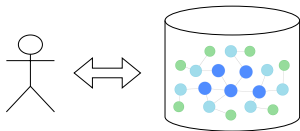


Query Algorithms for Relational Structures based on Homomorphism Counts

Balder ten Cate (FMT 2025)

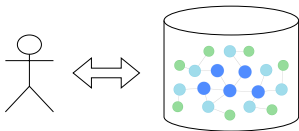
tC, Dalmau, Kolaitis, Wu (ICDT'24);
tC, Kolaitis, Kristjánsson (under submission)



Scenario: user wants to test if a (finite) structure satisfies a given property ...

- ▶ Is a given digraph **triangle-free**?
- ▶ Does a given digraph **contain a directed cycle**?

... by asking finitely many **Boolean CQs** $\exists \vec{x} \bigwedge_i \alpha_i$



Scenario: user wants to test if a (finite) structure satisfies a given property ...

- ▶ Is a given digraph **triangle-free**?
- ▶ Does a given digraph **contain a directed cycle**?

... by asking finitely many **Boolean CQs** $\exists \vec{x} \bigwedge_i \alpha_i$

- ▶ **Non-adaptive** (finite list of queries decided upfront) or
- ▶ **Adaptive** (each query may depend on answers to previous queries)

Non-adaptive case

	set semantics (true/false)	Bag semantics (# satisfying assgn)
triangle free?		
\exists directed cycle?		
\exists directed cycle of odd length?		
\exists directed cycle of length 2^{odd} ?		
≥ 3 edges?		

Non-adaptive case

	set semantics (true/false)	Bag semantics (# satisfying assgn)
triangle free?	Yes: suffices to ask $\exists xyz (Exy \wedge Eyz \wedge Ezx)$	
\exists directed cycle?		
\exists directed cycle of odd length?		
\exists directed cycle of length 2^{odd} ?		
≥ 3 edges?		

Non-adaptive case

	set semantics (true/false)	Bag semantics (# satisfying assgn)
triangle free?	Yes: suffices to ask $\exists xyz (Exy \wedge Eyz \wedge Ezx)$	Yes: same query
\exists directed cycle?		
\exists directed cycle of odd length?		
\exists directed cycle of length 2^{odd} ?		
≥ 3 edges?		

Non-adaptive case

	set semantics (true/false)	Bag semantics (# satisfying assgn)
triangle free?	Yes: suffices to ask $\exists xyz (Exy \wedge Eyz \wedge Ezx)$	Yes: same query
\exists directed cycle?	No: requires ∞ many queries (non-FO)	
\exists directed cycle of odd length?	No: requires ∞ many queries (non-FO)	
\exists directed cycle of length 2^{odd} ?	No: requires ∞ many queries (non-FO)	
≥ 3 edges?		

Non-adaptive case

	set semantics (true/false)	Bag semantics (# satisfying assgn)
triangle free?	Yes: suffices to ask $\exists xyz (Exy \wedge Eyz \wedge Ezx)$	Yes: same query
\exists directed cycle?	No: requires ∞ many queries (non-FO)	
\exists directed cycle of odd length?	No: requires ∞ many queries (non-FO)	
\exists directed cycle of length 2^{odd} ?	No: requires ∞ many queries (non-FO)	
≥ 3 edges?	No: not invariant for homomorphic equivalence	

Non-adaptive case

	set semantics (true/false)	Bag semantics (# satisfying assgn)
triangle free?	Yes: suffices to ask $\exists xyz (Exy \wedge Eyz \wedge Ezx)$	Yes: same query
\exists directed cycle?	No: requires ∞ many queries (non-FO)	
\exists directed cycle of odd length?	No: requires ∞ many queries (non-FO)	
\exists directed cycle of length 2^{odd} ?	No: requires ∞ many queries (non-FO)	
≥ 3 edges?	No: not invariant for homomorphic equivalence	Yes: suffices to ask $\exists xy (Exy)$

Non-adaptive case

	set semantics (true/false)	Bag semantics (# satisfying assgn)
triangle free?	Yes: suffices to ask $\exists xyz (Exy \wedge Eyz \wedge Ezx)$	Yes: same query
\exists directed cycle?	No: requires ∞ many queries (non-FO)	No: requires ∞ many queries
\exists directed cycle of odd length?	No: requires ∞ many queries (non-FO)	No: requires ∞ many queries
\exists directed cycle of length 2^{odd} ?	No: requires ∞ many queries (non-FO)	No: requires ∞ many queries
≥ 3 edges?	No: not invariant for homomorphic equivalence	Yes: suffices to ask $\exists xy (Exy)$

Homomorphism Counts

Let q be a Boolean CQ and A_q its canonical structure.

- ▶ The answer to $q(A)$ under the bag semantics is simply the number of homomorphisms $h : A_q \rightarrow A$ (where the counting is done in the semiring \mathbb{N} of natural numbers);
- ▶ The answer to $q(A)$ under the Boolean semantics is simply the number of homomorphisms $h : A_q \rightarrow A$ where the counting is done in the Boolean semiring $\mathbb{B} = \{0, 1\}$.

Therefore, we can cast the previous questions purely in terms of homomorphism counts.

Homomorphism Counts

Let q be a Boolean CQ and A_q its canonical structure.

- ▶ The answer to $q(A)$ under the bag semantics is simply the number of homomorphisms $h : A_q \rightarrow A$ (where the counting is done in the semiring \mathbb{N} of natural numbers);
- ▶ The answer to $q(A)$ under the Boolean semantics is simply the number of homomorphisms $h : A_q \rightarrow A$ where the counting is done in the Boolean semiring $\mathbb{B} = \{0, 1\}$.

Therefore, we can cast the previous questions purely in terms of homomorphism counts.

We will use the notation $\text{hom}_{\mathbb{N}}(A, B)$, and $\text{hom}_{\mathbb{B}}(A, B)$. We will also write $\text{hom}(A, B)$ for $\text{hom}_{\mathbb{N}}(A, B)$.

- ▶ $\text{hom}(K_3, G)$ = the number of triangles in graph G (times 6).
- ▶ $\text{hom}(G, K_3)$ = the number of 3-colorings of graph G .

Homomorphism Count Profiles

Fix an enumeration $All = \{A_1, A_2, \dots\}$ of all (finite) structures.

Definition: Let A be a finite structure

- ▶ The **left profile** of A is the vector
 $hom(All, A) := (hom(A_1, A), hom(A_2, A), \dots)$.
- ▶ The **right profile** of A is the vector
 $hom(A, All) := (hom(A, A_1), hom(A, A_2), \dots)$.

Left/Right Profiles and Isomorphism

Lovász's Theorem (1967):

For all structures A and B :

A and B are isomorphic iff $\text{hom}(\text{All}, A) = \text{hom}(\text{All}, B)$.

Chaudhuri-Vardi Theorem (1993):

For all structures A and B :

A and B are isomorphic iff $\text{hom}(A, \text{All}) = \text{hom}(B, \text{All})$.

Restricted Profiles

Definition:

Let $\mathcal{C} = \{B_1, B_2, \dots\}$ be a set of structures.

- ▶ The **left profile of A restricted to \mathcal{C}** is the vector $\text{hom}(\mathcal{C}, A) := (\text{hom}(B_1, A), \text{hom}(B_2, A), \dots)$
- ▶ The **right profile of A restricted to \mathcal{C}** is the vector $\text{hom}(A, \mathcal{C}) := (\text{hom}(A, B_1), \text{hom}(A, B_2), \dots)$

Each class \mathcal{C} gives rise to two equivalence relations:

- ▶ $A \equiv_{\mathcal{C}}^L B$ if A and B have the same left profile restricted to \mathcal{C} .
- ▶ $A \equiv_{\mathcal{C}}^R B$ if A and B have the same right profile restricted to \mathcal{C} .

Restricted Homomorphism Count Profiles for Graphs

	left profile restricted to ..	right profile restricted to ..
isomorphism	all graphs	all graphs
fractional isomorphism	all trees	
C^k -equivalence ($k \geq 2$)	graphs of $tw < k$	
quantum isomorphism	planar graphs	
\vdots		
chromatic equivalence		all cliques

Dvořák (2010)

Dell, Grohe and Rattan (2018)

Mančinska and Roberson (2020)

Atserias, Kolaitis, and Wu (2021)

...

What about **finite restrictions** of the left (or right) profile?

Query Algorithms (Chen, Flum, Liu, and Xun, 2022)

A **left query algorithm over \mathbb{N}** for a class \mathcal{C} is an algorithm that determines whether $A \in \mathcal{C}$ by asking finitely many homomorphism count queries $hom_{\mathbb{N}}(B, A)$.

Right query algorithms and **query algorithms over \mathbb{B}** are defined analogously.

The algorithm is **non-adaptive** if the queries are predetermined (i -th query does not depend on the answer to earlier queries).

Fact: The following are equivalent:

1. \mathcal{C} admits a **non-adaptive left query algorithm over \mathbb{N}** .
2. There is a finite set \mathcal{G} such that $hom_{\mathbb{N}}(\mathcal{G}, G)$ determines whether $G \in \mathcal{C}$.

Query Algorithms

Chen et al. 2022 proposed and studied **left/right query algorithms over \mathbb{N}** for **(undirected) graphs**.

Theorem:

- ▶ Every class of graphs definable by a **Boolean combination of universal FO-sentences** admits a **non-adaptive left query algorithm over \mathbb{N}** .
- ▶ The class of graphs satisfying $\exists x \forall y \neg E_{xy}$ does not.

Theorem:

- ▶ Every class of graphs **admits an adaptive left 3-query algorithm**.
- ▶ The same does not hold for right query algorithms. Indeed, some classes of graphs **do not admits an adaptive right k -query algorithm** for any k .

Non-Adaptive Query Algorithms over \mathbb{B} vs \mathbb{N}

tC, Dalmau, Kolaitis Wu (2024) studied **non-adaptive query algorithms** for **finite structures** over \mathbb{N} and \mathbb{B} .

Non-Adaptive Query Algorithms over \mathbb{B} vs \mathbb{N}

tC, Dalmau, Kolaitis Wu (2024) studied **non-adaptive query algorithms** for **finite structures** over \mathbb{N} and \mathbb{B} .

Fact: Let \mathcal{C} be a class of database instances. TFAE:

1. \mathcal{C} admits a non-adaptive left query algorithm over \mathbb{B} .
2. \mathcal{C} is definable by a Boolean combination of CQs
(in particular, \mathcal{C} is closed under homomorphic equivalence)

Non-Adaptive Query Algorithms over \mathbb{B} vs \mathbb{N}

tC, Dalmau, Kolaitis Wu (2024) studied **non-adaptive query algorithms** for **finite structures** over \mathbb{N} and \mathbb{B} .

Fact: Let \mathcal{C} be a class of database instances. TFAE:

1. \mathcal{C} admits a non-adaptive left query algorithm over \mathbb{B} .
2. \mathcal{C} is definable by a Boolean combination of CQs
(in particular, \mathcal{C} is closed under homomorphic equivalence)

Non-Adaptive Query Algorithms over \mathbb{B} vs \mathbb{N}

tC, Dalmau, Kolaitis Wu (2024) studied **non-adaptive query algorithms** for **finite structures** over \mathbb{N} and \mathbb{B} .

Fact: Let \mathcal{C} be a class of database instances. TFAE:

1. \mathcal{C} admits a non-adaptive left query algorithm over \mathbb{B} .
2. \mathcal{C} is definable by a Boolean combination of CQs
(in particular, \mathcal{C} is closed under homomorphic equivalence)

Theorem: For \mathcal{C} closed under homomorphism equivalence, TFAE:

1. \mathcal{C} admits a non-adaptive left query algorithm over \mathbb{N} .
2. \mathcal{C} admits a non-adaptive left query algorithm over \mathbb{B} .
3. \mathcal{C} is FO-definable (Rossman).

Non-adaptive case

	set semantics (true/false)	Bag semantics (# satisfying assgn)
triangle free?	Yes: suffices to ask $\exists xyz (Exy \wedge Eyz \wedge Ezx)$	Yes: same query
\exists directed cycle?	No: requires ∞ many queries (non-FO)	No: requires ∞ many queries
\exists directed cycle of odd length?	No: requires ∞ many queries (non-FO)	No: requires ∞ many queries
\exists directed cycle of length 2^{odd} ?	No: requires ∞ many queries (non-FO)	No: requires ∞ many queries
≥ 3 edges?	No: not invariant for homomorphic equivalence	Yes: suffices to ask $\exists xy (Exy)$

Adaptive Query Algorithms for Structures

tC, Kolaitis, Kristjánsson (2025): [adaptive query algorithms](#) for [finite structures](#) over \mathbb{N} and \mathbb{B} .

Adaptive Query Algorithms for Structures

tC, Kolaitis, Kristjánsson (2025): adaptive query algorithms for finite structures over \mathbb{N} and \mathbb{B} .

Examples:

- ▶ Every class admits an adaptive left query algorithm over \mathbb{N} .
- ▶ the class of digraphs containing a directed cycle admits an adaptive left query algorithm over \mathbb{B} .

Adaptive Query Algorithms for Structures

tC, Kolaitis, Kristjánsson (2025): **adaptive query algorithms** for **finite structures** over \mathbb{N} and \mathbb{B} .

Examples:

- ▶ Every class admits an **adaptive left query algorithm** over \mathbb{N} .
- ▶ the class of **digraphs containing a directed cycle** admits an **adaptive left query algorithm** over \mathbb{B} .

Results include:

- ▶ Not every class of structures admits an **adaptive k -query algorithm** over \mathbb{N} . However, every class of structures admits an **adaptive right 2-query algorithm** over \mathbb{N} .
- ▶ A finite-domain CSP admits an **adaptive left query algorithm** over \mathbb{B} iff it is FO.
- ▶ A Datalog program admits an **adaptive left query algorithm** over \mathbb{B} iff it is the intersection of its upper envelopes.

Adaptive case

	set semantics (true/false)	bag semantics (# satisfying assgn)
triangle free?		
\exists directed cycle?		
\exists directed cycle of odd length?		
\exists directed cycle of length 2^{odd} ?		
≥ 3 edges?		

Adaptive case

	set semantics (true/false)	bag semantics (# satisfying assign)
triangle free?	Yes	Yes
\exists directed cycle?	Yes	Yes
\exists directed cycle of odd length?	No (complement of a non-FO CSP)	Yes
\exists directed cycle of length 2^{odd} ?	No	Yes (but $\log \log n$ adaptive q's needed)
≥ 3 edges?	No (not invariant for homomorphic equivalence)	Yes

Synopsis

- ▶ Homomorphism count vectors capture interesting relaxations of isomorphism.
- ▶ “Query algorithms” tests for membership in a class of structures using *finitely many* homomorphism count queries.
- ▶ We studied the relative power of adaptive/non-adaptive left/right bounded/unbounded query algorithms over \mathbb{B}/\mathbb{N}

Some open question:

- ▶ Gap between $\log \log n$ and $\exp(n)$ bounds on the number of queries needed by adaptive left query algorithms over \mathbb{N}
- ▶ Bielecki and Van Den Bussche (ICDT'03): more general database interrogation scenario where the user may ask the *number of answers* to a (not necessarily Boolean) CQ.