a tuple of $R_i^{\mathbf{A}}$, for some i.

- For every w in A, let $N_d(\mathcal{G}(\mathbf{A}), w)$ be the set of all elements of A having distance less than d from w in $\mathcal{G}(\mathbf{A})$.

- A set $S \subseteq A$ is *d-scattered* if for every two distinct elements u and v of S,

$$N_d(\mathcal{G}(\mathbf{A}, u)) \cap N_d(\mathcal{G}(\mathbf{A}, v)) = \emptyset.$$

Density Lemma: Ajtai and Gurevich - 1994

Let φ be a FO-sentence that is preserved under homomorphisms on finite structures. There are d and m such that if $\mathbf{A}$ is a minimal model of φ, then $\mathcal{G}(\mathbf{A})$ does not contain a d-scattered set with m elements.

Extended Density Lemma

The Density Lemma can be extended as follows:

Extended Density Lemma: Let $\mathcal{C}$ be a class of finite structures closed under submodels and disjoint unions, and let φ be a FO-sentence that is preserved under homomorphisms on $\mathcal{C}$. For every $s \geq 0$, there are d and m such that if $\mathbf{A}$ is a minimal model of φ, then there is no $B \subseteq A$ such that $|B| \leq s$ and $\mathcal{G}(\mathbf{A}) - B$ contains a *d-scattered* set with m elements.

Proof: Use Gaifman's Locality Theorem - 1982.

Extended Density and Homomorphisms

Corollary: Let $\mathcal{C}$ be a class of finite structures having the following properties:

1. $\mathcal{C}$ is closed under submodels and disjoint unions.
2. There is $s \geq 0$ such that for all d and m there is N such that if $\mathbf{A} \in \mathcal{C}$ and $|A| > N$, then there is $B \subseteq A$ such that $|B| \leq s$ and $\mathcal{G}(\mathbf{A}) - B$ contains a *d-scattered* set with m elements.

Then the preservation-under-homomorphisms theorem holds for $\mathcal{C}$.

Proposition: The preservation-under-homs. theorem holds for every class $\mathcal{C}$ closed under sumbodels and disjoint unions, and such that for some k the Gaifman graph of every structure in $\mathcal{C}$ has maximum degree k.

Proof: Since $|N_d(\mathcal{G}(\mathbf{A}), w)| \leq k^d$, we can take $s = 0$, $N = mk^d$. ∎

Treewidth

Observation: Many algorithmic problems that are "hard" on arbitrary graphs become "easy" on trees.

Question: Can the concept of *tree* be weakened to a *tree-like* concept while maintaining "good" computational properties?

Answer: Robertson & Seymour – 1985–1995

Yes, the concept of *bounded treewidth* achieves this.

Note: Part of their work on *graph minors*.

Diestel: "Graph Theory"

Downey & Fellows: "Parametrized Complexity"

k-Trees, Partial k-Trees, Treewidth

Definition: Arnborg – 1987

Let k be a positive integer.

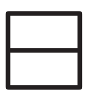
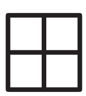
- A graph $\mathbf{H}$ is a *k-tree* if
  1. $\mathbf{H}$ is a $\mathbf{K}_{k+1}$-clique, or
  2. There are a k-tree $\mathbf{G}$, nodes $a_1, \ldots, a_k$ in $\mathbf{G}$ forming a $\mathbf{K}_k$-clique, and a node b in $\mathbf{H} - \mathbf{G}$ such that $\mathbf{H}$ is obtained from $\mathbf{G}$ by connecting b to each of the nodes $a_1, \ldots, a_k$ (thus, forming a $\mathbf{K}_{k+1}$-clique).

- A graph is a *partial k-tree* if it is a subgraph (not necessarily induced) of a k-tree.

- The *treewidth* of a graph $\mathbf{H}$, denoted by $\mathrm{tw}(\mathbf{H})$, is the smallest k such that $\mathbf{H}$ is a partial k-tree.

Treewidth: Examples

Fact:

- If $\mathbf{T}$ is a tree, then $\mathrm{tw}(\mathbf{T}) = 1$.
- If $\mathbf{C}$ is a cycle, then $\mathrm{tw}(\mathbf{C}) = 2$.
- tw() = 2.
- tw() = 3.
- $\mathrm{tw}(\mathbf{K}_k) = k - 1$, for every $k \geq 2$.
- The $k \times k$-grid has treewidth k.

Bounded Treewidth and Homomorphisms

Definition:

- If $\mathbf{A}$ is a structure, then $\mathrm{tw}(\mathbf{A}) = \mathrm{tw}(\mathcal{G}(\mathbf{A}))$.
- If $k \geq 2$, then $\mathcal{T}(k) = \{\mathbf{A} : \mathrm{tw}(\mathbf{A}) < k\}$.
- We say that a class $\mathcal{C}$ of finite structures has *bounded treewidth* if $\mathcal{C} \subseteq \mathcal{T}(k)$, for some $k \geq 2$.

Theorem: Let $\mathcal{C}$ be a class of finite structures having the following properties:

1. $\mathcal{C}$ is closed under submodels and disjoint unions.
2. $\mathcal{C}$ has bounded treewidth.

Then the preservation-under-homomorphisms theorem holds for class $\mathcal{C}$.

Bounded Treewidth and Homomorphisms

The proof of the preservation-under-homomorphisms theorem for classes of bounded treewidth uses the following combinatorial result.

Treewidth Density Lemma: For every k, d, and m, there is N such that for all graphs $\mathbf{G} = (V, E)$ with $\mathrm{tw}(\mathbf{G}) < k$ and $|V| > N$, there exists $B \subseteq V$ such that $|B| \leq k$ and $\mathbf{G} - B$ has a d-scattered set with m elements.

In turn, this uses the following result

Sunflower Lemma: Erdős and Rado - 1960

Let F be a collection of k-element subsets of a set A. If $|F| > k!(p-1)^k$, then F contains a *sunflower with p petals*, that is, a subcollection $F' \subseteq F$ such that every pair of distinct sets in F' has a common intersection B.

Graph Minors

Definition: A graph $\mathbf{G} = (V, E)$ is a *minor* of a graph $\mathbf{H} = (V', E')$, denoted $\mathbf{G} \prec \mathbf{H}$, if $\mathbf{G}$ can be obtained from $\mathbf{H}$ by a sequence of node deletions, edge deletions, and edge contractions.

Example: $\mathbf{K}_4 \prec \mathbf{K}_{3,3}$

Basic Facts:

- $\prec$ is a transitive relation
- If $\mathbf{G} \prec \mathbf{H}$, then $\mathrm{tw}(\mathbf{G}) \leq \mathrm{tw}(\mathbf{H})$.

  Hence,
- If $\mathrm{tw}(\mathbf{H}) < k$, then $\mathbf{K}_{k+1} \not\prec \mathbf{H}$.

Excluded Minors and Homomorphisms

Theorem: Let $\mathcal{C}$ be a class of finite structures having the following properties:

1. $\mathcal{C}$ is closed under submodels and disjoint unions.
2. There is a graph $\mathbf{G}$ that is not a minor of the Gaifman graph of some structure in $\mathcal{C}$.

Then the preservation-under-homomorphisms theorem holds for $\mathcal{C}$.

Proof: Use Ramsey's Theorem to prove a *Density Lemma* for classes of graphs that exclude a given graph as minor. ∎

Corollary: The preservation-under-homomorphisms theorem holds for the class of all planar graphs.

Proof: A graph $\mathbf{H}$ is planar if and only if $\mathbf{K}_5 \not\preceq \mathbf{H}$ and $\mathbf{K}_{3,3} \not\preceq \mathbf{H}$. ∎

Bounded Treewidth vs. Excluded Minors

Theorem: Robertson and Seymour - 1986

For every graph **G**, the following statements are equivalent:

1. **G** is planar.

2. The class of graphs excluding **G** as a minor has bounded treewidth.

Note: Thus, the preservation-under-homomorphisms theorem for classes excluding **G** as minor strictly subsumes the preservation-under-homomorphisms theorem for classes of bounded treewidth precisely when **G** is non-planar.

Synopsis

- The preservation-under-homomorphisms theorem holds for numerous classes of finite structures, including:
  - the classes of all finite structures of bounded treewidth, the class of all planar graphs, and, more generally,
  - the classes of all finite structures whose Gaifman graphs exclude a given graph as minor.
- Earlier work on finite model theory had focused on negative results for the class of all finite structures.

  However, positive results can be obtained by *relativizing* on classes of finite structures of graph-theoretic or combinatorial interest.

Datalog

Definition: A Datalog program is a function-free, negation-free, $\neq$-free logic program.

Example: Transitive Closure

$$\begin{array}{|lcl} T(x,y) & :- & E(x,y) \\ T(x,y) & :- & E(x,z) \wedge T(z,y) \end{array}$$

Fact: Datalog programs are existential positive first-order inductive definitions.

Example:

The Transitive Closure of E is the least fixed-point $\varphi^{\infty}(x,y)$ of the existential positive first-order formula

$$\varphi(x,y,T) \equiv E(x,y) \vee \exists z(E(x,z) \wedge T(z,y))$$

Datalog

Example: 2-COLORABILITY

Recall that a graph is 2-colorable if and only if it contains no odd cycle.

Datalog program for NON 2-COLORABILITY

$$\left|\begin{array}{lcl} O(X,Y) & :- & E(X,Y) \\ O(X,Y) & :- & O(X,Z), E(Z,W), E(W,Y) \\ Q & :- & O(X,X) \end{array}\right.$$

Fact: Every Datalog formula is preserved under homomorphisms.

Proof: Every Datalog formula is equivalent to an infinitary disjunction of existential positive first-order formulas.

Ajtai-Gurevich Theorem Revisited

Theorem: Ajtai and Gurevich - 1994

If a FO-sentence φ is equivalent to a Datalog sentence on finite structures, then it is also equivalent to an existential positive FO-sentence on finite structures.

New Proof: Main ingredients:

- Treewidth Density Lemma
- Connection between the number of variables of a Datalog program and the treewidth of its minimal models.

k-Datalog

Definition: A k-Datalog program is a Datalog program in which each rule

$$t_0 \ :- \ t_1, \ldots, t_m$$

has at most k distinct variables.

Example: Non 2-Colorability revisited

$$\left|\begin{array}{lll} O(X,Y) & :- & E(X,Y) \\ O(X,Y) & :- & O(X,Z), E(Z,W), E(W,Y) \\ Q & :- & O(X,X) \end{array}\right.$$

Therefore,

Non 2-Colorability is definable in 4-Datalog.

Datalog and Bounded Treewidth

Definition: Let k be a positive integer.

L^k is the collection of all FO-formulas with variables among $x_1, \ldots, x_k$, containing all atomic formulas in the variables $x_1, \ldots, x_k$, and closed under $\wedge$ and $\exists$.

Lemma 1: Every k-Datalog sentence φ is equivalent to an infinitary disjunction $\bigvee_{i=1}^{\infty} \theta_i$, where each θ_i is an L^k-sentence.

Lemma 2: Every minimal model of an L^k-sentence senten is the homomorphic image of another minimal model of treewidth less than k.

Extending the Ajtai-Gurevich Theorem

Definition:

$\bigvee \mathrm{L}^k$ is the collection of all infinitary disjunctions of L^k-sentences.

Note: $k-$Datalog $\subset \bigvee \mathrm{L}^k$

Theorem: If a FO-sentence φ is equivalent to a $\bigvee \mathrm{L}^k$-sentence on finite structures, then it is also equivalent to a finite disjunction of L^k-sentences.