Abstract of

Constraint Satisfaction with Infinite Domains

by Manuel Bodirsky

Dissertation, November 2004,
Humboldt-Universität zu Berlin, Germany

Referees:
Prof. Dr. Hans Jürgen Prömel,
Prof. Dr. Martin Grohe,
Prof. Dr. Jaroslav Nešetřil

Constraint satisfaction problems occur in many areas of computer science, most prominently in artificial intelligence including temporal or spacial reasoning, belief maintenance, machine vision, and scheduling. Other areas are graph theory, boolean satisfiability, type systems for programming languages, database theory, automatic theorem proving, and, as for some of the problems discussed in this thesis, computational linguistics and computational biology.

Many constraint satisfaction problems have a natural formulation as a homomorphism problem. For a fixed relational structure Γ we consider the following computational problem: Given a structure S with the same relational signature as Γ, is there a homomorphism from S to Γ? This problem is known as the constraint satisfaction problem CSP(Γ) for the so-called *template* Γ and is intensively studied for relational structures Γ with a finite domain. However, many constraint satisfaction problems can not be formulated with a finite template.

If we allow arbitrary infinite templates, constraint satisfaction is very expressive. We show that it contains undecidable problems, even if the constraint language is binary. In general, a computational problem can be described as the constraint satisfaction problem of an infinite template if and only if it is closed under inverse homomorphisms and disjoint unions. It is also easy to see that we can restrict our attention to countable templates.

In this thesis we study the computational complexity of constraint satisfaction with templates that are ω-*categorical*. A structure Γ is ω-*categorical* if all countable models of the first-order theory of Γ are isomorphic to Γ. This concept is central and well-studied in model-theory. On the one hand, ω-categoricity is a rather strong model-theoretic assumption on a relational

structure, and we can use them to show that many techniques for constraint satisfaction with finite templates extend to ω-categorical templates.

On the other hand, the class of computational problems that can be formulated as a constraint satisfaction problem with an ω-categorical template is large. In fact, it contains all the classes of constraint satisfaction problems that were previously studied systematically in the literature. In particular, the following list of well-known computational problems can be described with ω-categorical templates.

- Allen's interval algebra, and all its fragments (Allen, 1983)

- Problems in phylogenetic analysis (Gusfield, 1997, and Steel, 1992)

- Tree description constraints from computational linguistics (Cornell, 1994, B., 2002)

- All problems in the class *monotone monadic SNP without inequality* (Feder and Vardi, 1999)

The class strictly contains all constraint satisfaction problems with finite templates: all the examples mentioned above can not be formulated with finite templates.

The algebraic approach to constraint satisfaction. If a relational structure Γ is ω-categorical we can show that a relation is primitive positive definable in Γ if and only if it is preserved by the *polymorphisms* of Γ. This theorem is well-known for finite templates (Bodnarcuk et al., 1969, Geiger, 1968), and was the starting point of the so-called algebraic approach to study the complexity of constraint satisfaction with finite templates, described e.g. in (Jeavons et al., 1997). It shows that the complexity of a constraint satisfaction problem is determined by the (locally closed) *clone of polymorphisms* of the template. One example where the existence of a certain polymorphism of the template implies tractability are *near-unanimity operations*: If a (finite or infinite) template has a polymorphism f that satisfies $f(y, x, \ldots, x) = f(x, y, x, \ldots, x) = f(x, \ldots, x, y) = y$ it can be solved by a Datalog-program of width k. For finite templates, (Feder and Vardi, 1999) proved that every constraint satisfaction problem that can be solved by a Datalog program of width k can also be solved by the *canonical Datalog program* of width k. This is another result we can generalize to ω-categorical templates.

The notion of a *core* of a relational structure is another concept that turned out to be essential to the complexity analysis of constraint satsifaction with finite templates. In contrast to finite structures, a general infinite

structure might not have a core, or the core might not be unique. The thesis discusses a generalization of the notion of a core that shares many properties of the core of a finite structure that are used for constraint satisfaction. In particular, we prove that if we add a singleton relation to the signature of an ω-categorical core, we do not change the computational complexity of the corresponding constraint satisfaction problem. For finite cores this is one of the main results in (Bulatov et al., 2003) — our proof is simpler, although it applies to the much larger class of ω-categorical cores.

The following facts suggest that the class of ω-categorical templates is in a certain sense the largest class of structures where we can apply the algebraic approach to constraint satisfaction. One can show that the statement "The first-order definable relations of Γ are precisely the relations that are preserved by all automorphisms of Γ" is true if and only if Γ is ω-categorical. The existence of a canonical Datalog program of width k for a constraint satisfaction problem CSP(Γ), for all k, also implies that Γ is ω-categorical.

Forbidden induced substructures. In many cases an ω-categorical structure Γ can be described by a finite set of finite forbidden induced substructures – this holds for instance for the examples listed above. In this case the constraint satisfaction problem for Γ is contained in the class of computational problems called *monotone SNP*, which is a fragment of existential second order logic introduced in the context of constraint satisfaction in (Feder and Vardi, 1999). Conversely, we can use a recent result of (Cherlin, Shelah, and Shi, 1999) to show that every problem in the class *monotone monadic SNP* – another class introduced by Feder and Vardi – can be formulated as a constraint satisfaction problem with an ω-categorical template.

Many examples of ω-categorical structures come from the class of *homogeneous* structures. Homogeneous digraphs have been classified by (Cherlin, 1998). We determine the complexity of the constraint satisfaction problem in all cases where the template is a homogeneous digraph that is described by a finite number of finite forbidden induced substructures. All these problems are either tractable or NP-complete.

Tree-like templates. Finally, we study several concrete constraint satisfaction problems that have an ω-categorical *tree-like* template. Some of these problems have applications in computational linguistics (Koller et al., 2000) and computational biology (Gusfield, 1997). We present graph algorithms that solve these problems in polynomial time. In particular we present an algorithm with quadratic running time solving a problem posed in (Cornell, 1994); the algorithm is of a different type and not from one of the classes

of algorithms typically used for constraint satisfaction with finite domains. In particular, we show that the problem cannot be solved with the path-consistency technique.

The central algorithmic idea is the notion of a *free set of nodes* in a satisfiable constraint graph. It turns out that such sets can be computed efficiently using connectivity algorithms. The free sets then guide a decomposition strategy that allows to built a solution from recursively obtained subsolutions. Using similar algorithmic ideas we also derive a new and more efficient algorithm for *normal dominance constraints* (Althaus et al., 2001), which is a constraint language used in computational linguistics. Subquadratic running time can be achieved using decremental graph connectivity algorithms.

Published results. Some of the results concerning the general framework of constraint satisfaction with ω-categorical templates and the complexity classification for constraint satisfaction with homogeneous digraphs were published in (B. and Nešetřil, CSL 2003). A predecessor of the algorithm for Cornell's tree description language appeared in (B. and Kutz, STACS 2002). The new algorithm on normal dominance constraints is published in (B. et al., SODA 2004). The results on cores of ω-categorical structures are contained in (B., STACS 2005).