

Valued Constraint Satisfaction Problem and Resilience in Database Theory

Žaneta Semanišínová
joint work with Manuel Bodirsky and Carsten Lutz

Institute of Algebra
TU Dresden

Combinatorial Problems in Model Theory and Computer Science
7 Nov 2023



ERC Synergy Grant POCOCOP (GA 101071674)

Resilience of queries

Database: a relational structure \mathfrak{A}

Conjunctive query: a primitive positive formula q , i.e.

$\exists y_1, \dots, y_l (\psi_1 \wedge \dots \wedge \psi_m)$, where ψ_i are atomic

Resilience of queries

Database: a relational structure \mathcal{A}

Conjunctive query: a primitive positive formula q , i.e.

$\exists y_1, \dots, y_l (\psi_1 \wedge \dots \wedge \psi_m)$, where ψ_i are atomic

Definition (Resilience)

Fixed conjunctive query q .

Input: a finite database \mathcal{A}

Output: minimal number of tuples to be removed from relations of \mathcal{A} , so that $\mathcal{A} \not\models q$

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

Resilience of queries

Database: a relational structure \mathcal{A}

Conjunctive query: a primitive positive formula q , i.e.

$\exists y_1, \dots, y_l (\psi_1 \wedge \dots \wedge \psi_m)$, where ψ_i are atomic

Definition (Resilience)

Fixed conjunctive query q .

Input: a finite database \mathcal{A}

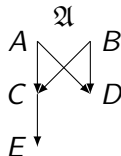
Output: minimal number of tuples to be removed from relations of \mathcal{A} , so that $\mathcal{A} \not\models q$

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 \mathcal{A} is 1 – remove (C, E) .



Resilience of queries

Database: a relational structure \mathfrak{A}

Conjunctive query: a primitive positive formula q , i.e.

$\exists y_1, \dots, y_l (\psi_1 \wedge \dots \wedge \psi_m)$, where ψ_i are atomic

Definition (Resilience)

Fixed conjunctive query q .

Input: a finite database \mathfrak{A}

Output: minimal number of tuples to be removed from relations of \mathfrak{A} , so that $\mathfrak{A} \not\models q$

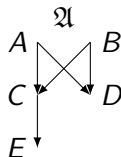
Appears first in the paper of Meliou, Gatterbauer, Moore, Suciu ('10).

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 .



Constraint satisfaction

Fixed τ -structure \mathfrak{A} (τ – finite relational signature)

Input: list of atomic τ -formulas (constraints)

Output:

- **CSP:** Decide whether there is a solution that satisfies **all** constraints.
- **MaxCSP:** Find the **maximal number** of constraints that can be satisfied at once.
- **VCSP:** Find the **minimal cost** with which the constraints can be satisfied (each constraint comes with a cost depending on the chosen values).

Constraint satisfaction

Fixed τ -structure \mathfrak{A} (τ – finite relational signature)

Input: list of atomic τ -formulas (constraints)

Output:

- **CSP:** Decide whether there is a solution that satisfies **all** constraints.
- **MaxCSP:** Find the **maximal number** of constraints that can be satisfied at once.
- **VCSP:** Find the **minimal cost** with which the constraints can be satisfied (each constraint comes with a cost depending on the chosen values).

Observation: VCSP **generalizes** CSP and MaxCSP.

Proof: Model the tuples in relations with cost 0 and outside with cost 1 (for MaxCSP) or ∞ (for CSP).

A **valued structure** Γ consists of:

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

A **valued structure** Γ consists of:

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

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

Question: Is

$$\inf_{\bar{a} \in D^n} \phi(\bar{a}) \leq u \text{ in } \Gamma?$$

Max-Cut as a VCSP

Example:

Input: $G = (V, E)$ – finite directed graph

Goal: Find a partition $A \cup B$ of V such that $E \cap (A \times B)$ is maximal.

Equivalently: $E \cap (A^2 \cup B^2 \cup B \times A)$ is minimal.

Example:

Input: $G = (V, E)$ – finite directed graph

Goal: Find a partition $A \cup B$ of V such that $E \cap (A \times B)$ is maximal.

Equivalently: $E \cap (A^2 \cup B^2 \cup B \times A)$ is minimal.

Let Γ_{MC} be a valued structure where:

- $D = \{0, 1\}$
- $\tau = \{R\}$, R binary

$$R(x, y) = \begin{cases} 0 & \text{if } x = 0 \text{ and } y = 1 \\ 1 & \text{otherwise} \end{cases}$$

Example:

Input: $G = (V, E)$ – finite directed graph

Goal: Find a partition $A \cup B$ of V such that $E \cap (A \times B)$ is maximal.

Equivalently: $E \cap (A^2 \cup B^2 \cup B \times A)$ is minimal.

Let Γ_{MC} be a valued structure where:

- $D = \{0, 1\}$
- $\tau = \{R\}$, R binary

$$R(x, y) = \begin{cases} 0 & \text{if } x = 0 \text{ and } y = 1 \\ 1 & \text{otherwise} \end{cases}$$

Take vertices of G as variables. The **size of a maximal cut** of G is

$$\min_{\bar{x} \in D^n} \sum_{(x_i, x_j) \in E} R(x_i, x_j). \text{ The partition of } V \text{ is given by the values 0 and 1.}$$

Example:

Input: $G = (V, E)$ – finite directed graph

Goal: Find a partition $A \cup B$ of V such that $E \cap (A \times B)$ is maximal.

Equivalently: $E \cap (A^2 \cup B^2 \cup B \times A)$ is minimal.

Let Γ_{MC} be a valued structure where:

- $D = \{0, 1\}$
- $\tau = \{R\}$, R binary

$$R(x, y) = \begin{cases} 0 & \text{if } x = 0 \text{ and } y = 1 \\ 1 & \text{otherwise} \end{cases}$$

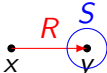
Take vertices of G as variables. The **size of a maximal cut** of G is

$$\min_{\bar{x} \in D^n} \sum_{(x_i, x_j) \in E} R(x_i, x_j). \text{ The partition of } V \text{ is given by the values 0 and 1.}$$

every instance of $VCSP(\Gamma_{MC})$ corresponds to a **digraph**

$\rightsquigarrow VCSP(\Gamma_{MC})$ is the **Max-Cut** problem (NP-hard)

Homomorphism duality

Example (canonical structure): $\exists x, y(R(x, y) \wedge S(y)) \rightsquigarrow$ 

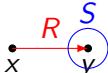
For a query q , take its canonical structure Ω .

Search for a structure \mathfrak{B}_q such that for every finite \mathfrak{A} :

$$\mathfrak{A} \not\models q \Leftrightarrow \Omega \not\rightarrow \mathfrak{A} \Leftrightarrow \mathfrak{A} \rightarrow \mathfrak{B}_q$$

\rightsquigarrow corresponds to $\text{CSP}(\mathfrak{B}_q)$ (if we represent the constraints by their canonical structure)

Homomorphism duality

Example (canonical structure): $\exists x, y(R(x, y) \wedge S(y)) \rightsquigarrow$ 

For a **query** q , take its **canonical structure** Ω .

Search for a structure \mathfrak{B}_q such that for **every finite** \mathfrak{A} :

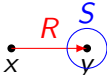
$$\mathfrak{A} \not\models q \Leftrightarrow \Omega \not\rightarrow \mathfrak{A} \Leftrightarrow \mathfrak{A} \rightarrow \mathfrak{B}_q$$

\rightsquigarrow corresponds to $\text{CSP}(\mathfrak{B}_q)$ (if we represent the constraints by their canonical structure)

Example: For every finite directed graph G we have:

$$\begin{array}{c} \uparrow \\ \uparrow \not\rightarrow G \Leftrightarrow G \rightarrow \uparrow \end{array}$$

Homomorphism duality

Example (canonical structure): $\exists x, y(R(x, y) \wedge S(y)) \rightsquigarrow$ 

For a **query** q , take its **canonical structure** Ω .

Search for a structure \mathfrak{B}_q such that for **every finite** \mathfrak{A} :

$$\mathfrak{A} \not\models q \Leftrightarrow \Omega \not\rightarrow \mathfrak{A} \Leftrightarrow \mathfrak{A} \rightarrow \mathfrak{B}_q$$

\rightsquigarrow corresponds to **CSP**(\mathfrak{B}_q) (if we represent the constraints by their canonical structure)

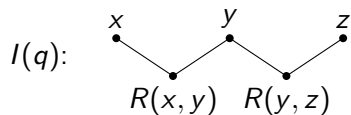
Example: For every finite directed graph G we have:

$$\uparrow \not\rightarrow G \Leftrightarrow G \rightarrow \uparrow$$

\rightsquigarrow existence of \mathfrak{B}_q enables studying **resilience** of q using the results about **(valued) constraint satisfaction problems**

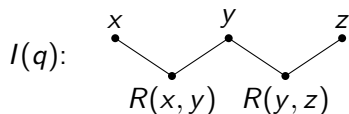
Existence of dual structures

Example (incidence graph): $q := \exists x, y, z(R(x, y) \wedge R(y, z))$



Existence of dual structures

Example (incidence graph): $q := \exists x, y, z(R(x, y) \wedge R(y, z))$

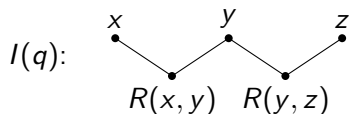


Theorem (Nešetřil, Tardiff ('00); Larose, Loten, Tardiff ('07))

A conjunctive query q has a *finite dual* if and only if $I(q)$ is a *tree*.

Existence of dual structures

Example (incidence graph): $q := \exists x, y, z(R(x, y) \wedge R(y, z))$



Theorem (Nešetřil, Tardiff ('00); Larose, Loten, Tardiff ('07))

A conjunctive query q has a *finite dual* if and only if $I(q)$ is a *tree*.

Theorem (Cherlin, Shelah, Shi ('99))

If $I(q)$ is *connected*, then q has a countable dual \mathfrak{B}_q , which can be chosen so that $\text{Aut}(\mathfrak{B}_q)$ is *oligomorphic*.

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

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

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).

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 $\text{VCSP}(\Gamma_q)$.

Remark: We have to consider bag databases – a database \mathfrak{A} might contain a tuple with multiplicity > 1 (differs from the original setting).

Example: Input $R(x, y) + R(x, y)$ for $\text{VCSP}(\Gamma)$ corresponds to a database with multiplicity 2 for $R(x, y)$.

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 $\text{VCSP}(\Gamma_q)$.

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

For every finite G :

$$\Omega = \begin{array}{c} \uparrow \\ \vdash \\ \uparrow \end{array} \not\rightarrow G \Leftrightarrow G \rightarrow \begin{array}{c} \uparrow \\ \vdash \\ \uparrow \end{array} = \mathfrak{B}_q$$

$\mathfrak{B}_q \rightsquigarrow \Gamma_{\text{MC}} = (\{0, 1\}; R)$

Resilience of $q = \text{VCSP}(\Gamma_{\text{MC}}) = \text{Max-Cut}$ is NP-hard

Complexity dichotomy for resilience of acyclic queries

Corollary: $I(q)$ is a tree $\Rightarrow \Gamma_q$ can be taken finite

Complexity dichotomy for resilience of acyclic queries

Corollary: $I(q)$ is a tree $\Rightarrow \Gamma_q$ can be taken finite

Follows from results of: Kozik, Ochremiak ('15); Kolmogorov, Rolínek, Krokhin ('15); Bulatov ('17); Zhuk ('17)

Theorem

For a *finite-domain* valued structure Γ , $\text{VCSP}(\Gamma)$ is in P or NP -complete.

Complexity dichotomy for resilience of acyclic queries

Corollary: $I(q)$ is a tree $\Rightarrow \Gamma_q$ can be taken finite

Follows from results of: Kozik, Ochremiak ('15); Kolmogorov, Rolínek, Krokhn ('15); Bulatov ('17); Zhuk ('17)

Theorem

For a *finite-domain* valued structure Γ , $\text{VCSP}(\Gamma)$ is in P or NP -complete.

Corollary (Bodirsky, Lutz, S.)

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

Complexity dichotomy for resilience of acyclic queries

Corollary: $I(q)$ is a tree $\Rightarrow \Gamma_q$ can be taken finite

Follows from results of: Kozik, Ochremiak ('15); Kolmogorov, Rolínek, Krokhin ('15); Bulatov ('17); Zhuk ('17)

Theorem

For a *finite-domain* valued structure Γ , $\text{VCSP}(\Gamma)$ is in P or NP -complete.

Corollary (Bodirsky, Lutz, S.)

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

Generalisation: theorem holds for q with acyclic $I(q)$

Complexity dichotomy for resilience of acyclic queries

Corollary: $I(q)$ is a tree $\Rightarrow \Gamma_q$ can be taken finite

Follows from results of: Kozik, Ochremiak ('15); Kolmogorov, Rolínek, Krokhin ('15); Bulatov ('17); Zhuk ('17)

Theorem

For a *finite-domain* valued structure Γ , $\text{VCSP}(\Gamma)$ is in P or NP -complete.

Corollary (Bodirsky, Lutz, S.)

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

Generalisation: theorem holds for q with acyclic $I(q)$

Question: For general queries, choose \mathfrak{B}_q such that $\text{Aut}(\mathfrak{B}_q)$ is oligomorphic \Rightarrow finitely many orbits of n -tuples for every n .

Can we use some results for finite domains?

Hard resilience problems

pp-construction – a notion of ‘expressing’ one valued structure in another (generalizes pp-constructions for classical structures)

Fact: If $\text{Aut}(\Gamma)$ is **oligomorphic** and Γ **pp-constructs** Δ , then $\text{VCSP}(\Delta)$ **reduces** to $\text{VCSP}(\Gamma)$ in **poly-time**.

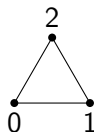
Hard resilience problems

pp-construction – a notion of ‘expressing’ one valued structure in another (generalizes pp-constructions for classical structures)

Fact: If $\text{Aut}(\Gamma)$ is **oligomorphic** and Γ **pp-constructs** Δ , then $\text{VCSP}(\Delta)$ **reduces** to $\text{VCSP}(\Gamma)$ in **poly-time**.

K_3 is the valued structure on $\{0, 1, 2\}$ with single binary relation E defined:

$$E(x, y) = \begin{cases} 0 & \text{if } x \neq y \\ \infty & \text{if } x = y \end{cases}$$



Observation: $\text{VCSP}(K_3)$ is the 3-colorability problem and hence NP-hard.

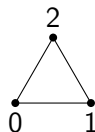
Hard resilience problems

pp-construction – a notion of ‘expressing’ one valued structure in another (generalizes pp-constructions for classical structures)

Fact: If $\text{Aut}(\Gamma)$ is **oligomorphic** and Γ **pp-constructs** Δ , then $\text{VCSP}(\Delta)$ **reduces** to $\text{VCSP}(\Gamma)$ in **poly-time**.

K_3 is the valued structure on $\{0, 1, 2\}$ with single binary relation E defined:

$$E(x, y) = \begin{cases} 0 & \text{if } x \neq y \\ \infty & \text{if } x = y \end{cases}$$



Observation: $\text{VCSP}(K_3)$ is the 3-colorability problem and hence NP-hard.

Corollary

If $\text{Aut}(\Gamma)$ is **oligomorphic** and Γ **pp-constructs** K_3 , then $\text{VCSP}(\Gamma)$ is **NP-hard**.

Fractional polymorphisms

polymorphism f of \mathfrak{B} – an operation $f : B^n \rightarrow B$ that preserves all relations of \mathfrak{B}

Idea: $\text{CSP}(\mathfrak{B})$ is tractable iff it has nice polymorphisms

Fractional polymorphisms

polymorphism f of \mathfrak{B} – an operation $f : B^n \rightarrow B$ that preserves all relations of \mathfrak{B}

Idea: $\text{CSP}(\mathfrak{B})$ is tractable iff it has nice polymorphisms

Definition (fractional polymorphism)

Γ – valued τ -structure with domain D

A **fractional polymorphism** of Γ of arity n is a **probability distribution** ω on operations $D^n \rightarrow D$ such that for every k -ary $R \in \tau$ and $a^1, \dots, a^n \in D^k$

$$\underbrace{E_{\omega}[f \mapsto R(f(a^1, \dots, a^n))]}_{\text{expected value}} \leq \underbrace{\frac{1}{n} \sum_{j=1}^n R(a^j)}_{\text{arithmetic mean}} .$$

Tractability conjecture

Known for finite-domain VCSPs:

Theorem

Γ – a *finite-domain* valued structure

- If Γ does not *pp-construct* K_3 , then Γ has *cyclic fractional polymorphism* (essentially Kozik, Ochremiak ('15)).
- If Γ has a *cyclic fractional polymorphism*, then $\text{VCSP}(\Gamma)$ is in P (Kolmogorov, Krokhin, Rolínek ('15)).

Tractability conjecture

Known for finite-domain VCSPs:

Theorem

Γ – a *finite-domain* valued structure

- If Γ does not *pp-construct* K_3 , then Γ has *cyclic fractional polymorphism* (essentially Kozik, Ochremiak ('15)).
- If Γ has a *cyclic fractional polymorphism*, then $\text{VCSP}(\Gamma)$ is in P (Kolmogorov, Krokhin, Rolínek ('15)).

Theorem (Bodirsky, Lutz, S.)

q – conjunctive query

If Γ_q has a *fractional polymorphism* which is *canonical* and *pseudo cyclic* with respect to $\text{Aut}(\Gamma_q)$, then $\text{VCSP}(\Gamma_q)$ is in P .

Tractability conjecture

Known for finite-domain VCSPs:

Theorem

Γ – a *finite-domain* valued structure

- If Γ does not pp-construct K_3 , then Γ has *cyclic fractional polymorphism* (essentially Kozik, Ochremiak ('15)).
- If Γ has a *cyclic fractional polymorphism*, then $\text{VCSP}(\Gamma)$ is in P (Kolmogorov, Krokhin, Rolínek ('15)).

Theorem (Bodirsky, Lutz, S.)

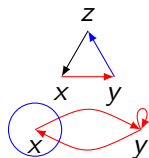
q – conjunctive query

If Γ_q has a *fractional polymorphism* which is *canonical* and *pseudo cyclic* with respect to $\text{Aut}(\Gamma_q)$, then $\text{VCSP}(\Gamma_q)$ is in P .

Conjecture: If Γ_q does not pp-construct K_3 , then the tractability theorem applies and $\text{VCSP}(\Gamma_q)$ and hence resilience of q is in P .

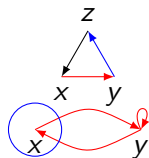
Examples

- $q_{MC} := \exists x, y, z (R(x, y) \wedge R(y, z))$
- $q_{path} := \exists x, y, z (R(x, y) \wedge S(y, z))$
- $q_{\Delta} := \exists x, y, z (R(x, y) \wedge S(y, z) \wedge T(z, x))$
- $q_{new} := \exists x, y (S(x) \wedge R(x, y) \wedge R(y, x) \wedge R(y, y))$



Examples

- $q_{MC} := \exists x, y, z (R(x, y) \wedge R(y, z))$
- $q_{path} := \exists x, y, z (R(x, y) \wedge S(y, z))$
- $q_{\Delta} := \exists x, y, z (R(x, y) \wedge S(y, z) \wedge T(z, x))$
- $q_{new} := \exists x, y (S(x) \wedge R(x, y) \wedge R(y, x) \wedge R(y, y))$



| query | complexity | VCSP proof |
|--------------|--------------------|---|
| q_{MC} | NP-hard (FGIM '20) | Max-Cut/pp-constructs K_3 |
| q_{path} | P (MG) | binary symmetric frac. polymorphism |
| q_{Δ} | NP-hard (FGIM '15) | pp-constructs K_3 |
| q_{new} | P (BLS) | binary pseudo cyclic frac. polymorphism |

References:

- Meliou, Gatterbauer, Moore, Suciu ('10)
- Freire, Gatterbauer, Immerman, Meliou ('15)
- Freire, Gatterbauer, Immerman, Meliou ('20)
- Makhija, Gatterbauer ('22)
- Bodirsky, Lutz, S.

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.