Valued Constraint Satisfaction Problem and Resilience in Database Theory

Žaneta Semanišinová
joint work with Manuel Bodirsky and Carsten Lutz

Institute of Algebra
TU Dresden

AAA103, 10.6.2023

ERC Synergy Grant POCOCOP (GA 101071674)

Resilience of queries

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

Conjunctive query: a primitive positive formula q, i.e.
$\exists y_1, \ldots, y_l \, (\psi_1 \wedge \cdots \wedge \psi_m)$, where ψ_i are atomic

Resilience of queries

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

Conjunctive query: a primitive positive formula q, i.e. $\exists y_1, \ldots, y_l\,(\psi_1 \wedge \cdots \wedge \psi_m)$, where ψ_i are atomic

Definition (Resilience)

Fixed conjunctive query q.

Input: a finite database $\mathfrak{A}$

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).

Resilience of queries

Database: a relational structure $\mathfrak{A}$
Conjunctive query: a primitive positive formula q, i.e.
$\exists y_1, \ldots, y_l \, (\psi_1 \wedge \cdots \wedge \psi_m)$, where ψ_i are atomic

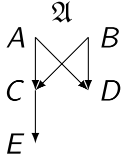
Definition (Resilience)

Fixed conjunctive query q.
Input: a finite database $\mathfrak{A}$
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

$$q = \exists x, y, z(\text{parent}(x, y) \wedge \text{parent}(y, z))$$



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

Might be considered with a threshold.

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).

Might be considered with a threshold.

Observation: VCSP generalizes CSP and MaxCSP.
Proof: Model the tuples in relations with cost 0 and outside with cost

- 1 and the same threshold (for MaxCSP);
- ∞ and threshold 0 (for CSP).

Focus on VCSP

A valued structure Γ consists of:

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

Focus on VCSP

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

$$\phi(x_1, \ldots, x_n) = \sum_i \psi_i,$$

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.

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.

Let Γ be a valued structure where:

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

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

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.
Let Γ be a valued structure where:

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

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

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

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

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.
Let Γ be a valued structure where:

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

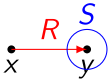
$$E(x, y) = \begin{cases} 0 \text{ if } x = 0 \text{ and } y = 1 \\ 1 \text{ otherwise} \end{cases}$$

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

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

Every instance of VCSP(Γ) corresponds to a digraph, hence VCSP(Γ) is the Max-Cut problem (NP-hard).

Homomorphism duality

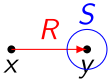
Example (canonical structure): $\exists x, y(R(x,y) \wedge S(y)) \rightsquigarrow$ 

For a query q, take its canonical structure $\mathfrak{Q}$.
Search for a structure $\mathfrak{B}_q$ such that for every finite $\mathfrak{A}$:

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

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

Homomorphism duality

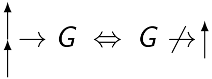
Example (canonical structure): $\exists x, y(R(x,y) \wedge S(y)) \rightsquigarrow$ 

For a query q, take its canonical structure $\mathfrak{Q}$.
Search for a structure $\mathfrak{B}_q$ such that for every finite $\mathfrak{A}$:

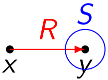
$$\mathfrak{A} \not\models q \Leftrightarrow \mathfrak{Q} \not\to \mathfrak{A} \Leftrightarrow \mathfrak{A} \to \mathfrak{B}_q$$

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

Example: For every finite directed graph G we have:



Homomorphism duality

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

For a query q, take its canonical structure $\mathfrak{Q}$.
Search for a structure $\mathfrak{B}_q$ such that for every finite $\mathfrak{A}$:

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

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

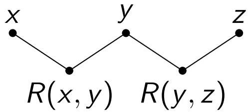
Example: For every finite directed graph G we have:

 $\to G \Leftrightarrow G \not\to$ 

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

Existence of dual structures

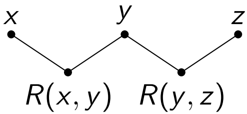
Example (incidence graph): $q := \exists x, y, z(R(x, y) \wedge R(y, z))$

$I(q)$: 

Existence of dual structures

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

$I(q)$:



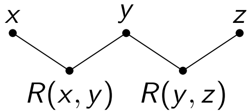
Theorem (Nešetřil,Tardiff ('00); Larose, Loten, Tardiff ('07))

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

Existence of dual structures

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

$I(q)$:



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 $\mathrm{Aut}(\mathfrak{B}_q)$ is oligomorphic.

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

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 VCSP(Γ_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 VCSP(Γ) 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 $\mathrm{VCSP}(\Gamma_q)$.

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 $\mathrm{VCSP}(\Gamma)$ corresponds to a database with multiplicity 2 for $R(x, y)$.

Corollary: $I(q)$ is a tree $\Rightarrow$ Γ_q can be taken finite
$\rightsquigarrow$ complexity dichotomy for finite-domain VCSPs applies
$\Rightarrow$ the resilience problem for q is in P or NP-complete (holds even when $I(q)$ is acyclic)

Hard resilience problems

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

Fact: If Aut(Γ) is oligomorphic and Γ pp-constructs Δ, then VCSP(Δ) reduces to VCSP(Γ) in poly-time.

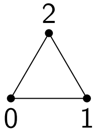
Hard resilience problems

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

Fact: If Aut(Γ) is oligomorphic and Γ pp-constructs Δ, then VCSP(Δ) reduces to VCSP(Γ) 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: VCSP(K_3) is the 3-colorability problem and hence NP-hard.

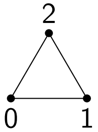
Hard resilience problems

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

Fact: If $\mathrm{Aut}(\Gamma)$ is oligomorphic and Γ pp-constructs Δ, then $\mathrm{VCSP}(\Delta)$ reduces to $\mathrm{VCSP}(\Gamma)$ 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: $\mathrm{VCSP}(K_3)$ is the 3-colorability problem and hence NP-hard.

Corollary

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

Fractional polymorphisms

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

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

Fractional polymorphisms

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

Idea: 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 probablity distribution ω on operations $D^n \to D$ such that for every k-ary $R \in \tau$ and $a^1, \ldots, a^n \in D^k$

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

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 of arity ≥ 2 (essentially Kozik, Ochremiak ('15)).*
- *If Γ has a cyclic fractional polymorphism of arity ≥ 2, then VCSP(Γ) 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 of arity ≥ 2 (essentially Kozik, Ochremiak ('15)).*
- *If Γ has a cyclic fractional polymorphism of arity ≥ 2, then $\mathrm{VCSP}(\Gamma)$ is in P (Kolmogorov, Krokhin, Rolínek ('15)).*

Theorem (Bodirsky, Lutz, S.)

q – conjunctive query
If Γ_q has a fractional polymorphism of arity ≥ 2 which is canonical and pseudo-cyclic with respect to $\mathrm{Aut}(\Gamma_q)$, then $\mathrm{VCSP}(\Gamma_q)$ 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 of arity ≥ 2 (essentially Kozik, Ochremiak ('15)).*
- *If Γ has a cyclic fractional polymorphism of arity ≥ 2, then* VCSP(Γ) *is in P (Kolmogorov, Krokhin, Rolínek ('15)).*

Theorem (Bodirsky, Lutz, S.)

q – *conjunctive query*
If Γ_q has a fractional polymorphism of arity ≥ 2 which is canonical and pseudo-cyclic with respect to $\mathrm{Aut}(\Gamma_q)$, *then* $\mathrm{VCSP}(\Gamma_q)$ *is in P.*

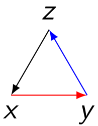
Conjecture: If Γ_q does not pp-construct K_3, then the tractability theorem applies and $\mathrm{VCSP}(\Gamma_q)$ and hence resilience of q is in P.

Examples

Example (hardness):

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

- resilience of q is known to be NP-hard (Freire, Gatterbauer, Immerman, Meliou ('15))
- Γ_q pp-constructs K_3

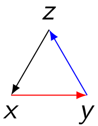


Examples

Example (hardness):

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

- resilience of q is known to be NP-hard (Freire, Gatterbauer, Immerman, Meliou ('15))
- Γ_q pp-constructs K_3



Example (tractability):

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

- complexity left open in Freire, Gatterbauer, Immerman, Meliou ('20)
- Γ_q has a canonical and pseudo-cyclic fractional polymorphism

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.