The Complexity of Resilience Problems via Valued Constraint Satisfaction Problems

Manuel Bodirsky

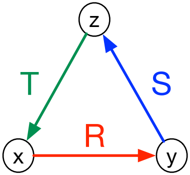
Joint work with Carsten Lutz and Žaneta Semanišinová
Institut für Algebra, TU Dresden

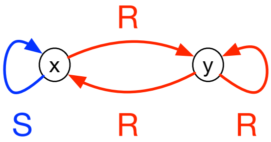
23.02.2024

European Research Council
Established by the European Commission

ERC Synergy Grant POCOCOP (GA 101071674).

Overview



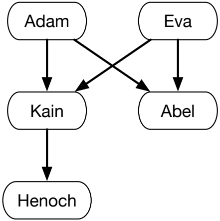


1. Resilience in database theory
2. Complexity of resilience
3. Connection with valued constraint satisfaction problems
4. Universal-algebraic approach
5. NP-hardness and polynomial-time tractability
6. Tractability conjecture
7. Comparison with previous results, examples

Conjunctive Queries

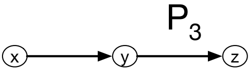
Database: relational structure $\mathfrak{A}$.

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

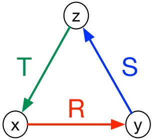
Resilience problem (for q): Given finite database $\mathfrak{A}$,
how many tuples must be removed from relations of $\mathfrak{A}$ s.t.

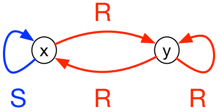
$$\mathfrak{A} \not\models q?$$

Computational complexity depends on q!

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

- $\exists x, y, z(R(x, y) \wedge S(y, z) \wedge T(z, x))$.
  'Triad': 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

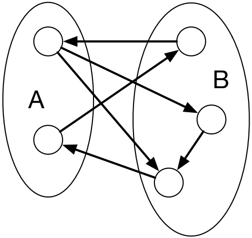
Given: a finite set of variables, a finite set of constraints.

- **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 $R^\Gamma \colon D^k \to \underbrace{\mathbb{Q} \cup \{\infty\}}_{\text{'costs'}}$.

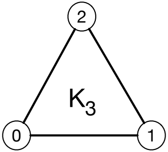
Example 1. Γ_{MC}.
$D = \{0, 1\}$.
$\tau = \{E\}$ where E is binary relation symbol.
$E^{\Gamma_{MC}} \colon D^2 \to \mathbb{Q} \cup \{\infty\}$ given by

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

P_2

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: valued structure Γ.

Definition (VCSP(Γ))

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

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

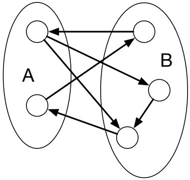
where each ψ_i is of the form $R(x_{i_1}, \dots, x_{i_k})$ for $R \in \tau$ of arity k and $i_1, \dots, i_k \in \{1, \dots, n\}$.

Question: $\phi \leq u$ in Γ?

Examples.

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

(Both problems NP-hard)



Finite-Domain VCSP Dichotomy

Γ: valued structure with a finite domain.

Theorem.

VCSP(Γ) is in P or NP-hard.

Guide to the literature:

- Cohen, Cooper, Jeavons (CP'2006): 'An algebraic characterisation of complexity for valued constraints'
- 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 $q := \exists x, y, z \big(E(x, y) \wedge E(y, z) \big)$ (bag semantics: the same tuple might appear multiple times in database)
- The VCSP for Γ_{MC} (Mac-Cut).

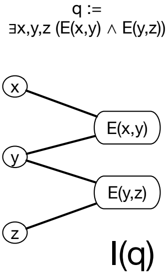
Consequence: Resilience problem for q is NP-hard.

Homomorphism Dualities

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.

Dichotomy for Acyclic Queries

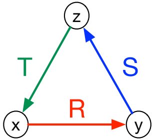
Theorem (B.+Lutz+Semanišinová).
Let q be a conjunctive query such that $I(q)$ is a tree.
Then the resilience problem for q is NP-hard or in P.

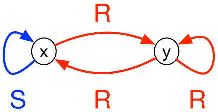
Proof idea: turn the finite dual $\mathfrak{B}_q$ of q into a valued structure Γ_q (all cost functions take values in $\{0, 1\}$).

Generalisations:

1. Presence of 'exogenous' tuples:
   the tuples for some specified relations R may not be removed.
   Use cost ∞ instead of 1 for valued relation R 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 such that $I(q)$ is connected.

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

q has a countable dual $\mathfrak{B}$ such that $\text{Aut}(\mathfrak{B})$ 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. $\text{Aut}(\mathbb{Q}; <)$ is oligomorphic.
(However, $(\mathbb{Q}; <)$ is not a dual of a single conjunctive query.)

Theorem (Hubicka+Nešetřil MVLSC'16).

There exists a dual $\mathfrak{B}$ of q which is reduct of a finitely bounded homogeneous structure.

- $\mathfrak{B}$ is finitely bounded if there exists a finite set of structures $\mathcal{F}$ such that $\mathfrak{A} \hookrightarrow \mathfrak{B}$ if and only if $\mathfrak{F} \not\hookrightarrow \mathfrak{A}$ for all $\mathfrak{F} \in \mathcal{F}$.
- $\mathfrak{B}$ is homogeneous if all isomorphisms between finite substructures of $\mathfrak{B}$ extend to automorphisms of $\mathfrak{B}$.

Arbitrary Resilience Problems as VCSPs

q: conjunctive query such that $I(q)$ is connected.
$\mathfrak{B}_q$: dual of q such that $\text{Aut}(\mathfrak{B}_q)$ is oligomorphic.
Γ_q: valued structure obtained from $\mathfrak{B}_q$.

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

Again:

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

Fractional Homomorphisms

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

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

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

1. $E_\omega[f \mapsto R^\Gamma(f(a))]$ exists (always exists if $\mathrm{Aut}(\Gamma)$ is oligomorphic), and
2. $E_\omega[f \mapsto R^\Gamma(f(a))] \leq R^\Delta(a)$.

Remarks.

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

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: $D^k \to \mathbb{Q} \cup \infty$.

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.

Example 1

P$_3$

Recall: Resilience problem for $q = \exists x, y, z(R(x, y) \wedge R(y, z))$

P$_2$

equals VCSP(Γ_{MC}), i.e., MaxCut, and is NP-hard.

Γ_{MC} pp-constructs K_3: Suffices to show that Γ_{MC} pp-constructs

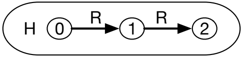
$$\text{NAE} := (\{0,1\}; \{0,1\}^3 \setminus \{(0,0,0),(1,1,1)\})$$

because it is known that NAE pp-constructs K_3.

$$\text{NAE}(x, y, z) = \text{Opt}\big(R(x, y) + R(y, z) + R(z, x)\big)$$

Example 2

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



Fact: q has homogeneous dual $\mathfrak{B}_q$ (Reason: Gaifman graph of q is clique)
Note: No finite dual!
Γ_q: corresponding valued structure.

Claim: Γ_q pp-constructs K_3.
Fact: Suffices to show that Γ_q pp-constructs

$$\text{OIT} := \big(\{0, 1\}; \{(0, 0, 1), (0, 1, 0), (1, 0, 0)\}\big).$$

Define

$$S(x, y, z) := R(x, y) + R(y, z) + \text{Opt}(H(x, y, z))$$

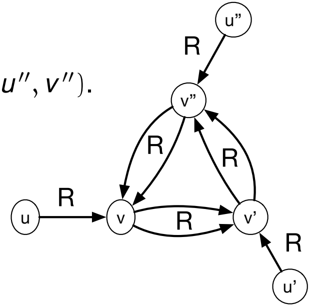
Example 2, pp-construction

$$G(u,v,u',v',u'',v'') := \mathrm{Opt}\Big(\overbrace{S(u,v,v') + S(u',v',v'') + S(u'',v'',v)}^{\geq 3} \\ + \underbrace{S(v,v',v'') + S(v',v'',v) + S(v'',v,v')}_{\geq 2}\Big)$$

pp-power Δ with domain B^2
and $\mathrm{OIT}^\Delta((u,v),(u',v'),(u'',v'')) := G(u,v,u',v',u'',v'')$.

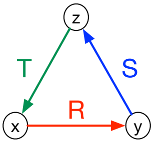
Δ is homomorphically equivalent to $(\{0,1\}; \mathrm{OIT})$:

- $h(u,v) := \begin{cases} 0 & \text{if} (u,v) \in R^{\mathfrak{B}} \\ 1 & \text{otherwise} \end{cases}$
- Pick $u,v,x,y \in B$ such that $\{u,v,x,y\}^3 \subseteq H^{\mathfrak{B}}$, $(u,v) \in R^{\mathfrak{B}}$, $(x,y) \notin R^{\mathfrak{B}}$.
  Define $g(0) := (u,v)$, $g(1) := (x,y)$.



Example 3

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



Known: NP-hard (Freire, Gatterbauer, Immerman, Meliou VLDB'2015)
'Self-join free': every relation symbol appears at most once.

Fact. Γ_q pp-constructs K_3.

Consequences.

- VCSP(Γ_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

$$R^{\Gamma^\ell}((a_1^1, \dots, a_\ell^1), \dots, (a_1^k, \dots, a_\ell^k)) := \frac{1}{\ell} \sum_{i \in \{1,\dots,\ell\}} R^\Gamma(a_i^1, \dots, a_i^k).$$

Idea:

Expected cost of a k-tuple obtained from applying ω to ℓ tuples $\leq$ the average cost of these tuples.

Example. $\pi_i^\ell \colon D^\ell \to D$ given by $\pi_i^\ell(x_1, \dots, x_\ell) = x_i$.

Id_ℓ given by $\mathrm{Id}_\ell(\{\pi_i^\ell\}) := \frac{1}{\ell}$ 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

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

Theorem.

Γ: valued structure over finite domain. Then

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

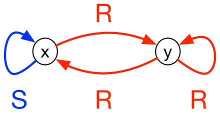
- VCSP(Γ_q) is in P and
- the resilience problem for q is in P
  (for bag semantics, therefore also for set semantics)

Theorem (B.,Lutz,Semanišinová).

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

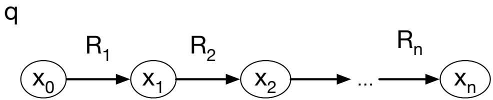
Proof by reduction to the finite, similarly as for CSPs in B.+Mottet (LICS'16).

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



Γ_q has such a polymorphism.
Hence: Resilience problem for q is in P.

Example 4



q acyclic and self-join free!

- has finite dual $\mathfrak{D}$.
- is linear: Resilience problem for q is in P (Gatterbauer, Immerman, Meliou VLDB'2015)

$\mathfrak{D}$: vertex v_S for every $S \subseteq \{1, \ldots, n-1\}$.
Idea: v_S satisfies $\phi_i(x)$ for all $i \in S$, where

$$\phi_i(x) := \exists x_{i+1}, \ldots, x_n \big(R_{i+1}(x, x_{i+1}) \wedge \cdots \wedge R_n(x_{n-1}, x_n) \big).$$

$R_i(v_S, v_T)$ holds iff

- $i < n$ and $i \notin T$, or
- $i > 1$ and $(i-1) \in S$.

Example 4: Fractional Polymorphisms!

- $f\colon D^2 \to D\colon (v_S, v_T) \mapsto v_{S \cup T}$.
- $g\colon D^2 \to D\colon (v_S, v_T) \mapsto v_{S \cap T}$.
- ω: binary fractional polymorphism defined by

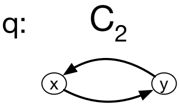
$$\omega(f) = \omega(g) := \frac{1}{2}$$

Verify:

- ω is cyclic!
- ω is fractional polymorphism of Γ_q.

In this case: cost functions are submodular.

Example 5: Cyclic and Self-joins

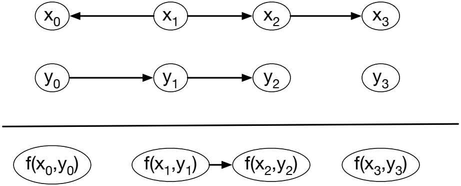


q cyclic and with self-join. No finite dual!
Infinite dual $\mathfrak{D}$: 'the random oriented graph', i.e., countable homogeneous universal oriented graph.
Exists $f \colon \mathfrak{D}^2 \hookrightarrow \mathfrak{D}$.

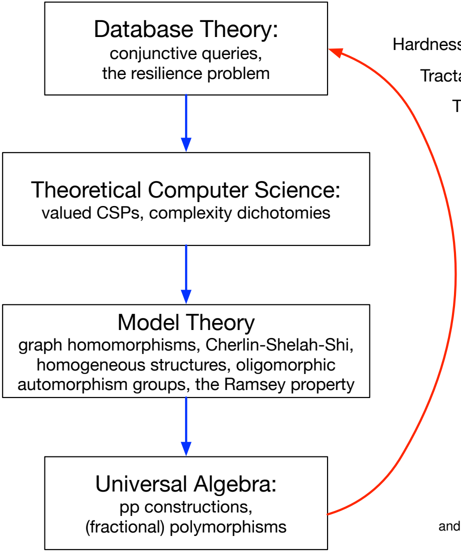
- f injective, not cyclic!
- f is pseudo-cyclic: there exist $\alpha, \beta \in \text{Aut}(\mathfrak{D})$ such that

$$\alpha(f(x,y)) = \beta(f(y,x)).$$

- f is canonical: for $s, t \in D^n$, the orbit of $f(s,t)$ in $\text{Aut}(\mathfrak{D})$ only depends on the orbit of s and the orbit of t in $\text{Aut}(\mathfrak{D})$.
- ω defined by $\omega(f) := 1$ is canonical pseudo-cyclic fractional polymorphism of Γ_q.



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.