Valued Constraint Satisfaction Problem and Resilience in Database Theory

Žaneta Semanišinová
joint work with Manuel Bodirsky and Carsten Lutz

Institute of Algebra
TU Dresden

LICS
11 Jul 2024

ERC Synergy Grant POCOCOP (GA 101071674)

Resilience of queries

Database: a relational structure $\mathfrak{A}$

Conjunctive query: a formula q of the form $\exists y_1, \ldots, y_l \, (\psi_1 \wedge \cdots \wedge \psi_m)$, where ψ_i are atomic

Resilience of queries

Database: a relational structure $\mathfrak{A}$

Conjunctive query: a formula q of the form $\exists y_1, \ldots, y_l \, (\psi_1 \wedge \cdots \wedge \psi_m)$, where ψ_i are atomic

Definition (Resilience)

Fixed conjunctive query q.

Input: a finite database $\mathfrak{A}$, $u \in \mathbb{N}$

Output: Can we remove $\leq u$ tuples from relations of $\mathfrak{A}$ so that $\mathfrak{A} \not\models q$?

Appears first in the paper of Meliou, Gatterbauer, Moore, Suciu ('10).

Resilience of queries

Database: a relational structure $\mathfrak{A}$
Conjunctive query: a formula q of the form $\exists y_1, \ldots, y_l \, (\psi_1 \wedge \cdots \wedge \psi_m)$, where ψ_i are atomic

Definition (Resilience)

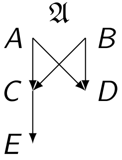
Fixed conjunctive query q.
Input: a finite database $\mathfrak{A}$, $u \in \mathbb{N}$
Output: Can we remove $\leq u$ tuples from relations of $\mathfrak{A}$ so that $\mathfrak{A} \not\models q$?

Appears first in the paper of Meliou, Gatterbauer, Moore, Suciu ('10).

Example: The resilience of

$$q = \exists x, y, z(R(x, y) \wedge R(y, z))$$

with respect to $\mathfrak{A}$ is 1 – remove (C, E).



Resilience of queries

Database: a relational structure $\mathfrak{A}$
Conjunctive query: a formula q of the form $\exists y_1, \ldots, y_l \, (\psi_1 \wedge \cdots \wedge \psi_m)$, where ψ_i are atomic

Definition (Resilience)

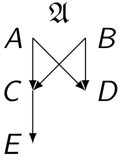
Fixed conjunctive query q.
Input: a finite database $\mathfrak{A}$, $u \in \mathbb{N}$
Output: Can we remove $\leq u$ tuples from relations of $\mathfrak{A}$ so that $\mathfrak{A} \not\models q$?

Appears first in the paper of Meliou, Gatterbauer, Moore, Suciu ('10).

Example: The resilience of

$$q = \exists x, y, z(R(x, y) \wedge R(y, z))$$

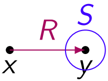
with respect to $\mathfrak{A}$ is 1 – remove (C, E).

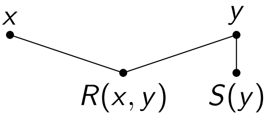


Goal: Classify complexity of resilience for all q.

Translation to a dual problem

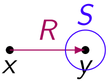
Example:
$q := \exists x, y(R(x, y) \wedge S(y))$



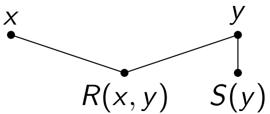


Translation to a dual problem

Example:
$q := \exists x, y(R(x,y) \wedge S(y))$



canonical structure



incidence graph I(q)

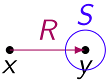
Theorem (Cherlin, Shelah, Shi '99)

Let q be a query and $\mathfrak{Q}$ its canonical structure. If $I(q)$ is connected, then there exists a structure $\mathfrak{B}_q$, such that for every finite $\mathfrak{A}$:

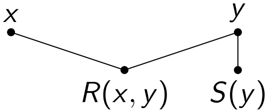
$$\mathfrak{A} \not\models q \Leftrightarrow \mathfrak{Q} \not\to \mathfrak{A} \Leftrightarrow \mathfrak{A} \to \mathfrak{B}_q$$

Translation to a dual problem

Example:
$q := \exists x, y(R(x,y) \wedge S(y))$



canonical structure



incidence graph I(q)

Theorem (Cherlin, Shelah, Shi '99)

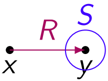
Let q be a query and $\mathfrak{Q}$ its canonical structure. If $I(q)$ is connected, then there exists a structure $\mathfrak{B}_q$, such that for every finite $\mathfrak{A}$:

$$\mathfrak{A} \not\models q \Leftrightarrow \mathfrak{Q} \not\to \mathfrak{A} \Leftrightarrow \mathfrak{A} \to \mathfrak{B}_q$$

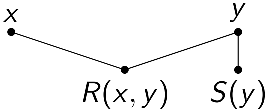
- $\mathfrak{B}_q$ *can be chosen so that* $\mathrm{Aut}(\mathfrak{B}_q)$ *is oligomorphic.*

Translation to a dual problem

Example:
$q := \exists x, y(R(x, y) \wedge S(y))$



canonical structure



incidence graph I(q)

Theorem (Cherlin, Shelah, Shi '99)

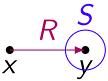
Let q be a query and $\mathfrak{Q}$ its canonical structure. If $I(q)$ is connected, then there exists a structure $\mathfrak{B}_q$, such that for every finite $\mathfrak{A}$:

$$\mathfrak{A} \not\models q \Leftrightarrow \mathfrak{Q} \not\to \mathfrak{A} \Leftrightarrow \mathfrak{A} \to \mathfrak{B}_q$$

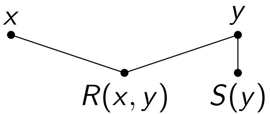
- *$\mathfrak{B}_q$ can be chosen so that $\mathrm{Aut}(\mathfrak{B}_q)$ is oligomorphic.*
- *$\mathfrak{B}_q$ can be chosen finite iff $I(q)$ is a tree. (Nešetřil, Tardiff '00; Larose, Loten, Tardiff '07)*

Translation to a dual problem

Example:
$q := \exists x, y(R(x, y) \wedge S(y))$



canonical structure



incidence graph I(q)

Theorem (Cherlin, Shelah, Shi '99)

Let q be a query and $\mathfrak{Q}$ its canonical structure. If $I(q)$ is connected, then there exists a structure $\mathfrak{B}_q$, such that for every finite $\mathfrak{A}$:

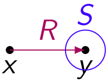
$$\mathfrak{A} \not\models q \Leftrightarrow \mathfrak{Q} \not\to \mathfrak{A} \Leftrightarrow \mathfrak{A} \to \mathfrak{B}_q$$

- $\mathfrak{B}_q$ *can be chosen so that* $\mathrm{Aut}(\mathfrak{B}_q)$ *is oligomorphic.*
- $\mathfrak{B}_q$ *can be chosen finite iff $I(q)$ is a tree. (Nešetřil, Tardiff '00; Larose, Loten, Tardiff '07)*

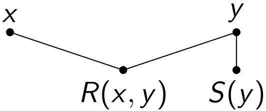
oligomorphic – countable domain B_q and the action of $\mathrm{Aut}(\mathfrak{B}_q)$ on B_q^n has finitely many orbits for every $n \geq 1$

Translation to a dual problem

Example:
$q := \exists x, y(R(x,y) \wedge S(y))$



canonical structure



incidence graph I(q)

Theorem (Cherlin, Shelah, Shi '99)

Let q be a query and $\mathfrak{Q}$ its canonical structure. If $I(q)$ is connected, then there exists a structure $\mathfrak{B}_q$, such that for every finite $\mathfrak{A}$:

$$\mathfrak{A} \not\models q \Leftrightarrow \mathfrak{Q} \not\to \mathfrak{A} \Leftrightarrow \mathfrak{A} \to \mathfrak{B}_q$$

- *$\mathfrak{B}_q$ can be chosen so that $\mathrm{Aut}(\mathfrak{B}_q)$ is oligomorphic.*
- *$\mathfrak{B}_q$ can be chosen finite iff $I(q)$ is a tree. (Nešetřil, Tardiff '00; Larose, Loten, Tardiff '07)*

Example: For every finite directed graph G we have:

$$\uparrow\!\uparrow \not\to G \Leftrightarrow G \to \uparrow$$

Valued Constraint Satisfaction Problem

CSP – satisfiability of a conjunction of atomic formulas
$\hookrightarrow$ generalize by giving values to constraints

Valued Constraint Satisfaction Problem

CSP – satisfiability of a conjunction of atomic formulas
$\hookrightarrow$ generalize by giving values to constraints

A valued structure Γ consists of:

- (countable) domain D
- (finite, relational) signature τ
- for each $R \in \tau$ of arity k, a function $R^\Gamma\colon D^k \to \mathbb{Q} \cup \{\infty\}$

Valued Constraint Satisfaction Problem

CSP – satisfiability of a conjunction of atomic formulas
$\hookrightarrow$ generalize by giving values to constraints

A valued structure Γ consists of:

- (countable) domain D
- (finite, relational) signature τ
- for each $R \in \tau$ of arity k, a function $R^\Gamma \colon D^k \to \mathbb{Q} \cup \{\infty\}$

Definition (VCSP(Γ))

Input: $u \in \mathbb{Q}$, an expression
$$\phi(x_1, \dots, x_n) = \sum_i \psi_i,$$
where each ψ_i is an atomic τ-formula
Output: Is
$$\inf_{\bar{a} \in D^n} \phi(\bar{a}) \leq u \text{ in } \Gamma?$$

Connection of resilience and VCSPs

query q with $I(q)$ connected (WLOG) $\rightsquigarrow$ obtain the dual structure $\mathfrak{B}_q$ $\rightsquigarrow$ turn it into a valued structure Γ_q with cost functions taking values 0 and 1

Connection of resilience and VCSPs

query q with $I(q)$ connected (WLOG) $\rightsquigarrow$ obtain the dual structure $\mathfrak{B}_q$ $\rightsquigarrow$ turn it into a valued structure Γ_q with cost functions taking values 0 and 1

Theorem (Bodirsky, Lutz, S.)

The resilience problem for q equals $\mathrm{VCSP}(\Gamma_q)$.

Connection of resilience and VCSPs

query q with $I(q)$ connected (WLOG) $\rightsquigarrow$ obtain the dual structure $\mathfrak{B}_q$ $\rightsquigarrow$ turn it into a valued structure Γ_q with cost functions taking values 0 and 1

Theorem (Bodirsky, Lutz, S.)

The resilience problem for q equals $\mathrm{VCSP}(\Gamma_q)$.

Remark: We need to consider bag databases: tuples have multiplicity ≥ 1.

Connection of resilience and VCSPs

query q with $I(q)$ connected (WLOG) $\rightsquigarrow$ obtain the dual structure $\mathfrak{B}_q$ $\rightsquigarrow$ turn it into a valued structure Γ_q with cost functions taking values 0 and 1

Theorem (Bodirsky, Lutz, S.)

The resilience problem for q equals $\mathrm{VCSP}(\Gamma_q)$.

Remark: We need to consider bag databases: tuples have multiplicity ≥ 1.

Fact: If the domain of Γ is finite, VCSP(Γ) is in P or NP-complete.
[Kozik, Ochremiak ('15); Kolmogorov, Rolínek, Krokhin ('15); Bulatov ('17); Zhuk ('17)]

Connection of resilience and VCSPs

query q with $I(q)$ connected (WLOG) $\rightsquigarrow$ obtain the dual structure $\mathfrak{B}_q$ $\rightsquigarrow$ turn it into a valued structure Γ_q with cost functions taking values 0 and 1

Theorem (Bodirsky, Lutz, S.)

The resilience problem for q equals VCSP(Γ_q).

Remark: We need to consider bag databases: tuples have multiplicity ≥ 1.

Fact: If the domain of Γ is finite, VCSP(Γ) is in P or NP-complete.
[Kozik, Ochremiak ('15); Kolmogorov, Rolínek, Krokhin ('15); Bulatov ('17); Zhuk ('17)]

Corollary (Bodirsky, Lutz, S.)

Let q be a conjunctive query such that $I(q)$ is acyclic. Then the resilience problem for q is in P or NP-complete.

Results on VCSPs and resilience

Main focus: develop the theory of infinite-domain VCSPs

Results on VCSPs and resilience

Main focus: develop the theory of infinite-domain VCSPs

- algebraic condition based on fractional polymorphisms sufficient for tractability of VCSP(Γ)

Results on VCSPs and resilience

Main focus: develop the theory of infinite-domain VCSPs

- algebraic condition based on fractional polymorphisms sufficient for tractability of VCSP(Γ)
- algebraic condition based on pp-constructability sufficient for NP-hardness of VCSP(Γ)

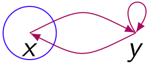
Results on VCSPs and resilience

Main focus: develop the theory of infinite-domain VCSPs

- algebraic condition based on fractional polymorphisms sufficient for tractability of VCSP(Γ)
- algebraic condition based on pp-constructability sufficient for NP-hardness of VCSP(Γ)

$\hookrightarrow$ answers open question [Freire, Gatterbauer, Immerman, Meliou '20] about the resilience of

$$q := \exists x, y \big(S(x) \wedge R(x, y) \wedge R(y, x) \wedge R(y, y) \big)$$



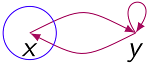
Results on VCSPs and resilience

Main focus: develop the theory of infinite-domain VCSPs

- algebraic condition based on fractional polymorphisms sufficient for tractability of VCSP(Γ)
- algebraic condition based on pp-constructability sufficient for NP-hardness of VCSP(Γ)

↪ answers open question [Freire, Gatterbauer, Immerman, Meliou '20] about the resilience of

$$q := \exists x, y \big(S(x) \wedge R(x, y) \wedge R(y, x) \wedge R(y, y) \big)$$



Conjecture: For a query q, whenever Γ_q does not satisfy the hardness condition, it satisfies the tractability condition.

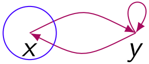
Results on VCSPs and resilience

Main focus: develop the theory of infinite-domain VCSPs

- algebraic condition based on fractional polymorphisms sufficient for tractability of VCSP(Γ)
- algebraic condition based on pp-constructability sufficient for NP-hardness of VCSP(Γ)

$\hookrightarrow$ answers open question [Freire, Gatterbauer, Immerman, Meliou '20] about the resilience of

$$q := \exists x, y\big(S(x) \wedge R(x,y) \wedge R(y,x) \wedge R(y,y)\big)$$



Conjecture: For a query q, whenever Γ_q does not satisfy the hardness condition, it satisfies the tractability condition.

- the conjecture is true for every Γ_q on a finite domain (derived from [Kozik, Ochremiak '15] and [Kolmogorov, Krokhin, Rolínek '15])

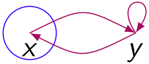
Results on VCSPs and resilience

Main focus: develop the theory of infinite-domain VCSPs

- algebraic condition based on fractional polymorphisms sufficient for tractability of VCSP(Γ)
- algebraic condition based on pp-constructability sufficient for NP-hardness of VCSP(Γ)

$\hookrightarrow$ answers open question [Freire, Gatterbauer, Immerman, Meliou '20] about the resilience of

$$q := \exists x, y\big(S(x) \wedge R(x,y) \wedge R(y,x) \wedge R(y,y)\big)$$



Conjecture: For a query q, whenever Γ_q does not satisfy the hardness condition, it satisfies the tractability condition.

- the conjecture is true for every Γ_q on a finite domain (derived from [Kozik, Ochremiak '15] and [Kolmogorov, Krokhin, Rolínek '15])
- equivalently, for every query q such that $I(q)$ is a tree

Thank you for your attention

Funding statement: Funded by the European Union (ERC, POCOCOP, 101071674).
Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.