Notes AIM CSP workshop.

Please don't take my notes for granted, I don't feel very confortable with many of the details, being an outsider to the field. So, it's not unlikely that I made some mistake somewhere. It would be good if somebody will go through it and check for blatant mistakes. -b.

1 Monday

1.1 Tutorials by Benoit Larose (Algebra I) and Phokion Kolaitis (Logic I)

1.2 Problem generating session (moderator: Moshe Vardi)

The aim of this session is to collect working problems, rather than just open problems that one would like to see solved.

1. Width hierarchy
   (a) Clarify the cacophony of notions of width
   (b) Find candidate counterexamples
   (c) Relationships with arity
   (d) Robustness of notions of width. In particular, preservation under pp-definitions
2. Algorithmic classification
   (a) Decidability of width k. (Perhaps the following powerful decidability result can be of use: containment of Datalog queries in unions of conjunctive queries. Follows from Courcelles theorem by bounded treewidth)
   (b) Decidability of bounded width
   (c) Even without solving either of the above, establish an effective upper bound on width, so that a positive solution to the first would imply a positive solution to the second
   (d) Relationship with bounded treewidth machinery
   (e) Complexity of omitting types (is related to bounded with by an algebraic conjecture). Known bounds: NP-hard; in Π_p^2.
   (f) Counting CSP: complexity of classification (dichotomy known)
3. Special classes of instances
   (a) In particular, CSP(Planar, B) and CSP(Outdegree¡=k, B)
   (b) Are algebraic techniques still applicable?

(c) Does dichotomy becomes easier or harder?

(d) Even the Boolean case is interesting

4. Beyond CD(4)

   (the issue is to tell for B in CD(k) whether CSP(B) is expressible in Datalog. Solved for k=4.)

   (a) Alternative proofs for CD(4) might have a better chance of generalizing to ¿4

   (b) Relationship with bounded width

   (c) Does the hierarchy for CD relate to width hierarchy? (source of candidates?)

5. Fragments

   (a) Lower complexity classes

   (b) Datalog fragments

   (c) Corresponding algebra (some necessary conditions are known but few sufficient conditions)

   (d) nuf terms versus Jonsson terms

6. Various other suggestions:

   - Local versus global tractability. If every finite subset of an infinite constraint language is tractable, then is the whole constraint language tractable? We know that a negative answer would imply that there is a infinite constraint language (infinite set of relations over finite domain) with intermediate complexity (in the global sense).
   - Candidate structures: templates that omit types 1 and/or 2 with unknown width (in general and on digraphs); tractable CSPs with no known bound on width
   - Analysis of probabilistic templates
   - Assuming some conjectures such as the algebraic conjecture concerning bounded width, can we use them as a black box for proving the dichotomy conjecture?
   - Does taking products preserve tractability? Interesting mostly in its own right, though a negative answer would have strong implications.

7. Possible tutorials (+ how many people would be disappointed if they would have to miss it)

   - The digraphs without sources and sinks result (8)
   - Datalog: more concrete techniques, e.g., for proving width lower bounds (11)

- Counting CSP (7)
- Duality theory (5)
- Results of Bulatov et al about edge colored graphs where colors correspond to types?? (13)

2 Tuesday

2.1 Tutorials by Matt Valeriote (Algebra II) and Anuj Dawar (Logic II)

2.2 Working group reports

Albert Atserias led a small Q&A session on Datalog for people who are less familiar with this topic. Other groups focussed on specific open problems. The other groups focussed on specific open problems (concerning the width hierarchy, algorithmic classification problems, and restricted classes of structures). At the end of the day, the working groups reported to each other:

Algorithmic classification (Moshe Vardi reports) The group looked at PP-definability, and more concretely the following problem:

> Given a structure $\mathbf{B}$ and a relation R, is R PP-definable in $\mathbf{B}$?

Consider the uniform version of this question (everything is part of the input). In the boolean case, this problem can be solved in polynomial time. The group has formulated a conjecture, namely that in general and already on 3 valued domains, the problem is co-NExpTime-complete. The upper bound is easily obtained ("guess the polymorphism"), for the lower bound perhaps a tiling argument can be used.

The group also looked at another question: we know that $datalog \models MSO$ is decidable, and $datalog \models datalog$ is undecidable. How about entailments between SNP and Datalog? No new results are to be reported on this yet.

Restricted classes of structures (Martin Grohe reports) The working group focussed on the case of planar graphs. Examples of tractable CSPs in this context: K_1, K_2, and K_n for $n \geq 4$. Examples of NP-hard CSPs: K_3, odd cycles, Penny graph, Odd wheels. Are there tractaible planar templates that are not homomorphically equivalent to $K_1, K_2, K_4, \ldots$?

A systematic analysis faces many technical difficulties. For one, the usual arguments use relations of increasing arity. It is not clear what would be the right notion of a planar hypergraph in this context. And, if we restrict attention to binary relations only, it is not clear whether the polymorphisms still determine the complexity. Finally, note that PP-definable relations on planar graphs are not necessarily planar.

Width hierarchy (Miklos Maroti reports) An attempt was made to compare and unify the different notions width (phrased in terms of relational structures). No new results were obtained yet.

I've left out the concrete details here, as I noticed that there was a mistake in the definition of k-minimal strategies, so it would only contribute to the cacophony.

3 Wednesday

3.1 Tutorials by Eric Allender (Complexity classes below P) and Martin Grohe (treewidth)

Eric gave a nice overview of what is known about complexity classes below PTime. The general theme was "invention versus discovery".

Martin discussed tree width, and its relation to two other notions, namely *linkedness* and *bramble number*. Linkedness makes concrete the intuitive connection between tree width and how densely connected a graph is, while brambles are a convenient tool for establishing lower bounds on the tree with of a graph.

3.2 Working groups reports (Thursday morning)

The Zadori conjecture (Ross Willard reports) The Zadori conjecture says that for every finite structure $\mathbf{B}$, if the associated algebra $\mathcal{A}(\mathbf{B})$ has Jónsson terms then it has an NUF. The working group looked at the special case where the associated algebra is conservative, $\{0,1\} \subseteq \mathbf{B}$ and there is no NUF that restricts to this two-element subset. No new results were obtained yet.

MM_1SNP (Manuel Bodirski reports) MM_1SNP is the fragment of MMSNP with only one set variable. For given a relational vocabulary τ, the sentences of MM_1SNP are of the form $\exists P \forall \vec{x}.\phi(\vec{x})$, with ϕ a quantifier free formula of $FO[\tau \cup \{P\}]$, in which all occurrences of τ-symbols are negative in ϕ.

The question is: given an MM_1SNP sentence ϕ, what is the complexity of deciding for a finite τ-structure S whether $S \models \phi$?

The working group looked at the question whether the algebraic techniques from CSP can be applied to MM_1SNP problems by encoding them as infinite templates. No new results were obtained yet.

LFP versus Datalog (Moshe Vardi reports) The working group looked at the question whether $LFP \cap \neg CSP \subseteq^? Datalog$. Some starting points were identified:

- $LFP \cap Hom \not\subseteq Datalog$ (a recent result announced by Anuj Dawar).
- modal μ-calculus $\cap$ $\neg$CSP $\subseteq$ Datalog (Moshe Vardi and Albert Aterias said they are reasonably confident that this holds).

- $\exists L^{\omega,+}_{\infty\omega} \cap \neg CSP \subseteq Datalog$ (Kolaitis-Vardi)
- $\exists L^{\omega}_{\infty\omega} \cap Hom \subseteq \exists L^{\omega,+}_{\infty\omega}$ (Feder-Vardi)
- $FO \cap Hom \subseteq \exists FO^{+}$ (Rossman)

No new results were obtained yet.

Edith reports on CSPs below PTime A new result is obtained for quantified Boolean CSPs. Recall that there are two versions of Schaefer's theorem, one that distinguishes between PTime and NP, and one that considers various subclasses of PTime. In the case of Quantified CSPs, the known dichotomy only distinguishes between PTime and NP. What happens inside PTime? The answer: exactly the same picture as in the unquantified case. This can be shown by looking at Hubie Chen's reduction from $QCSP(\mathbf{B})$ to $CSP(\mathbf{B})$ in the PTime case, and notificing that it is an AC^0-reduction. In other words, tractable quantified CSPs are no harder than their unquantified counterparts.

Width hierarchy (?? Petar reports?) Consider the algebraic formulation of relational width. Suppose B is a finite idempotent algebra and H a k-minimal strategy from A to B, but H does not induce a homomorphism from A to B. Then is then the case that one can find a simple $B' \in HS(B)$ already shows and a k-minimal strategy H' from A to B' not inducing a homomorphism from A to B'? It may be possible to prove this, and it would tell us where to look in order to determine relational width.

Width hierarchy (Matt Valeriote reports) The working group tried to find possible indicators for high width. We know that relational width k implies the existence of a k-ary weak NU term, hence, the absence of such a term would imply high relational width. Likewise, relational width 3 can be shown to imply the existence of a term t satisfying the equations $t(yyyxxx) = t(yxxyyx) = t(xyxyxy) = t(xxyxxy)$. No further results can be reported yet.

4 Thursday

4.1 Tutorials by Aldrei Bulatov (edge colored graphs) and Victor Dalmau (duality)

Andrei presented his techniques involving edge colored graphs. Towards the end of his talk, he announced results, which would imply that, if $HS(\mathbb{A}(\mathbf{B}))$ omits types 1,2 and 3, then $CSP(\mathbb{B})$ has relational width 3.

Victor in his talk discussed the four fundamental ways to classify CSPs, makely in terms of logical definability, in terms of obstruction sets, in terms of properties of the associated algebras, and in terms of complexity. Victor gave an overview of classes of CSPs for which several such characterizations are known.

During Victor's talk, Martin Grohe raised a new conjecture related to chassifications in terms of obstruction sets, namely that bounded tree width is the strongest purely graph theoretic condition on obstruction sets that implies tractability.

4.2 Working group reports (friday morning)

LFP versus Datalog (Moshe Vardi reports) In order to show that $LFP \cap \neg CSP \subseteq Datalog$, it is enough to show that whenever $\mathbf{A} \leq_k \mathbf{B}$ (meaning that Duplicator has a winning strategy in the one-sided k-pebble game) then there are structures $\mathbf{A}', \mathbf{B}'$ such that $\mathbf{A} \to \mathbf{A}'$, $\mathbf{B}' \to \mathbf{B}$, and $\mathbf{A}' \equiv_k \mathbf{B}'$ (meaning that Duplicator has a winning strategy in the two-sided k-pebble game). The working group is currently trying to prove the latter.

Zadori conjecture (?? Benoit Larose reports?) The working group continued to work on the Zadori conjecture. A related question is the following:

> If $D = Pol(R_1, \ldots, R_n)$ with $R_1 = \{0, 1\}$, and $C = \{f \restriction \{0,1\} \mid f \in D\}$, then is it the case that C can be presented by finitely many relations?

A negative answer would imply that the Zadori conjecture fails, while a positive answer would show that it holds at least in the 2-convervative case. No new results can be reported yet.

MM$_1$SNP (Manuel Bodirski reports) A partial result is obtained a special case with only 0-obstructions and 1-obstructions, and where the instances are assumed to have no loops. In this case, testing whether a structure satisfies the formula is PTime when the cores of all obstructions have size at most 2, and NP-hard otherwise.

PP-equivalence (Matt Valeriote reports) The PP-equivalence problem for a fixed structure $\mathbb{B}$ takes as input two PP-formulas with the same free variables and asks whether they are equivalent over $\mathbb{B}$. The PP-isorphism problem is similar, but allows for a permutation of the variables. The working group looked at these problems. No results were obtained yet.

5 Friday

5.1 Tutorials by Ross Willard (k-edge tems) and Ralph McKinsey (dichotomy for smooth digraphs)

Ross described a polynomial time algorithm for CSPs that have k-edge terms. This provides a common generalization of CSPs with Mal'tsev terms and those with NU terms.

Ralph gave some insights into the proof of the celebrated dichotomy theorem for digraphs without sources and sinks.

5.2 Working group reports

The PP-definability problem (Matt reports) Continuing on the same problem (given a finite structure and a relation, is the relation PP-definable in the structure), the working group first managed to show that the problem is co-NP-hard by a reduction from 3-colorability, and next in fact PSpace-hard by a reduction from the analogous definability problem for unary functions. Recall that the best known upper bound at present is NExpTime.

Exponential time hypothesis (Martin Grohe reports) There is a well known hypothesis which says that no algorithm can solve 3SAT in time $2^{o(n)}$, i.e., in less than exponential time. This is known as the *exponential time hypothesis* (ETH), and is commonly used as an assumption in lowerbound proofs. The working group showed that, assuming EHT, for any structure $\mathbf{B}$, if the associated algebra $\mathbb{A}(\mathbf{B})$ omits the unary type, then $CSP(\mathbf{B})$ cannot be solved in time $2^{o(n)}$. The main work was to check that there are *linear* reductions from 3SAT.

LFP versus datalog (Anuj Dawar reports) The working group continued to work on the same question —does $LFP \cap \neg CSP \subseteq Datalog$– and obtained a positive result in the case of arbitrary, not necessarily finite, structures. The idea is to show that, whenever $A \leq_k B$ (i.e., Duplicator has a winning strategy in the one sided k-pebble game) then there is a (possibly infinite) homomorphic extension A' of A and a (possible infinite) homomorphic preimage B' of B such that $A' \equiv_k B'$ (i.e., Duplicator has a winning strategy in the two-sided k-pebble game).

$\mathbf{MM_1SNP}$ problem (Manuel Bodirski reports) A complete classification is presented for the following special case: given a set $S \subseteq \{1, \ldots, k\}$, determine for which graphs $G = (V, E)$ it is the case that G can be 2-colored such that on every (directed) k-path the number of 1's belongs to S.

5.3 Final wrap-up

There was a brief discussion as to whether it makes sense to set up a Wiki with open problems. The opinions seemed mostly negative, as such a webpage would likely quickly become outdated. On the other hand, seems to be some interest for a central annotated bibliography.