Valued Constraint Satisfaction Problems and Resilience in Database Theory

Manuel Bodirsky

Joint work with Carsten Lutz and Žaneta Semanišinová
Institut für Algebra, TU Dresden

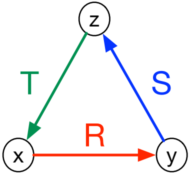
14.03.2023

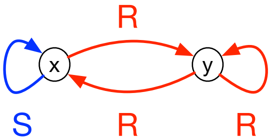
European Research Council
Established by the European Commission

ERC Synergy Grant POCOCOP (GA 101071674).

Ho!

Overview



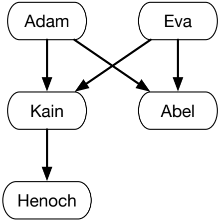


1. Resilience in Database Theory
2. Complexity of Resilience
3. Connection with Valued Constraint Satisfaction Problems
4. Universal-Algebraic Approach
5. NP-hardness and polynomial-time tractability
6. Tractability Conjecture

Conjunctive Queries

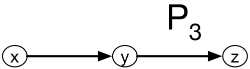
Database: relational structure $\mathfrak{A}$.

x is parent	of y
Adam	Kain
Eva	Kain
Adam	Abel
Eva	Abel
Kain	Henoch



Conjunctive query: primitive positive formula q, e.g.

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



In our example:

$$\mathfrak{A} \models q$$

$$P_3 \to \mathfrak{A}$$

Resilience

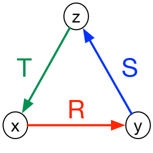
Resilience problem: How many tuples must be removed from relations of $\mathfrak{A}$ s.t.

$$\mathfrak{A} \not\models q?$$

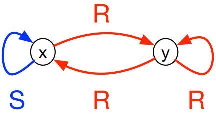
Computational complexity depends on q!

Examples. Meliou+Gatterbauer+Moore+Suciu (DVLDB'10), Freire+Gatterbauer+Immerman+Meliou (VLDB'2015,PODS'20).

- $\exists x, y, z(R(x,y) \wedge S(y,z) \wedge T(z,x))$.
  Resilience problem is NP-hard.



- $\exists x, y(R(x,y) \wedge R(y,y) \wedge R(y,x) \wedge S(x))$
  Complexity left open in PODS'20.



Research Goal:
Classify complexity of resilience for all conjunctive queries q!

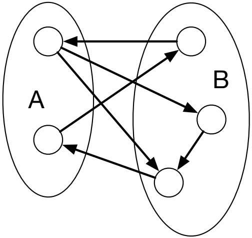
Valued Constraint Satisfaction Problems

Given: a finite set of variables, a finite set of constraints.

- **CSP (Constraint Satisfaction Problem):**
  decide whether there exists a solution that satisfies all constraints.
- **Max CSP:** find a solution that satisfies as many constraints as possible.
- **Valued CSP:** Find solution of minimal cost: each constraint comes with costs depending on the chosen values.

Example. Max Cut (NP-hard)
Given a finite directed graph (V, E), find a partition A, 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.



Valued Structures

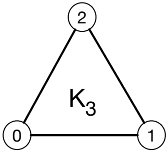
Γ: valued structure.
(Countable) domain D.
(Finite, relational) signature τ.
For each $R \in \tau$ of arity k, function $R^\Gamma \colon D^k \to \underbrace{\mathbb{Q} \cup \{\infty\}}_{\text{'costs'}}$.

Example 1. Γ_{MC}.
$D = \{0, 1\}$.
$\tau = \{E\}$ where E is binary relation symbol.
$E^{\Gamma_{MC}} \colon D^2 \to \mathbb{Q} \cup \{\infty\}$ given by

$$E^{\Gamma_{MC}}(a, b) = \begin{cases} 0 & \text{if } a = 0 \text{ and } b = 1, \\ 1 & \text{otherwise.} \end{cases}$$

Example 2. K_3. $D = \{0, 1, 2\}$, $\tau = \{E\}$.

$$E^{K_3}(a, b) = \begin{cases} 0 & \text{if } a \neq b, \\ \infty & \text{otherwise.} \end{cases}$$



VCSPs, Formal Definition

Fixed: Γ.

Definition (VCSP(Γ))

Input: $u \in \mathbb{Q}$, and an expression ϕ of the form

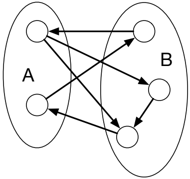
$$\inf_{x \in D^n} \sum_{i \in \{1,\dots,m\}} \psi_i$$

where each ψ_i is of the form $R(x_{i_1}, \dots, x_{i_k})$
for $R \in \tau$ of arity k and $i_1, \dots, i_k \in \{1, \dots, n\}$.

Question: $\phi \leq u$ in Γ?

Examples.

- VCSP(Γ_{MC}) is the Max Cut Problem!
- VCSP(K_3) is 3-colorability Problem!



VCSP Dichotomy

Γ: valued structure with a finite domain.

Theorem.

VCSP(Γ) is in P or NP-hard.

Guide to the literature:

- Živný+Thapper (STOC'13): proof if no ∞ costs.
- Kozik+Ochremiak (ICALP'15): hardness condition.
  If hardness condition does not apply:
  Γ has cyclic fractional polymorphism of arity at least two.
- Kolmogorov+Rolínek+Krokhin (FOCS'15): in this case, VCSP(Γ) is in P if the finite-domain Feder-Vardi CSP dichotomy conjecture is true.
- Bulatov (FOCS'17), Zhuk (FOCS'17):
  proof of Feder-Vardi conjecture.

Resilience Problems as VCSPs

Homomorphism duality: for every finite digraph G we have

$$P_3 \not\to G \text{ if and only if } G \to P_2$$

Turn P_2 into a valued structure Γ with signature $\{E\}$: define

$$E^\Gamma(a, b) := \begin{cases} 0 & \text{if } (a, b) \in E \\ 1 & \text{otherwise} \end{cases}$$

Note: $\Gamma = \Gamma_{MC}$!
Consequence: The following problems are identical:

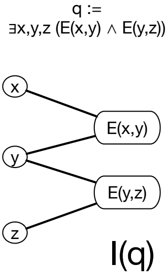
- The resilience problem for $q := \exists x, y, z \big(R(x, y) \wedge R(y, z) \big)$ (the same tuple might appear multiple times in the database)
- The VCSP for Γ_{MC}.

Consequence: Resilience problem for q is NP-hard.

For which queries q is there a dual structure $\mathfrak{B}$ such that for every finite structure $\mathfrak{A}$

$$\mathfrak{A} \not\models q \text{ if and only if } \mathfrak{A} \to \mathfrak{B} \text{ ?}$$

Definition. incidence graph $I(q)$:
bipartite undirected multigraph.
First colour class: variables of q.
Second colour class: conjuncts of q.
Edges link conjuncts with their variables.



Theorem (Nešetřil+Tardiff'00; Larose+Loten+Tardif'07; Foniok'07).
A conjunctive query q has a finite dual if and only if $I(q)$ is a tree.

Consequences

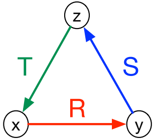
Theorem (B.+Lutz+Semanišinová).
Let q be a conjunctive query such that $I(q)$ is a tree.
Then the resilience problem for q is NP-hard or in P.

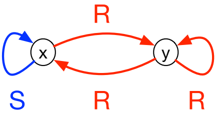
Proof idea: turn the finite dual $\mathfrak{B}_q$ of q into a valued structure Γ_q (all cost functions take values in $\{0, 1\}$).

Generalisations:

1. Presence of 'exogenous' tuples:
   the tuples for some specified relations may not be removed.
   use cost ∞ instead of 1 in the dual.
2. '(Finite) unions of conjunctive queries' instead of conjunctive queries.
3. It suffices that $I(q)$ is acyclic.

But what if $I(q)$ contains cycles?





Cherlin-Shelah-Shi

q: conjunctive query.

Theorem (Cherlin+Shelah+Shi Adv.Appl.Math'99).

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

A permutation group G on a countably infinite set B is called oligomorphic if $G \curvearrowright B^n$ has finitely many orbits for every $n \geq 1$.

Example. $\mathrm{Aut}(\mathbb{Q}; <)$ is oligomorphic.
(However, $(\mathbb{Q}; <)$ is not a dual of a single conjunctive query.)

Fact. If G is oligomorphic, then $|G| = 2^{\aleph_0}$.

Model-Complete Cores

q: conjunctive query.

Theorem (B.'06).

The dual $\mathfrak{B}$ of q can be chosen so that it is

- model complete: every first-order formula is in $\mathfrak{B}$ equivalent to an existential formula;
- a core: every homomorphism from $\mathfrak{B}$ to $\mathfrak{B}$ is an embedding.

Moreover, $\mathfrak{B}$ is up to isomorphism uniquely described by these properties, and $\text{Aut}(\mathfrak{B})$ is oligomorphic.

Example. Let $q = \exists x, y \big(E(x, y) \wedge E(y, x) \big)$.
Then $\mathfrak{B}$ is the so-called random tournament:
the up to isomorphic unique model of the almost-sure theory of the uniform distribution on finite tournaments of size n.

Consequences

q: conjunctive query such that $I(q)$ is connected.
$\mathfrak{B}_q$: model-complete core dual of q.
Γ_q: valued structure obtained from $\mathfrak{B}_q$.

Theorem (B., Lutz, Semanišinová).
The resilience problem for q equals $\mathrm{VCSP}(\Gamma_q)$.

Again:

- Also works with exogeneous tuples.
- Also works for unions of conjunctive queries.
- Assumption that $I(q)$ is connected can be made wlog.

Expressive Power of Valued Structures

Γ: valued structure with domain D and signature τ.
ϕ: τ-expression $\sum_{i \in \{1,\dots,m\}} \psi_i$.
$R \colon D^k \to \mathbb{Q} \cup \infty$.

Definition. $\phi(x_1, \dots, x_k, y_1, \dots, y_l)$ expresses R in Γ if for all $a \in D^k$

$$R(a) = \inf_{b \in D^k} \phi^\Gamma(a, b)$$

Fact. If $\mathrm{Aut}(\Gamma)$ is oligomorphic, then $\mathrm{VCSP}(\Gamma, R)$ reduces to $\mathrm{VCSP}(\Gamma)$.

Other complexity-preserving expansions of Γ:

- $R_\emptyset(a) := \infty$ for all $a \in D$.
- $R_=(a, b) := 0$ if $x = y$ and $R_=(a, b) = \infty$ otherwise.
- non-negative scaling: $r \cdot R$ for $r \in \mathbb{Q}_{\geq 0}$.
- shifting: $R + s$ for $s \in \mathbb{Q}$.
- $\mathrm{Feas}(R) := \{a \in D^k \mid R(a) < \infty\}$.
- $\mathrm{Opt}(R) := \{a \in \mathrm{Feas}(R) \mid R(a) \leq R(b) \text{ for every } b \in D^k\}$.

Hardness

Definition

- $\langle\Gamma\rangle$: valued structure obtained from Γ by adding $R_\emptyset$ and $R_=$ and closing under expressibility, non-negative scaling, shifting, Feas, and Opt.
- d-th pp-pwer of Γ: valued structure Δ with domain D^d such that for every R of arity k in Δ there exists S of arity dk in $\langle\Gamma\rangle$ such that

$$R\big((a_1^1,\dots,a_d^1),\dots,(a_1^k,\dots,a_d^k)\big) = S(a_1^1,\dots,a_d^1,\dots,a_1^k,\dots,a_d^k).$$

- Γ pp-constructs Δ if Δ is fractionally homomorphically equivalent to a pp-power of Γ.

Fact. If $\mathrm{Aut}(\Gamma)$ is oligomorphic and Γ pp-constructs Δ, then $\mathrm{VCSP}(\Delta)$ reduces to $\mathrm{VCSP}(\Gamma)$.

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

Fractional Homomorphisms

Definition. A fractional map from D to C is a probability distribution

$$\big(C^D, \underbrace{\mathcal{B}(C^D)}_{\text{Borel } \sigma\text{-algebra}}, \omega\colon \mathcal{B}(C^D) \to [0,1]\big).$$

A fractional homomorphism between valued structures Δ to Γ with the same signature τ and domains D and C is a fractional map from D to C such that for every $R \in \tau$ of arity k and every $a \in D^k$

$$E_\omega[f \mapsto R^\Gamma(f(a))]$$

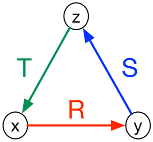
exists (always exists if $\mathrm{Aut}(\Gamma)$ is oligomorphic) and

$$E_\omega[f \mapsto R^\Gamma(f(a))] \leq R^\Delta(a).$$

Remarks.

- Fractional homomorphisms compose.
- Hence: may define fractional homomorphic equivalence.
- Fractional homomorphic equivalence preserves complexity of VCSP.

Example



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

Claim. Γ_q pp-constructs Γ_{K_3}.

Consequences.

- $\mathrm{VCSP}(\Gamma_q)$ is NP-hard.
- Resilience problem for q is NP-hard.

Fractional Polymorphisms

Γ: valued structure with domain D and signature τ.
Fractional polymorphism of Γ:
fractional homomorphism ω from specific pp power Γ^ℓ to Γ:
for every $R \in \tau$ of arity k

$$R^{\Gamma^\ell}((a^1_1, \dots, a^1_\ell), \dots, (a^k_1, \dots, a^k_\ell)) := \frac{1}{\ell} \sum_{i \in \{1,\dots,\ell\}} R^\Gamma(a^1_i, \dots, a^k_i).$$

Idea. Expected cost of a k-tuple obtained from applying ω to ℓ tuples is at most the average cost of these tuples.

Example. $\pi^\ell_i \colon D^\ell \to D$ given by $\pi^\ell_i(x_1, \dots, x_\ell) = x_i$.
Id_ℓ given by $\mathrm{Id}_\ell(\{\pi^\ell_i\}) := \frac{1}{\ell}$ for every $i \in \{1, \dots, \ell\}$
is fractional polymorphism for every Γ.

Polynomial-time Tractability

$f\colon D^\ell \to D$ is cyclic if for all $x_1, \dots, x_\ell \in D$:

$$f(x_1, \dots, x_\ell) = f(x_2, \dots, x_\ell, x_1).$$

ω is called cyclic if for every $A \in \mathcal{B}(D^{D^\ell})$ we have

$$\omega(A) = \omega(\{f \in A \mid f \text{ is cyclic}\})$$

Theorem.

Γ: valued structure over finite domain. Then

- If K_3 has no pp-construction in Γ, then Γ has cyclic fractional polymorphism of arity $\ell \geq 2$ (essentially Kozik+Ochremiak).
- If Γ has cyclic fractional polymorphism of arity $\ell \geq 2$, then VCSP(Γ) is in P (Kolmogorov+Krokhin+Rolínek)

Tractability Conjecture

q: conjunctive query.

Conjecture. If K_3 does not have a pp-construction in Γ_q, then

- $\mathrm{VCSP}(\Gamma_q)$ is in P and
- the resilience problem for q is in P.

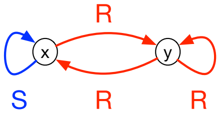
Theorem (B.,Lutz,Semanišinová).

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

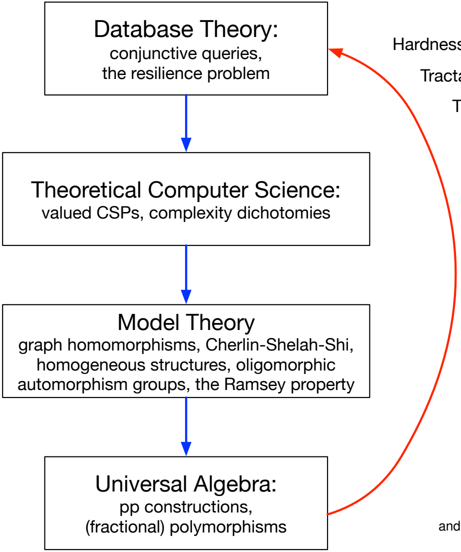
Proof by reduction to the finite, similarly as in B.+Mottet (LICS'16).

Example $\exists x, y \big(R(x,y) \wedge R(y,y) \wedge R(y,x) \wedge S(x)\big)$
Complexity left open at PODS'20.

Has such a polymorphism.



Summary



Hardness Conditions

Tractability Conditions

Tractability Conjecture

Dichotomy for Acylic Conjunctive Queries

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.