

Homomorphism Preservation Theorem

Albert Atserias

Universitat Politècnica de Catalunya

Barcelona, Spain

Structure of the talk

1. Classical preservation theorems
2. Preservation theorems in finite model theory and CSPs
3. Rossman's Theorem
4. More on preservation theorems in finite model theory

Classical Preservation Theorems

Existential Formulas (1)

The smallest class of **formulas** that

- contains all atomic formulas $R(x_1, \dots, x_r)$ and $x_1 = x_2$,
- contains all negated atomic formulas $\neg R(x_1, \dots, x_r)$ and $x_1 \neq x_2$,
- is closed under conjunction $\varphi_1 \wedge \varphi_2$,
- is closed under disjunction $\varphi_1 \vee \varphi_2$,
- is closed under existential quantification $(\exists x)(\varphi(x))$,
- is closed under universal quantification $(\forall x)(\varphi(x))$.

Existential Formulas (2)

In the vocabulary of **graphs** $\sigma = \{E\}$:

Example 1:

There is an **induced** pentagon.

$$(\exists x_1) \cdots (\exists x_5) (E(x_1, x_2) \wedge E(x_2, x_3) \wedge \cdots \wedge \neg E(x_5, x_2) \wedge \neg E(x_5, x_3)).$$

Example 2:

Either there is an **induced** pentagon or an **induced** co-pentagon.

$$\exists C_5 \vee \exists \overline{C_5}.$$

Existential Formulas (3)

Easy observation:

if $A \models \varphi$ and A is an induced substructure of B , then $B \models \varphi$.

Proof by induction on the structure of φ .

Question:

Are existential formulas the only first-order formulas that are preserved under extensions?

Existential Formulas (4)

Łoś-Tarski Theorem

Let φ be a first-order formula. The following are equivalent:

- φ is preserved under extensions,
- φ is equivalent to an existential formula.

Proof uses the Compactness Theorem for first-order logic.

Existential Positive Formulas (1)

The smallest class of **formulas** that

- contains all atomic formulas $R(x_1, \dots, x_r)$ and $x_1 = x_2$,
- contains all negated atomic formulas $\neg R(x_1, \dots, x_r)$ and $x_1 \neq x_2$,
- is closed under conjunction $\varphi_1 \wedge \varphi_2$,
- is closed under disjunction $\varphi_1 \vee \varphi_2$,
- is closed under existential quantification $(\exists x)(\varphi(x))$,
- is closed under universal quantification $(\forall x)(\varphi(x))$.

Existential Positive Formulas (2)

In the vocabulary of **graphs** $\sigma = \{E\}$:

Example 1:

There is a **homomorphic copy** of the pentagon.

$$(\exists x_1) \cdots (\exists x_5) (E(x_1, x_2) \wedge E(x_2, x_3) \wedge \cdots \wedge E(x_5, x_1)).$$

Example 2:

Either there is a **homomorphic copy** of a triangle
or a **homomorphic copy** of a pentagon.

$$\exists^+ C_3 \vee \exists^+ C_5.$$

Existential Positive Formulas (3)

Easy observation:

if $\mathbf{A} \models \varphi$ and there is a homomorphism $\mathbf{A} \rightarrow \mathbf{B}$, then $\mathbf{B} \models \varphi$.

Proof by induction on the structure of φ .

Question:

Are existential positive formulas the only first-order formulas that are preserved under homomorphisms?

Existential Positive Formulas (4)

Lyndon-Łoś-Tarski Theorem

Let φ be a first-order formula. The following are equivalent:

- φ is preserved under homomorphisms,
- φ is equivalent to an existential positive formula.

Proof uses the Compactness Theorem for first-order logic.

(Rossman gave a new proof! Coming soon)

Primitive Positive Formulas (1)

The smallest class of **formulas** that

- contains all atomic formulas $R(x_1, \dots, x_r)$ and $x_1 = x_2$,
- contains all negated atomic formulas $\neg R(x_1, \dots, x_r)$ and $x_1 \neq x_2$,
- is closed under conjunction $\varphi_1 \wedge \varphi_2$,
- is closed under disjunction $\varphi_1 \vee \varphi_2$,
- is closed under existential quantification $(\exists x)(\varphi(x))$,
- is closed under universal quantification $(\forall x)(\varphi(x))$.

Primitive Positive Formulas (2)

Easy observation:

if $\mathbf{A} \models \varphi$ and there is a homomorphism $\mathbf{A} \rightarrow \mathbf{B}$, then $\mathbf{B} \models \varphi$,
if $\mathbf{A} \models \varphi$ and $\mathbf{B} \models \varphi$, then $\mathbf{A} \times \mathbf{B} \models \varphi$.

Proof by induction on the structure of φ .

Note: Primitive positive formulas (in prenex normal form) are precisely the conjunctive queries.

Primitive Positive Formulas (3)

Corollary to Lyndon-Łoś-Tarski

Let φ be a first-order formula. The following are equivalent:

- φ is preserved under products and homomorphisms,
- φ is equivalent to a primitive positive formula.

Primitive Positive Formulas (3)

Corollary to Lyndon-Łoś-Tarski

Let φ be a first-order formula. The following are equivalent:

- φ is preserved under products and homomorphisms,
- φ is equivalent to a primitive positive formula.

Proof: As φ is preserved under homomorphisms, we have

$$\varphi \equiv \exists^+ \mathbf{A}_1 \vee \dots \vee \exists^+ \mathbf{A}_m.$$

Then $\mathbf{A}_1, \dots, \mathbf{A}_m$ are models of φ , so $\mathbf{A}_1 \times \dots \times \mathbf{A}_m \models \varphi$. Hence, $\mathbf{A}_i \rightarrow \mathbf{A}_1 \times \dots \times \mathbf{A}_m$ for some i , and $\mathbf{A}_i \rightarrow \mathbf{A}_j$ for every j , so

$$\mathbf{A} \models \varphi \Rightarrow \mathbf{A} \models \exists^+ \mathbf{A}_i \Rightarrow \mathbf{A} \models \varphi.$$

Hence $\varphi \equiv \exists^+ \mathbf{A}_i$. \square

Other Classical Preservation Theorems

Many other similar results are known:

- Preserved under surjective hom. : positive formulas (Lyndon).
- Preserved under unions of chains : Π_2 formulas.
- Preserved under reduced products : Horn formulas.
- ...

Preservation Theorems in Finite Model Theory and CSP

Compactness Theorem Fails on Finite Models

Compactness Theorem

Let Γ be a collection of first-order formulas. If every finite subset of Γ has a model, then Γ has a model.

Failure on the finite:

- $\varphi_n = (\exists x_1) \cdots (\exists x_n) (\bigwedge_{i \neq j} x_i \neq x_j)$.
- $\Gamma = \{\varphi_n : n \geq 1\}$.

Every finite subset has a finite model, but all models of Γ are infinite.

But do Preservation Theorems Fail?

Most fail, indeed:

- Tait (1959!): Preservation under extensions fails.
- Ajtai-Gurevich (1987): Preservation under surjective hom. fails.
- ...
- Preservation under homomorphisms: [does it hold?](#)

Remained open for many years... solved in 2005 (holds) by Rossman.

While Waiting for Ben ...

Let us restrict attention to classes of the form

$$\neg\text{CSP}(\mathbf{H}) = \{A : A \not\rightarrow \mathbf{H}\}.$$

Note: $\neg\text{CSP}(\mathbf{H})$ is automatically closed under homomorphisms:

If $A \not\rightarrow \mathbf{H}$ and $A \rightarrow B$, then $B \not\rightarrow \mathbf{H}$.

Question:

Suppose $\neg\text{CSP}(\mathbf{H})$ is first-order definable.
Is it definable by an **existential-positive formula**?

Preservation Theorem for CSPs

Theorem (A. 2005)

Yes.

- Much **weaker** result than Rossman's.
- But proof much **simpler** as well.
- Proof uses a **key result** in graph-theoretic approach to CSP.
- Proof builds on well-known previous work.
- ...

Overview of Proof

Fix \mathbf{H} and suppose φ is a first-order formula that defines $\neg\text{CSP}(\mathbf{H})$.

Ingredients:

- Random preimage lemma.
- Scattered sets on large-girth structures.
- Density lemma.
- Putting it all together.

Definitions

Definitions: Let \mathbf{A} be a finite structure:

- \mathbf{A} is a **minimal model** of φ if $\mathbf{A} \models \varphi$ and no proper substructure $\mathbf{B} \subset \mathbf{A}$ is a model of φ .
- **Gaifman graph** $\mathcal{G}(\mathbf{A})$ of \mathbf{A} : vertices are elements of \mathbf{A} , edges are pairs of elements that appear together in some tuple of \mathbf{A} .
- A **d -scattered set** is a set of elements that are pairwise at distance $> d$ in the Gaifman graph.

Ingredient 1 : Random Preimage Lemma

Interesting history:

[Erdős 1959] There exist graphs of arbitrary **large girth** and arbitrary **large chromatic number**.

[Nešetřil-Rödl 1979] Sparse Incomparability Lemma: for every non-bipartite graph $G \rightarrow H$ there is an **incomparable** $G' \not\rightarrow G$ with $G' \rightarrow H$ and **large girth**.

[Feder-Vardi 1993-1998] Same proof technique as in Sparse Incomparability Lemma but slightly different statement and for arbitrary structures. We call it the **Random Preimage Lemma**.

Random Preimage Lemma

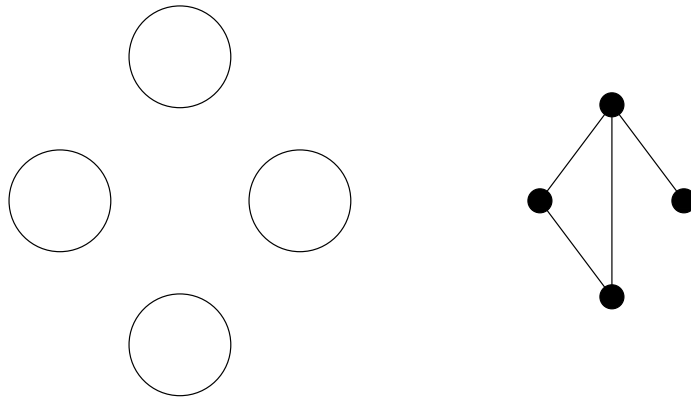
Random Preimage Lemma (Feder-Vardi 1993-1998)

For every g and k and for every finite structure \mathbf{A} , there exists a finite structure \mathbf{A}' such that:

- (preimage) $\mathbf{A}' \rightarrow \mathbf{A}$,
- (equivalence) $\mathbf{A}' \rightarrow \mathbf{H}$ iff $\mathbf{A} \rightarrow \mathbf{H}$, whenever $|\mathbf{H}| \leq k$,
- (large girth) the girth of \mathbf{A}' as a hypergraph is at least g .

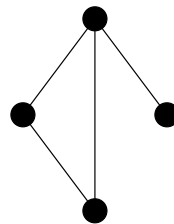
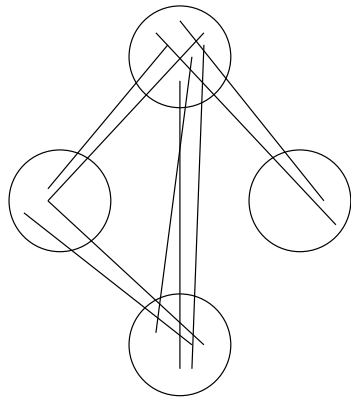
Why is the Random Preimage Lemma true?

1. Replace every point of A by a **big independent set** (a potato).



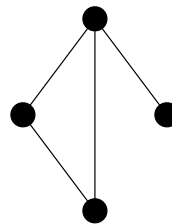
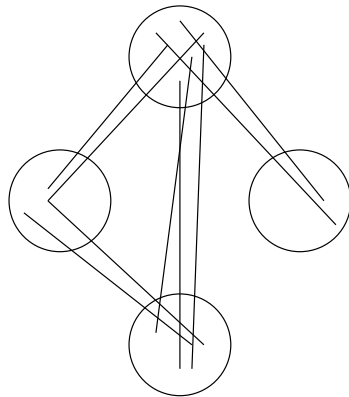
Why is the Random Preimage Lemma true?

1. Replace every point of A by a **big independent set** (a potato).
2. Throw **random edges** between independent sets that are edges in A .



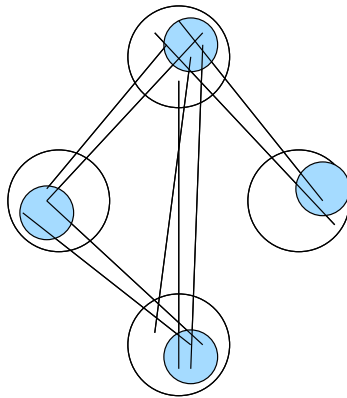
Why is the Random Preimage Lemma true?

1. Replace every point of A by a **big independent set** (a potato).
2. Throw **random edges** between independent sets that are edges in A .
3. Remove one edge from every short cycle to **force large girth**.



Pigeonhole Argument

Every homomorphism to a small \mathbf{H} , $|\mathbf{H}| \leq k$, must be **constant** on a large portion of every independent set.



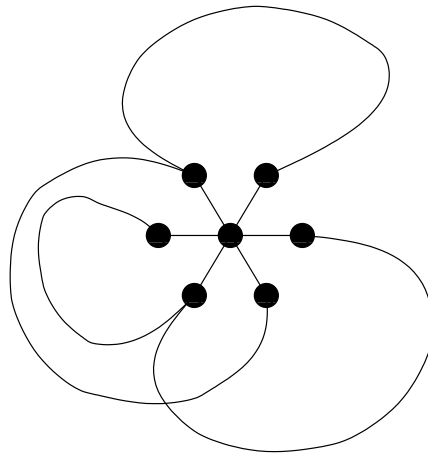
Hence,

$$\begin{aligned} \mathbf{A}' \rightarrow \mathbf{H} &\text{ implies } \mathbf{A} \rightarrow \mathbf{H}. \\ \mathbf{A} \rightarrow \mathbf{H} &\text{ implies } \mathbf{A}' \rightarrow \mathbf{A} \rightarrow \mathbf{H}. \end{aligned}$$

Ingredient 2 : Scattered Sets

Lemma: Let d , m and g be integers. For every large finite structure A of girth at least g , there exists a such that $A \setminus a$ contains a d -scattered of size m .

- Case of **small degree** is easy.
- Case of **large degree**:



Ingredient 3: Density Lemma

Density Lemma (Ajtai-Gurevich 1988)

Let φ be preserved under homomorphisms on finite models. There exist d and m such that if \mathbf{A} is a minimal model of φ , then \mathbf{A} has no d -scattered set of size m . Moreover, neither does $\mathbf{A} \setminus a$ for every $a \in A$.

Proof depends on [Gaifman's Locality Theorem](#).

Ingredient 4 : Putting it all together

Suppose $\neg\text{CSP}(\mathbf{H})$ definable by φ . We prove φ has **finitely many minimal models** $\mathbf{A}_1, \dots, \mathbf{A}_m$. This is enough because then

$$\varphi \equiv \exists^+ \mathbf{A}_1 \vee \dots \vee \exists^+ \mathbf{A}_m.$$

We bound the size of minimal models.

Suppose \mathbf{A} is a **very large** minimal model.

Using Ingredient 1

A large minimal model, in particular $A \not\rightarrow H$

\Downarrow (Ingredient 1: random preimage)

there is a large minimal model $A'' \subseteq A' \rightarrow A$ of large girth

Using Ingredient 2

A'' has large girth

⇓ (Ingredient 2: scattered sets)

there is an $a \in A''$ such that $A'' \setminus a$ has a **large scattered set**

Using Ingredient 3

there is an $a \in A''$ such that $A'' \setminus a$ has a large scattered set

↓ (Ingredient 3: density lemma)

A'' is **not a minimal model**. Contradiction. \square

Rossman's Theorem

Finite Homomorphism Preservation Theorem

Theorem (Rossman 2005)

Let φ be a first-order formula. The following are equivalent:

- φ is **preserved under homomorphisms** on finite structures,
- φ is equivalent to an **existential positive formula** on finite structures.

Proof (cook) plan:

- Ingredient 1: Existential positive types.
- Ingredient 2: Existential positive saturation.
- Ingredient 3: Finitization and back and forth argument.

Ingredient 1 : Existential positive types

Definitions: Let \mathbf{A} be a structure, let \mathbf{a} be a k -tuple, let n be an integer.

- the **quantifier rank** of a formula is the nesting-depth of quantifiers.
- fo_n : collection of all first-order formulas of quantifier rank $\leq n$.
- pp_n : collection of all primitive positive formulas of quantifier rank $\leq n$.
- $\text{fo}_n(\mathbf{A}, \mathbf{a})$: collection of all $\varphi \in \text{fo}_n$ such that $\mathbf{A} \models \varphi(\mathbf{a})$.
- $\text{pp}_n(\mathbf{A}, \mathbf{a})$: collection of all $\varphi \in \text{pp}_n$ such that $\mathbf{A} \models \varphi(\mathbf{a})$.
- we write $\mathbf{A} \leq_n^{\text{fo}} \mathbf{B}$ if $\text{fo}_n(\mathbf{A}) \subseteq \text{fo}_n(\mathbf{B})$.
- we write $\mathbf{A} \leq_n^{\text{pp}} \mathbf{B}$ if $\text{pp}_n(\mathbf{A}) \subseteq \text{pp}_n(\mathbf{B})$.

Main Property of Types

Fact: Let \mathcal{C} be a class of structures. The following are equivalent:

- \mathcal{C} is preserved under \leq_n^{pp} ,
- \mathcal{C} is definable by a finite disjunction of formulas in pp_n .

Fact: Let \mathcal{C} be a class of structures. The following are equivalent:

- \mathcal{C} is preserved under \leq_n^{fo} ,
- \mathcal{C} is definable by a finite disjunction of formulas in fo_n .

Proofs: Follows easily from

the **number of formulas** of quantifier rank $\leq n$,
up to logical equivalence, is **finite**.

Goal

Given $\varphi \in \text{fo}_n$ preserved under homomorphisms, we want to find n' so that for every \mathbf{A} and \mathbf{B} with $\mathbf{A} \leq_{n'}^{\text{pp}} \mathbf{B}$, there exist \mathbf{A}^* and \mathbf{B}^* so that

$$\begin{array}{ccc} \mathbf{A} & \leq_{n'}^{\text{pp}} & \mathbf{B} \\ \downarrow & & \uparrow \\ \mathbf{A}^* & \leq_n^{\text{fo}} & \mathbf{B}^* \end{array}$$

Ingredient 2 : Existential Positive Saturation

Definition: Let A be a structure and let n and k be integers. We say that A is **existential-positively n -saturated** if for every $k \leq n$ we have

$$(\forall \mathbf{a} \in A^{k-1})(\forall a \in A)(\forall T, \text{pp}_{n-k}(A, \mathbf{a}) \subseteq T \subseteq \text{pp}_{n-k}(A, \mathbf{a}a)) \\ (\exists b \in A)(\text{pp}_{n-k}(A, \mathbf{a}b) = T).$$

Main Property: If A and B are existential-positively n -saturated, then

$$A \equiv_n^{\text{pp}} B \Rightarrow A \equiv_n^{\text{fo}} B.$$

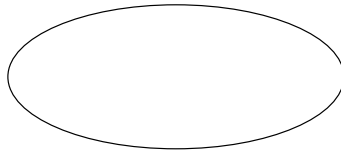
Proof: By induction on $n - k$ prove, for every $\mathbf{a} \in A^k$ and $\mathbf{b} \in B^k$:

$$\text{pp}_{n-k}(A, \mathbf{a}) = \text{pp}_{n-k}(B, \mathbf{b}) \Rightarrow \text{fo}_{n-k}(A, \mathbf{a}) = \text{fo}_{n-k}(B, \mathbf{b}).$$

But can we saturate structures anyway?

Theorem: Yes. But oops, they become infinite!

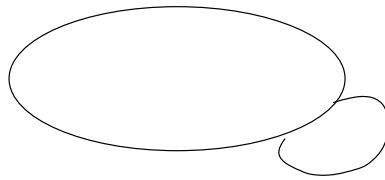
Proof by iterated ear-construction:



But can we saturate structures anyway?

Theorem: Yes. But oops, they become infinite!

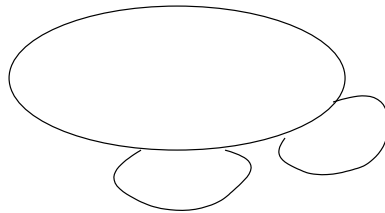
Proof by iterated ear-construction:



But can we saturate structures anyway?

Theorem: Yes. But oops, they become infinite!

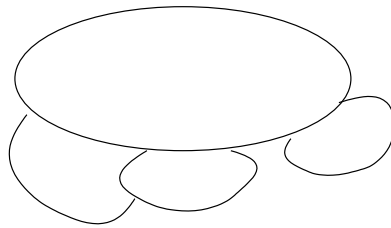
Proof by iterated ear-construction:



But can we saturate structures anyway?

Theorem: Yes. But oops, they become infinite!

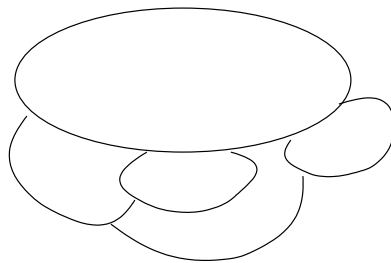
Proof by iterated ear-construction:



But can we saturate structures anyway?

Theorem: Yes. But oops, they become infinite!

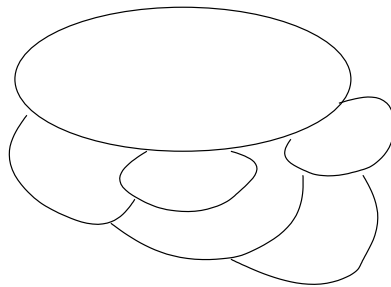
Proof by iterated ear-construction:



But can we saturate structures anyway?

Theorem: Yes. But oops, they become infinite!

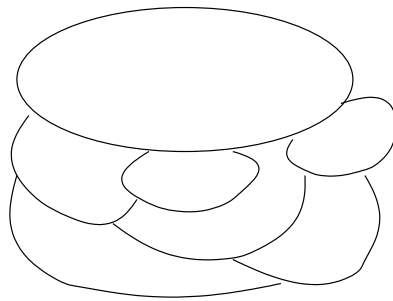
Proof by iterated ear-construction:



But can we saturate structures anyway?

Theorem: Yes. But oops, they become infinite!

Proof by iterated ear-construction:



Ingredient 3 : Finitization and back and forth

Theorem: For every n , there exists $n' \geq n$ such that if \mathbf{A} and \mathbf{B} are finite structures with $\mathbf{A} \leq_{n'}^{\text{pp}} \mathbf{B}$, then there exist finite \mathbf{A}^* and \mathbf{B}^* such that

$$\begin{array}{ccc} \mathbf{A} & \leq_{n'}^{\text{pp}} & \mathbf{B} \\ \downarrow & & \uparrow \\ \mathbf{A}^* & \leq_n^{\text{fo}} & \mathbf{B}^* \end{array}$$

Note: \mathbf{A}^* and \mathbf{B}^* are *sufficiently* saturated.

More preservation theorems in finite model theory

Restricted Classes of Finite Structures

Let \mathcal{T} be the class of finite **oriented forests**. Suppose φ is a first-order formula that is preserved under homomorphisms on \mathcal{T} .

Is φ equivalent, on \mathcal{T} , to an existential-positive formula?

Note: Doesn't seem to follow from Rossman's result.

But the answer is **yes**. **Why?**: every large forest has a large scattered set after removing (at most) one point, so it cannot be a minimal model of φ by the Density Lemma. \square

Restricted Classes of Finite Structures

Let \mathcal{T}_k be the class of finite structures whose Gaifman graphs have **treewidth** at most k .

Theorem (A., Dawar and Kolaitis 2004)

The homomorphism preservation property holds on \mathcal{T}_k .

More generally.

Theorem (A., Dawar and Kolaitis 2004)

The homomorphism preservation property holds on any class of finite structures whose Gaifman graphs **exclude some minor** and is closed under substructures and disjoint unions.

Preservation under Extensions?

Theorem (A., Dawar and Grohe 2005)

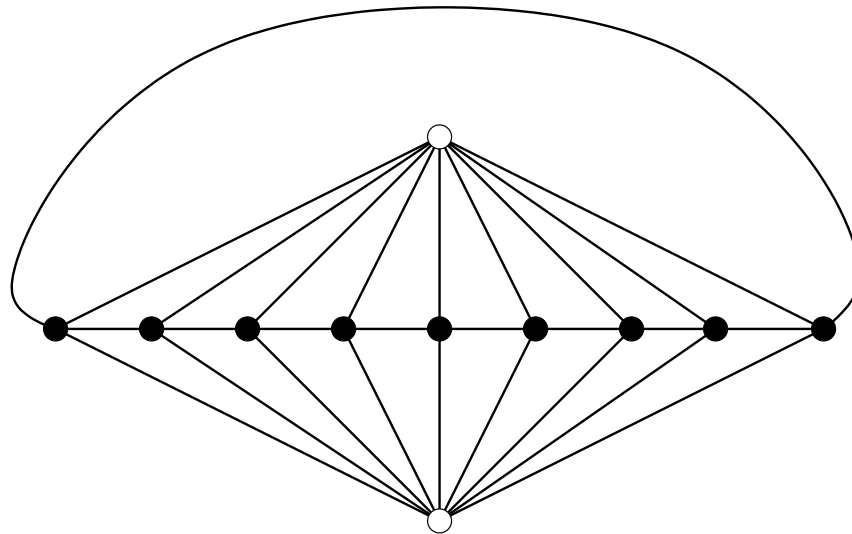
The extension preservation property holds on the following classes:

- graphs of **bounded degree**,
- **acyclic** structures,
- \mathcal{T}_k for every $k \geq 1$,

Remarkably, it **fails** for the class of **planar graphs**.

Counterexample on Planar Graphs

there are at least two different white points such that
either some point is not connected to both,
or every black point has exactly two black neighbors.



Conclusions

Conclusions

Homomorphisms are flexible enough to carry over [model-theoretic constructions](#), and in the context of CSPs, the finite model theory is even easier.

Open Question 1:

- Can we finitize Rossman's saturation using randomness?

Open Question 2:

- Does $\text{LFP} \cap \text{HOM} = \text{Datalog}$?
- Does $\text{LFP} \cap \text{co-CSP} = \text{Datalog}$?

It is known that:

- $\text{Datalog}(\neg, \neq) \cap \text{HOM} = \text{Datalog}$ (Feder-Vardi 2003)