Valued Constraint Satisfaction Problem and Resilience in Database Theory

Žaneta Semanišinová
joint work with Manuel Bodirsky and Carsten Lutz

Institute of Algebra
TU Dresden

Dagstuhl Seminar 25081
Semirings in Databases, Automata, and Logic
20 Feb 2025

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

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

$\hookrightarrow$ core problem in reverse data management
$\hookrightarrow$ appears first in [Meliou, Gatterbauer, Moore, Suciu '10]

Resilience of queries

database – a relational structure $\mathfrak{A}$
conjunctive query – a formula q of the form $\exists y_1, \dots, y_l\, (\psi_1 \wedge \dots \wedge \psi_m)$, where ψ_i are atomic

Definition (resilience)

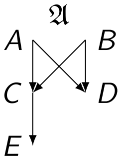
Fixed conjunctive query q.
Input: a finite database $\mathfrak{A}$, $u \in \mathbb{N}$
Output: Can we remove $\leq u$ tuples from relations of $\mathfrak{A}$, so that $\mathfrak{A} \not\models q$?

$\hookrightarrow$ core problem in reverse data management
$\hookrightarrow$ appears first in [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 $\mathfrak{A} \not\models q$?

$\hookrightarrow$ core problem in reverse data management

$\hookrightarrow$ appears first in [Meliou, Gatterbauer, Moore, Suciu '10]

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.
- **MinCSP**: Find the minimal number of constraints to violate so that the remaining constraints are satisfiable simultaneously.
- **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.
- **MinCSP**: Find the minimal number of constraints to violate so that the remaining constraints are satisfiable simultaneously.
- **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 MinCSP.
Proof: Model the tuples in relations with cost 0 and outside with cost 1 (for MinCSP) or ∞ (for CSP).

Focus on VCSP

A valued structure Γ consists of:

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

the concrete choice of the ordered semiring is not crucial

Focus on VCSP

A valued structure Γ consists of:

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

the concrete choice of the ordered semiring is not crucial

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

$$\inf_{t \in C^n} \phi(t) \leq u \text{ in } \Gamma?$$

Example: Max-Cut as a VCSP

Input: $G = (V, E)$ – finite directed (multi)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: Max-Cut as a VCSP

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

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

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

Example: Max-Cut as a VCSP

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

- $C = \{0, 1\}$
- $\tau = \{R\}$, 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

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

Example: Max-Cut as a VCSP

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

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

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

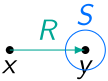
Take vertices of G as variables. The size of a maximal cut of G is

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

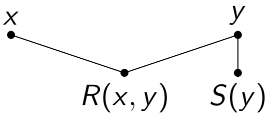
every instance of VCSP(Γ_{MC}) corresponds to a directed multigraph
$\rightsquigarrow$ VCSP(Γ_{MC}) is the Max-Cut problem (NP-hard)

Translation to a dual problem

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



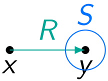
canonical structure

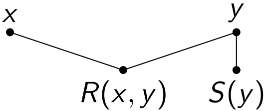


incidence graph I(q)

Translation to a dual problem

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





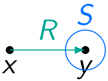
Theorem (Cherlin, Shelah, Shi '99)

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 $\mathfrak{A}$:

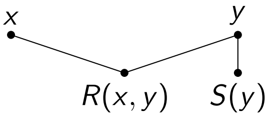
$$\mathfrak{A} \not\models q \;\Leftrightarrow\; \mathfrak{Q} \not\to \mathfrak{A} \;\Leftrightarrow\; \mathfrak{A} \to \mathfrak{B}_q$$

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)

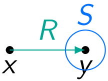
Let q be a query and $\mathfrak{Q}$ its canonical structure. If $I(q)$ is connected, then there exists a structure $\mathfrak{B}_q$, such that for every finite $\mathfrak{A}$:

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

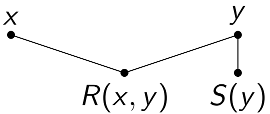
- *$\mathfrak{B}_q$ can be chosen so that $\mathrm{Aut}(\mathfrak{B}_q)$ is oligomorphic.*

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)

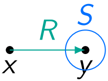
Let q be a query and $\mathfrak{Q}$ its canonical structure. If $I(q)$ is connected, then there exists a structure $\mathfrak{B}_q$, such that for every finite $\mathfrak{A}$:

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

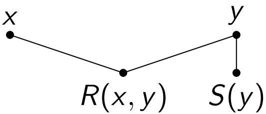
- *$\mathfrak{B}_q$ can be chosen so that $\mathrm{Aut}(\mathfrak{B}_q)$ is oligomorphic.*
- *$\mathfrak{B}_q$ can be chosen finite iff q is homomorphically equivalent to q' such that $I(q')$ is a tree. (Nešetřil, Tardiff '00; Larose, Loten, Tardiff '07)*

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 $\mathfrak{A}$:

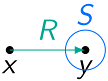
$$\mathfrak{A} \not\models q \;\Leftrightarrow\; \mathfrak{Q} \not\to \mathfrak{A} \;\Leftrightarrow\; \mathfrak{A} \to \mathfrak{B}_q$$

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

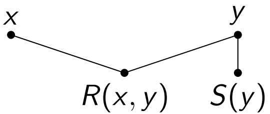
oligomorphic – countable domain B_q and the action of $\mathrm{Aut}(\mathfrak{B}_q)$ on B_q^n has finitely many orbits for every $n \geq 1$

Translation to a dual problem

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



canonical structure



incidence graph I(q)

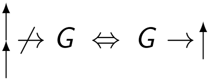
Theorem (Cherlin, Shelah, Shi '99)

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 q is homomorphically equivalent to q' such that $I(q')$ is a tree. (Nešetřil, Tardiff '00; Larose, Loten, Tardiff '07)*

Example: For every finite directed graph G we have:



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, S., Lutz '24)

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

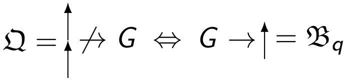
Connection of resilience and VCSPs

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, S., Lutz '24)

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

Example: $q := \exists x, y, z (R(x, y) \wedge R(y, z))$
For every finite G:



$\mathfrak{B}_q \rightsquigarrow \Gamma_{\mathrm{MC}} = (\{0, 1\}; R)$
Resilience of $q = \mathrm{VCSP}(\Gamma_{\mathrm{MC}}) =$ Max-Cut is NP-hard

Connection of resilience and VCSPs

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

Theorem (Bodirsky, S., Lutz '24)

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 $R(x, y)$.

Complexity dichotomy for resilience of acyclic queries

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

Complexity dichotomy for resilience of acyclic queries

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

Follows from [Kozik, Ochremiak '15; Kolmogorov, Krokhin, Rolínek '15; Bulatov '17; Zhuk '17]:

Theorem

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

Complexity dichotomy for resilience of acyclic queries

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

Follows from [Kozik, Ochremiak '15; Kolmogorov, Krokhin, Rolínek '15; Bulatov '17; Zhuk '17]:

Theorem

For a finite-domain valued structure Γ, VCSP(Γ) *is in P or NP-complete.*

Corollary (Bodirsky, S., Lutz '24)

Let q be a conjunctive query such that I(q) is acyclic. 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 $\Rightarrow$ Γ_q can be taken finite

Follows from [Kozik, Ochremiak '15; Kolmogorov, Krokhin, Rolínek '15; Bulatov '17; Zhuk '17]:

Theorem

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

Corollary (Bodirsky, S., Lutz '24)

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

Question: For general queries, choose $\mathfrak{B}_q$ such that $\mathrm{Aut}(\mathfrak{B}_q)$ 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 Aut(Γ) is oligomorphic and Γ pp-constructs Δ, then VCSP(Δ) reduces to VCSP(Γ) in poly-time.

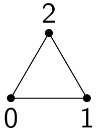
Hard resilience problems

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.

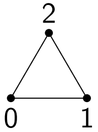
Hard resilience problems

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

Fact: If 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.

Corollary

If Aut(Γ) *is oligomorphic and* Γ *pp-constructs* K_3, *then* VCSP(Γ) *is NP-hard.*

Fractional polymorphisms

polymorphism of a relational structure $\mathfrak{A}$ – $f : A^n \to A$ such that for all relations R of $\mathfrak{A}$ and $t^1, \dots, t^n \in R$, $f(t^1, \dots, t^n) \in R$ (applied row-wise)

Fractional polymorphisms

polymorphism of a relational structure $\mathfrak{A}$ – $f : A^n \to A$ such that for all relations R of $\mathfrak{A}$ and $t^1, \dots, t^n \in R$, $f(t^1, \dots, t^n) \in R$ (applied row-wise)

Example: The operation min is a polymorphism of $(\mathbb{Q}; <)$.

$$\begin{pmatrix} 1 \\ \wedge \\ 5 \end{pmatrix} \begin{pmatrix} 2 \\ \wedge \\ 3 \end{pmatrix} \begin{matrix} \overset{\min}{\to} \\ \\ \overset{\min}{\to} \end{matrix} \begin{pmatrix} 1 \\ \wedge \\ 3 \end{pmatrix}$$

Fractional polymorphisms

polymorphism of a relational structure $\mathfrak{A}$ – $f : A^n \to A$ such that for all relations R of $\mathfrak{A}$ and $t^1, \ldots, t^n \in R$, $f(t^1, \ldots, t^n) \in R$ (applied row-wise)

Example: The operation min is a polymorphism of $(\mathbb{Q}; <)$.

$$\begin{pmatrix} 1 \\ \wedge \\ 5 \end{pmatrix} \begin{pmatrix} 2 \\ \wedge \\ 3 \end{pmatrix} \begin{matrix} \overset{\min}{\to} \\ \\ \overset{\min}{\to} \end{matrix} \begin{pmatrix} 1 \\ \wedge \\ 3 \end{pmatrix}$$

Definition (fractional polymorphism)

A fractional polymorphism of Γ of arity n is a probability distribution ω on the maps $f : C^n \to C$ such that for every k-ary $R \in \tau$ and $t^1, \ldots, t^n \in C^k$

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

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* $\mathrm{VCSP}(\Gamma)$ *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 VCSP(Γ) is in P [Kolmogorov, Krokhin, Rolínek '15].*

Theorem (Bodirsky, S., Lutz '24)

If Γ_q has a fractional polymorphism which is canonical and pseudo cyclic, then VCSP(Γ_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 VCSP(Γ) is in P [Kolmogorov, Krokhin, Rolínek '15].*

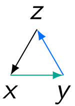
Theorem (Bodirsky, S., Lutz '24)

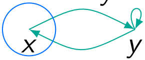
If Γ_q has a fractional polymorphism which is canonical and pseudo cyclic, then VCSP(Γ_q) is in P.

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

Examples

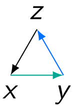
- $q_{\mathrm{MC}} := \exists x, y, z\big(R(x, y) \wedge R(y, z)\big)$
- $q_{\mathrm{path}} := \exists x, y, z\big(R(x, y) \wedge S(y, z)\big)$
- $q_{\triangle} := \exists x, y, z\big(R(x, y) \wedge S(y, z) \wedge T(z, x)\big)$
- $q_{\mathrm{new}} := \exists x, y\big(S(x) \wedge R(x, y) \wedge R(y, x)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

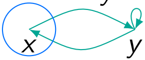




Examples

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





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

References:

- Freire, Gatterbauer, Immerman, Meliou '15
- Freire, Gatterbauer, Immerman, Meliou '20
- Makhija, Gatterbauer '22
- Bodirsky, S., Lutz '24

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.