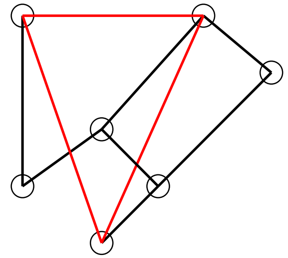
Datalog and Constraint Satisfaction with Infinite Templates

Manuel Bodirsky, Humboldt-Universität zu Berlin

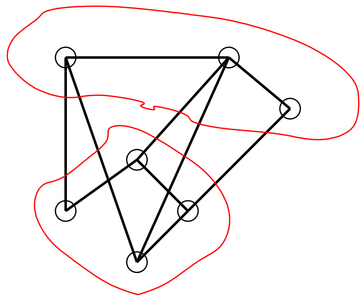
23.3.2006, Oxford

Examples of Computational Problems

1.	Triangle-freeness
Given	Graph G
Question	Is there no triangle in G?



2.	No-mono-tri
Given	A graph $G = (V; E)$
Question	Can we partition V into A and B such that both A and B induce triangle-free graphs?



3.	Cyclic-ordering
Given	Variables V, set of triples $(x_1, x_2, x_3) \in V^3$
Question	Is there an assignment s.t. for each triple either $x_1 < x_2 < x_3$ or $x_2 < x_3 < x_1$ or $x_3 < x_1 < x_2$?

Constraint Satisfaction with Infinite Domains

Template $\Gamma = (D; R_1, R_2, \ldots, R_n)$ relational structure

CSP(Γ)

Given A finite relational structure S

Question $S \to \Gamma$? I.e., is there a homomorphism h from S to Γ?

Example Cyclic-ordering is CSP(($\mathbb{Q}$; Cyc)) where Cyc is $\{(x_1, x_2, x_3) \in \mathbb{Q}^3 \mid x_1 < x_2 < x_3 \vee x_2 < x_3 < x_1 \vee x_3 < x_1 < x_2\}$.

Observation A computational problem can be formulated as a CSP if and only if it is closed under disjoint unions and its complement is closed under homomorphisms.

ω-Categorical Structures

Question For which infinite templates does the algebraic approach to constraint satisfaction work?

Theorem [Engeler/Ryll-Nardzewski/Svenonius – see Hodges'97]:
For a countable relational structure Γ, tfae:

1. Aut(Γ) has finitely many orbits of n-tuples, for all n
2. There are finitely many inequivalent first-order formulas with n free variables in Γ, for all n
3. The first-order theory of Γ has only one countable model, up to isomorphism
4. Γ is ω-categorical

Examples $(\mathbb{Q}; <)$
$(\mathbb{Q}$; Cyc$)$

Datalog

Logic programming	Datalog = Prolog - function symbols
Database theory	Datalog = conjunctive queries + recursion
Example	Datalog program Φ:

$$
\begin{array}{rcl}
\mathsf{tc}(x,y) & \leftarrow & x<y \\
\mathsf{tc}(x,y) & \leftarrow & \mathsf{tc}(x,u),\ \mathsf{tc}(u,y) \\
\mathsf{false} & \leftarrow & \mathsf{tc}(x,x)
\end{array}
$$

Terminology

$<$: input relation symbol

tc, false: IDBs

The program Φ has width $(2,3)$

The program Φ solves CSP$((\mathbb{Q},<))$

Datalog for Constraint Satisfaction

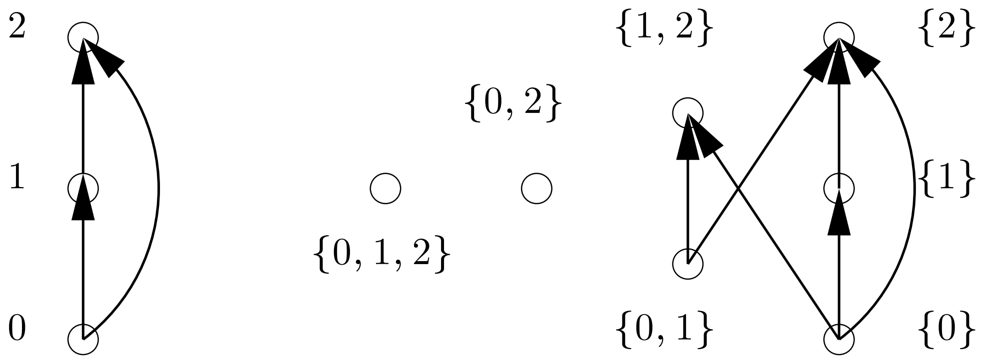
References	Feder+Vardi'93, Kolaitis+Vardi'98
Observation	Datalog programs can be evaluated in polynomial time
Definition	A problem C has width l iff it can be solved by a Datalog program of width (l, k) for some $k \geq l$
Examples	CSP$(\mathbb{Q}, <)$ has width 2 CSP$(\mathbb{Q}, \text{Cyc})$ has unbounded width
Questions	Which CSPs can be solved by Datalog programs? Which Datalog queries can be formulated as CSPs?

Bounded Width for Finite Templates

Width 0 — CSP(T) has width 0 iff its complement is in FO (Atserias'04).

Width 1 — Feder+Vardi'93, Dalmau+Pearson'99:
CSP(T) has width 1 iff $P(T) \to T$.

$P(T)$ — Vertices: non-empty subsets of vertices of T.
Edges: link A and B in $P(T)$ if for all $a \in A$ there is $b \in B$ and for all $b \in B$ there is $a \in A$ such that $ab \in E$.



Width 2 — For finite T not known to be decidable

From Monadic Datalog to MMSNP

An SNP sentence is a second order sentence of the form:

$$\exists R_1, \ldots, R_l . \forall x_1, \ldots, x_k . \Phi, \quad \Phi \text{ quantifier free}$$

Observation — If Γ has finite domain, CSP(Γ) is contained in monadic monotone SNP (MMSNP)

Theorem — [Feder+Vardi'93,Kun'05]: CSP with finite templates has a dichotomy if and only if MMSNP has a dichotomy

Example — MMSNP is strictly larger than CSP with finite templates

$$\forall x, y, z . \neg(E(x,y) \land E(y,z) \land E(x,z))$$

Fact — Every monadic Datalog query can be formulated with a MMSNP sentence

From MMSNP to CSP

Theorem [B.+Dalmau STACS06] Let C be a non-empty MMSNP problem that is closed under disjoint unions. Then $C = \text{CSP}(\Gamma)$ with ω-categorical Γ

Theorem [Cherlin+Shelah+Shi'03, Covington] For every finite set $\mathcal{N}$ of finite connected structures there is an ω-categorical structure Δ that is universal for Forb($\mathcal{N}$)

This is, for all *countable* Δ' we have
$\Delta' \subseteq_{\text{induced}} \Delta$ iff $N \not\to \Delta'$ for all $N \in \mathcal{N}$

Examples For $\mathcal{N} = \{K_3\}$ we get the homogeneous universal triangle-free graph $\not\triangle$.
For $\mathcal{N} = \emptyset$ get the countable random graph.
Δ is not always homogeneous: $\mathcal{N} = \{C_5\}$

Canonical Datalog Programs

Feder+Vardi'93 Canonical Datalog programs for finite templates

Now Let the template Γ be ω-categorical

Definition The canonical Datalog (l, k)-program for CSP(Γ)
- contains an IDB for every at most l-ary primitive positive definable relation in Γ
- contains a rule $R \leftarrow R_1, \dots, R_s$ iff the corresponding implication is valid in Γ, and contains at most k variables

Example (Part of) the canonical program for CSP$(\mathbb{Q}, <)$

tc$(x, y) \leftarrow x < y$
tc$(x, y) \leftarrow$ tc(x, u), tc(u, y)
false $\leftarrow$ tc(x, x)

Theorem CSP(Γ) can be solved with an (l, k)-Datalog program iff the canonical (l, k)-Datalog program solves CSP(Γ)

Bounded Width Characterizations

Width 0 $C :=$ CSP(Γ) has width 0
iff C be be described by forbidden obstructions
iff the complement of C is in FO (Rossmann'05)

Width 1 CSP(Γ) has width one if and only if for some k the structure $P(\Gamma, k)$ homomorphically maps to Γ, where $P(\Gamma, k)$ is constructed as follows:

- Let Φ be the canonical $(1, k)$-program for CSP(Γ)
- View Φ as a MMSNP query
- Formulate this query as a CSP as shown before
- the corrsponding ω-categorical template is $P(\Gamma, k)$

Width 2 Not clear

Strict Bounded Width

Remark: The canonical Datalog program for CSP(Γ) computes an instance of CSP(Γ'), where Γ' is the expansion of Γ by all primitive positive definable relations

Definition An instance S of CSP(Γ) is called globally consistent iff every partial homomorphism from S to Γ can be extended to a full homomorphism from S to Γ.

Definition CSP(Γ) has strict width l iff for some k the canonical (l, k)-program computes on all instances S of CSP(Γ) a globally consistent instance S'.

Example: CSP$(\mathbb{Q}, <)$ has strict width 2.

Strict Width l

Definition We say that a k-ary operation f preserves a structure Γ iff f is a homomorphism from Γ^k to Γ.

An operation f is a weak near-unanimity operation iff
$f(y, x, \ldots, x) = f(x, y, x, \ldots, x) = \ldots$
$\quad = f(x, \ldots, x, y) = f(x, \ldots, x)$

Example $(\mathbb{Q}; <)$ preserved by the ternary median operation
$\not\lhd$ has no nu-operation (Larose+Tardiff'01)
But $\not\lhd$ has a weak nu-operation

Theorem (Dalmau+B. STACS06) Let Γ be an ω-categorical. Then CSP(Γ) has strict width l if and only if it is preserved by a $l{+}1$-ary weak near-unanimity operation.

Summary

- MMSNP queries that are closed under disjoint unions can be formulated as constraint satisfaction problems with ω-categorical templates

- For ω-categorical templates, have notion of canonical Datalog programs

- Characterizations of (strict) bounded width:
  - for templates that are ω-categorical
  - width 0
  - width 1
  - strict width k

Problems

A. Atserias	LFP $\cap$ HOM $=$ Datalog?
New	LFP $\cap$ coCSP $=$ Datalog? for infinite templates
Equivalently	LFP $\cap$ HOM $\cap$ coUnions $=$ Datalog?
A. Atserias	LFP $\cap$ coCSP $=$ Datalog? for finite templates

Is the width-hirarchy strict for ω-categorical templates?

References

Introductory book

Hodges'97 *A shorter model theory*, Cambridge University Press

Papers referenced in this talk

Atserias'04 *On Digraph Coloring Problems and Treewidth Duality*, LICS'04

B.,Dalmau'06 *Datalog and Constraint Satisfaction with Infinite Templates*, STACS'06

Cherlin,Shelah,Shi'99 *Universal graphs with forbidden subgraphs and algebraic closure*, Advances in Applied Mathematics 22, p. 454-491

Covington'90 *Homogenizable Relational Structures*, Illinois Journal of Mathematics 34, p.731-743

Dalmau+Pearson'99 *Closure Functions and Width 1 Problems*, CP'99

Feder+Vardi'93 *Monotone Monadic SNP and Constraint Satisfaction*, STOC'93

Kolaitis,Vardi'98 *Conjunctive-query Containment and Constraint Satisfaction*, PODS'98

Kun'05 *Every Problem in MMSNP is equivalent to a CSP*, unpublished

Larose,Tardiff'01 *Strongly Rigid Graphs and Projectivity*, Multiple-Valued Logic 7, p. 339-361.

Rossmann'05 *Existential Positive Types and Preservation under Homomorphisms*, LICS'05