SIGACT News Complexity Theory Column 49

Lane A. Hemaspaandra
Dept. of Computer Science, University of Rochester
Rochester, NY 14627, USA lane@cs.rochester.edu

Introduction to Complexity Theory Column 49

Warmest thanks to Neil Immerman for this issue's column, which is on Progress in Descriptive Complexity. (As you can see in this column, Neil is the rare writer whose italic lines truly deserve the italics.) This issue should reach your hands shortly before winter intersession, and I'd highly recommend that, if you don't already own Neil's wonderful book *Descriptive Complexity* ([I] in this column's bibliography), you consider visiting your school's library and bringing it home for holidays.

Good news: A new conference—Theory and Applications of Models of Computation (TAMC)—is joining the constellation of theory meetings. From that conference's web site, http://gcl.iscas.ac.cn/accl06/TAMC06_Home.htm: "TAMC06 is a new annual conference focusing on theory and applications of computation. It is organized as part of the Grand China NSF International Joint Project after which the conference is named... The scope of the conference will include algorithms, complexity, models of computation, and computability. The conference will... bring together researchers and students with an interest in computer science, mathematics and logic, and applications to the physical sciences."

Upcoming articles include Piotr Faliszewski and me on *Open Questions in the Theory of Semifeasible Computation*, Oded Goldreich on *Bravely, Moderately: A Common Theme in Four Recent Results*, and Lance Fortnow and Rahul Santhanam on hierarchies of semantic classes.

Guest Column: Progress in Descriptive Complexity[1]

Neil Immerman[2]

Abstract

The *raison d'être* of descriptive complexity is the following

> *The Computational Complexity of checking if an input has a property can be understood as the Logical Complexity needed to describe the property.*

In this guest column I give an overview of descriptive complexity, and survey some current progress and directions.

[2]Dept. of Computer Science, UMass, Amherst, MA 01003, USA, immerman@cs.umass.edu. Supported by NSF Grant CCF-0514621.

The Spectrum Problem and Fagin's Theorem

In August, 2005, I attended a workshop at Oxford University celebrating the fiftieth anniversary of the Asser's Spectrum Problem (www.univ-paris12.fr/lacl/durand/workshop.html). This meeting gave me the opportunity to take an historic view in my talk at the workshop and in this related guest column.

The **spectrum** of a first-order formula, φ, is the set of cardinalities of finite models of φ. I write $|\mathcal{A}|$ for the universe of structure $\mathcal{A}$, and $\|\mathcal{A}\|$ for the cardinality of its universe. I will only consider finite structures throughout, thus,

$$\mathrm{spec}(\varphi) \quad = \quad \{\|\mathcal{A}\| \mid \mathcal{A} \models \varphi\} \; .$$

In 1952, Scholz posed the problem of characterizing the set of first-order spectra [S]. In 1955, Asser asked whether the set of spectra is closed under complementation [A]. That is, is it the case that for every first-order spectrum $S = \mathrm{spec}(\varphi)$, there is another first-order formula, ψ, such that $\overline{S} = \mathbf{Z}^+ - S = \mathrm{spec}(\psi)$?

In the early 1970's, the set of spectra was characterized as the sets of positive integers (input as usual in binary) that can be accepted in $\mathbf{NE} = \mathbf{NTIME}[2^{O(n)}]$. This was proved independently by three groups of researchers: Alan Selman and Neil Jones, Ron Fagin, and Claude Christen[3]. Thus, fifty years after it was stated, Asser's spectrum problem remains open and is equivalent to the question whether $\mathbf{NE} = \text{co-}\mathbf{NE}$.

Fagin presented his result in a particularly captivating way. He defined a **generalized spectrum** to be the set of finite structures of some vocabulary, Σ, that can be expanded to be a model of some first-order formula, φ, by interpreting any new relation symbols that occur in φ but not in Σ. Fagin proved that the set of generalized spectra is exactly equal to the computational class **NP**.

In descriptive complexity, we think of a (computational decision) **problem**, S, as a set of finite structures of some vocabulary, Σ, i.e., $S \subseteq \mathrm{STRUC}_{\mathrm{fin}}[\Sigma]$. Recall that in first-order logic we quantify over the elements of the universe, whereas in second-order logic we may also quantify over relations over the universe. SO$\exists$ denotes those second-order sentences that consist of some second-order existential quantifiers followed by a first-order formula. For example, the following formula, $\Phi \in \mathrm{SO}\exists$, expresses the **NP**-complete property, three-colorability of graphs:

$$\begin{aligned}\Phi \quad \equiv \quad & \exists R^1\, Y^1\, B^1\, \forall x\, y\, \Big((R(x) \vee Y(x) \vee B(x)) \;\wedge \\ & \quad (E(x,y) \rightarrow (\neg(R(x) \wedge R(y)) \;\wedge\; \neg(Y(x) \wedge Y(y)) \;\wedge\; \neg(B(x) \wedge B(y))))\Big)\end{aligned}$$

Φ says that there exist three unary relations, i.e., subsets of the universe, red, yellow, and blue, such that every vertex belongs to at least one of these subsets and no two adjacent vertices belong to the same subset. (In the sequel we will sometimes consider the arity of the quantified relations. In Φ's case these are all monadic, so we say that $\Phi \in \mathrm{SO}\exists$(monadic).)

[3] Christen considered himself scooped by the other researchers and didn't publish. I heard about his contribution for the first time in Janos Makowsky's talk at the spectrum workshop.

My favorite statement of Fagin's theorem is, "$\mathrm{SO}\exists = \mathbf{NP}$", i.e., a problem is recognizable in **NP** iff it is describable as a second-order, existential sentence. The amazing insight is that an important complexity class can thus be characterized via a simple and natural logical language. Fagin's theorem was the beginning of descriptive complexity.

Descriptive Complexity

All important complexity classes have natural descriptive characterizations. Here are two of my favorites. First,

Theorem: [IB] For $v = 1, 2, \ldots,\quad \mathbf{DSPACE}[n^v] \;=\; \mathbf{VAR}[v+1]$

Here $\mathbf{VAR}[k]$ is the set of problems expressible by a sequence of first-order sentences that may use (and reuse) at most k distinct variables. Intuitively, since each variable ranges over the n-element universe, v variables can express about n^v memory locations. Different parallel machines have different connection patterns, so there is nothing as precise as the above theorem, but it is still true to say that v variables corresponds to about n^v gates:

> *Number of variables corresponds to amount of hardware and, in particular, a bounded number of variables corresponds to polynomially much hardware.*

A second fundamental descriptive principle is that,

> *Parallel time, inductive depth, and quantifier depth all coincide.*

More specifically:

Theorem: [I2]

1. For all constructible, polynomially bounded $t(n)$,
$$\mathbf{CRAM}[t(n)] \;=\; \mathbf{IND}[t(n)] \;=\; \mathbf{FO}[t(n)]$$
2. For all $t(n)$, even beyond polynomial, $\quad \mathbf{CRAM}[t(n)] \quad = \quad \mathbf{FO}[t(n)]$

Here are a few definitions to explain the above theorem.

$\mathbf{CRAM}[t(n)]$ is the set of problems recognizable by a concurrent parallel random access machine using polynomially much hardware, i.e., memory and processors, in parallel time $O(t(n))$.

$\mathbf{IND}[t(n)]$ is the set of problems describable by first-order inductive definitions that close in $t(n)$ steps. For example, consider the following inductive definition of the transitive closure of the relation, E:

$$E^+(x,y) \quad \equiv \quad E(x,y) \;\vee\; \exists z(E^+(x,z) \wedge E^+(z,y))\ .$$

Note that the above inductive definition closes in $1 + \log n$ steps for graphs of n vertices. Thus the graph reachability problem, REACH $= \{G = (V, E, s, t) \mid G \models E^+(s,t)\}$ is in $\mathbf{IND}[\log n]$.

LFP is a least fixed-point operator which formalizes the definition of new relations by induction. Continuing the above example, let

$$\varphi(R, x, y) \quad \equiv \quad E(x, y) \ \vee \ \exists z(R(x, z) \wedge R(z, y)) \ .$$

Thus φ is a monotone operator taking binary relations to binary relations. It's least fixed point, $\mathbf{LFP}(\varphi)$, is the transitive closure, E^+. Thus, for example, $\mathbf{LFP}(\varphi)(s,t)$ expresses REACH $\in \mathbf{IND}[\log n] \subseteq \mathbf{FO}(\mathbf{LFP})$.

$\mathbf{FO}[t(n)]$ is the set of problems describable by a block of restricted quantifiers that is repeated $t(n)$ times for structures of universe size n. More explicitly, a quantifier block is a formula fragment, $\mathrm{QB} = [(Q_1 x_1.M_1) \cdots (Q_k x_k.M_k)]$, where each M_i is quantifier-free. Here $(\forall x.M)\varphi$ means $\forall x(M \rightarrow \varphi)$ and $(\exists x.M)\varphi$ means $\exists x(M \wedge \varphi)$. The formula expressing the $\mathbf{FO}[t(n)]$ property for structures of size at most n is

$$\varphi_n \quad = \quad \underbrace{[\mathrm{QB}][\mathrm{QB}] \cdots [\mathrm{QB}]}_{t(n)} M_0$$

Continuing the above example, we can rewrite the above monotone operator, φ, as follows: $\varphi(R, x, y) \ \equiv \ [\mathrm{QB}_\varphi](x, y)$, where, $\mathrm{QB}_\varphi \ \equiv \ (\forall z.M_1)(\exists z)(\forall uv.M_2)(\exists xy.M_3)$, and

$$M_1 \ \equiv \ \neg E(x, y); \quad M_2 \ \equiv \ (u = x \wedge v = z) \vee (u = z \wedge v = y); \ \text{ and, } \ M_3 \ \equiv \ (x = u \wedge y = v) \ .$$

Applying the operator φ is equivalent to writing the quantifier block, QB_φ. Thus,

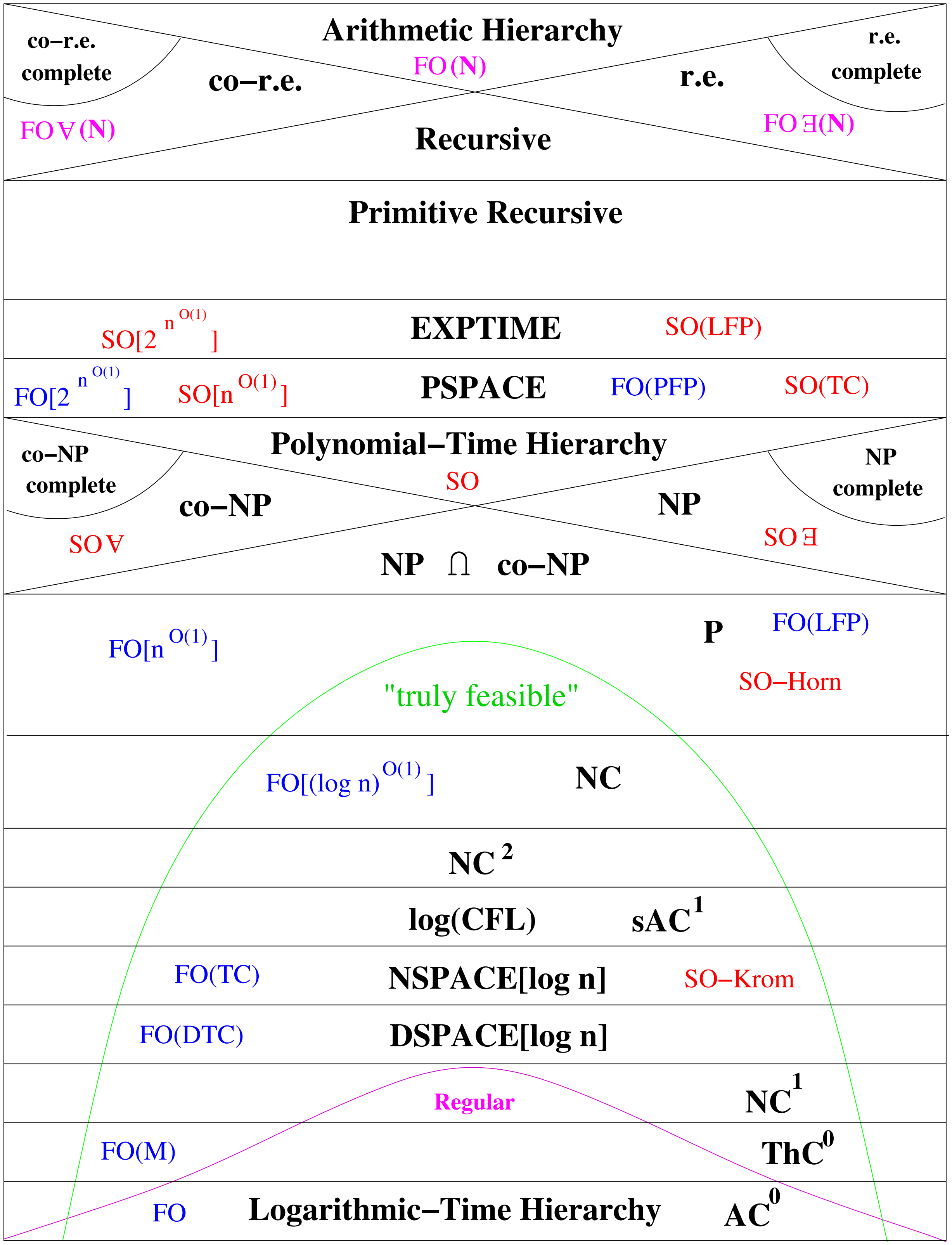
$$\mathbf{LFP}(\varphi) \quad \equiv \quad [\mathrm{QB}_\varphi]^{1+\log n}(x \neq x) \ .$$

While **LFP** is defined over monotone operators and thus closes in polynomial steps, **PFP** is a partial fixed-point operator, i.e., an arbitrary iteration operator which may not close in fewer than exponentially many steps. Thus from the previous theorem we have the following,

$$\begin{array}{lclclclcl}
\mathbf{FO}(\mathbf{LFP}) & = & \bigcup_{k=1}^{\infty} \mathbf{CRAM}[n^k] & = & \bigcup_{k=1}^{\infty} \mathbf{IND}[n^k] & = & \bigcup_{k=1}^{\infty} \mathbf{FO}[n^k] & = & \mathbf{P} \\
\mathbf{FO}(\mathbf{PFP}) & = & \bigcup_{k=1}^{\infty} \mathbf{CRAM}[2^{n^k}] & = & \bigcup_{k=1}^{\infty} \mathbf{FO}[2^{n^k}] & = & \mathbf{PSPACE} & &
\end{array}$$

The Descriptive World Diagram gives a view of most important computational complexity classes. The class, "truly feasible", does not have a precise definition but it represents those problems for which we can compute the exact answer for all moderately-sized inputs using an affordable amount of time and hardware.

Part of the power of computational complexity is that the mathematically precise complexity classes tend to provide good characterizations of the issues that we care about in



The Descriptive World

practice. In particular, **P** is a good mathematical wrapper for the truly feasible problems in the sense that most "natural" problems in **P** are in fact truly feasible. Of course it is easy to present unnatural problems that are in **P** but require, e.g., time n^{1000} and so are infeasible.

Much practical work in algorithms and complexity consists in precisely defining problems and ascertaining which complexity class they are complete for. If the problem is above **P** then we must search for a simplification or an approximate solution to the problem that is itself feasible.

Uniformity

In the circuit model of computation, a family of boolean circuits, C_n, $n = 1, 2, \ldots$ recognizes problems represented as n-bit inputs. The depth of the circuit is parallel time and the width is the amount of hardware. We say that the circuit family is **uniform** if the circuits all consist of one program on varying size inputs. However, this cannot be defined formally for circuits so we instead require that the map from 1^n to C_n is very easy to compute.

A problem in $\mathbf{FO}[t(n)]$ corresponds to a **syntactically uniform** family of $\mathbf{AC}[t(n)]$ circuits and so the artificial definition of uniformity is obviated. Note that the class $\mathbf{SO}[t(n)]$ consists of circuit problems of $O(t(n))$ depth but exponential width. Let $\mathbf{CRAMEH}[t(n)]$ consist of those problems solved by CRAM's running in parallel time $O(t(n))$ but using exponentially much hardware.

The following equalities are not new, but they still amaze me for the trade-offs that they hint at between parallel time and hardware and for the fact that the mysteries about the relationships between **LOGSPACE**, **P**, **NP**, and **PSPACE**, could all be understood by understanding the power gained by repeatedly writing down a quantifier block.

$$
\begin{array}{ccccccc}
\mathbf{LH} & = & \mathbf{FO} & = & \mathbf{CRAM}[1] & = & \mathbf{AC}^0 \\
 & & \mathbf{FO}[\log n] & = & \mathbf{CRAM}[\log n] & = & \mathbf{AC}^1 \\
 & & \mathbf{FO}[\log^{O(1)} n] & = & \mathbf{CRAM}[\log^{O(1)} n] & = & \mathbf{NC} \\
\mathbf{FO}(\mathbf{LFP}) & = & \mathbf{FO}[n^{O(1)}] & = & \mathbf{CRAM}[n^{O(1)}] & = & \mathbf{P} \\
\mathbf{FO}(\mathbf{PFP}) & = & \mathbf{FO}[2^{n^{O(1)}}] & = & \mathbf{CRAM}[2^{n^{O(1)}}] & = & \mathbf{PSPACE} \\
 & & & & & & \\
\mathbf{PH} & = & \mathbf{SO} & = & \mathbf{CRAMEH}[1] & = & \mathbf{PH} \\
\mathbf{SO}(\mathbf{TC}) & = & \mathbf{SO}[n^{O(1)}] & = & \mathbf{CRAMEH}[n^{O(1)}] & = & \mathbf{PSPACE} \\
\mathbf{SO}(\mathbf{LFP}) & = & \mathbf{SO}[2^{n^{O(1)}}] & = & \mathbf{CRAMEH}[2^{n^{O(1)}}] & = & \mathbf{EXPTIME} \\
\mathbf{SO}(\mathbf{PFP}) & = & \mathbf{SO}[2^{2^{n^{O(1)}}}] & = & \mathbf{CRAMEH}[2^{2^{n^{O(1)}}}] & = & \mathbf{EXPSPACE}
\end{array}
$$

In the above, **LH** is the logarithmic-time hierarchy, **PH** is the polynomial-time hierarchy and **TC** is the transitive closure operator. The class **SO**(**LFP**) is analogous to **FO**(**LFP**) but **LFP** may be used to inductively define relations on second-order variables.

In 1987 Steve Homer asked if there is a "complete quantifier block", QB_c, in the sense that for all functions $t(n)$, the problem consisting of QB_c iterated $t(n)$ times is complete for $\mathbf{FO}[t(n)]$. The answer is, "yes", as was shown in [HI]. Thus, the above fundamental questions in complexity could be understood if we could determine the power of repeatedly writing down the particular quantifier block QB_c.

We Need an Ordering on the Universe

The above first-order descriptive characterizations of computational complexity classes assume that there is an available total ordering on the universe. For example, for a graph, $G = (V, E)$, its universe is assumed to be $V = \{0, 1, \ldots, n-1\}$ and the usual ordering, $\leq$, on V is given.

> *An unordered graph makes sense mathematically, but you can't store such an object in a computer.*

If you remove the ordering then the first-order descriptive characterizations fail. For example, we cannot even cycle through the vertices of a graph, counting them mod 2. Fagin, Jones, and Selman didn't run into this problem because in SO∃ you can guess an ordering (unless we are dealing with SO∃(monadic)).

Ehrenfeucht-Fraïssé Games

Ehrenfeucht-Fraïssé games are useful combinatorial games for proving lower bounds on what is expressible in a given quantifier rank and with a given number of variables. (Quantifier rank is the depth of nesting of quantifiers.) These games have been used to prove many beautiful lower bounds without the ordering, and thus separate most descriptive classes without ordering. The following are a few such results among others:

- Grädel and McColm proved lower bounds separating the relative power of the operators: **DTC**, **TC**, and **LFP** [GM].
- Grohe proved tight arity hierarchies for **TC** and **LFP** [G].
- Etessami and I proved lower bounds with counting and one-way local orderings [EI, EI2].
- Libkin and several coauthors introduced new methods for proving that certain logics can only express local properties. See [LN] for a survey.
- Grohe and Schwentick proved the locality of order-independent first-order queries [GS2].

One Historical Thread

I have a vivid memory of reading Fagin's papers [F, F2] in the winter of 1977-78, when I was a graduate student at Cornell. Fagin hinted at a possible proof that **NP** is not closed under complementation by proving via Ehrenfeucht-Fraïssé games that graph connectivity is not in SO∃(monadic). It is easy to see that the complementary problem, non-connectivity, is in SO∃(monadic) because we can exhibit a non-empty set of vertices that is closed under adjacency, but does not include all vertices.

I despaired of proving second-order lower bounds beyond monadic, and turned instead to examining computational problems expressed via sequences of first-order sentences. This became my Ph.D. thesis.

Meanwhile, others extended Fagin's lower bound on connectivity. First, de Rougemont proved that connectivity is still not expressible in SO∃(monadic) in the presence of a successor relation [dR]. Later, Schwentick proved that this holds true in the presence of an ordering relation on the universe [S2]. Schwentick's proof, a precursor of the Grohe-Schwentick paper mentioned above, was a tour de force. The difficulty is that in the presence of ordering, every two vertices are distance at most one in the Gaifman graph, because they are related by the ordering relation. (The Gaifman graph places an edge between two elements of the universe if they occur in this same tuple of some relation. In locality arguments, local means close in the Gaifman graph.)

The Schwentick lower bound still does not say anything about the nondeterministic time needed to check graph connectivity. For this, we need to look at Lynch's Theorem: $\mathbf{NTIME}[n^k] \subseteq \mathrm{SO}\exists(+)(\text{arity } k)$ [Ly]. Fagin had proved that $\mathbf{NTIME}[n^k] \subseteq \mathrm{SO}\exists(\text{arity } 2k)$. Lynch tightened this to the optimal arity k, in the presence of the addition relation. Thus, if Schwentick's lower bound could be extended to include addition, it would follow that connectivity cannot be checked in nondeterministic linear time on a Turing machine.

Abiteboul and Vianu

Twenty-five years ago I used Ehrenfeucht-Fraïssé games to prove the following lower bounds:

1. $\mathrm{CVP} \in \mathbf{FO}(\mathrm{wo}\leq)[n] - \mathrm{qr}(\mathrm{wo}\leq)[2^{\sqrt{\log n}}]$

2. $\mathrm{CLIQUE}[k] \in \mathbf{VAR}(\mathrm{wo}\leq)[k] - \mathbf{VAR}(\mathrm{wo}\leq)[k-1]$

(1) says that the **P**-complete circuit value problem (CVP) which can be expressed in $\mathbf{FO}[n]$ without use of the ordering relation, cannot be expressed in quantifier rank $2^{\sqrt{\log n}}$ without the ordering. If this result went through with ordering it would separate **P** from **NC**, and even from $\mathbf{SPACE}[(\log n)^k]$ for all k.

(2) Says that while the existence of a complete subgraph of size k can be expressed very simply without ordering, using k variables, $x_1, \ldots, x_k$, this is impossible with fewer variables. If this result went through with ordering, it would follow that $\mathbf{P} \neq \mathbf{NP}$.

As mentioned above, many other lower bounds on languages without ordering have been proved. However, one place we couldn't get such lower bounds was for $\mathbf{FO}(\mathrm{wo}\leq)(\mathbf{LFP})$ vs. $\mathbf{FO}(\mathrm{wo}\leq)(\mathbf{PFP})$, the languages without ordering that correspond to **P** and **PSPACE**.

Abiteboul and Vianu explained why:

Theorem: [AV] $\mathbf{FO}(\mathrm{wo}\leq)(\mathbf{LFP}) = \mathbf{FO}(\mathrm{wo}\leq)(\mathbf{PFP}) \Leftrightarrow \mathbf{P} = \mathbf{PSPACE}$

In their proof, Abiteboul and Vianu showed that $\mathbf{FO}(\mathrm{wo}\leq)(\mathbf{LFP})$ suffices to capture the "essence of a given unordered structure".

Order-Independent P

While $\mathbf{FO}(\mathbf{LFP})$ captures **P**, it includes many problems that depend on the arbitrary ordering in which the input is presented. For example, we can easily test in polynomial time whether there is an edge from the first vertex to the last vertex. That is an order-dependent problem. Natural graph problems such as connectivity, three-colorability, etc. are order independent: they do not depend on the order in which the vertices are presented.

A long-standing open problem of great interest is whether there is a descriptive language that captures order-independent **P**, i.e., $\{S \in \mathbf{P} \mid \forall \mathcal{A}, \mathcal{B}(\mathcal{A} \cong \mathcal{B} \Rightarrow (\mathcal{A} \in S \Leftrightarrow \mathcal{B} \in S))\}$.

Note that given a Turing machine, M, it is undecidable whether M runs in polynomial time on all inputs. However, using clocks, it is easy to give a recursively enumerable (r.e.) listing of Turing machines accepting exactly all polynomial-time problems.

It is open whether or not there is such an r.e. listing of order independent **P**. Dawar observed that these two questions are equivalent: There is a language for order-independent **P** iff there is an r.e. listing of order-independent **P** [D].

Another interesting observation is that finding a language for order-independent **P** is closely related to the graph isomorphism problem. In particular,

Observation: If there is a polynomial-time graph canonization algorithm, then there is a language for order-independent **P**.

The proof is that all order-independent **P** problems would then be of the form: first convert the input graph to its canonical form, and then run a timed, polynomial-time algorithm on the canonical graph.

The language $\mathbf{FO}(\mathrm{wo}\leq)(\mathbf{LFP})$ is strictly contained in order-independent **P**. In particular, testing whether or not there is an even number of edges is not in $\mathbf{FO}(\mathrm{wo}\leq)(\mathbf{LFP})$.

It thus was natural to consider the language $\mathbf{FO}(\mathbf{LFP}, \mathbf{COUNT})$ which considers a separate domain of natural numbers, with the usual ordering, of cardinality the same as the number of vertices, and connected to the graph via counting quantifiers.

Lander and I considered comparing graphs using $\mathbf{FO}(\mathbf{LFP}, \mathbf{COUNT})$ [IL]:

Observation: Given a graph, G, on n vertices and a number k we can write a canonical description of the $vc(k)$-type of G, i.e., it's complete k-variable description in $\mathbf{FO}(\mathbf{COUNT})$, in time $n^k \log n$.

The $vc(k)$-type of a graph corresponds to the stable coloring of its $k-1$ tuples. In particular, stable vertex coloring computes exactly the $vc(2)$-type. (To compute the stable vertex coloring of a graph, start by coloring all vertices black. Then color them according to how many black neighbors they have. Then color them according to how many neighbors of each color they have, etc.) Babai and Kučera proved that almost all random graphs have a linear-time canonization algorithm [BK]:

Theorem: With very high probability, every vertex in a random graph has a unique stable color after three iterations, i.e, a unique name of qr(3) in $vc(2)$.

There was a large body of work asking whether stable coloring of k-tuples for a fixed k would give a polynomial-time general graph isomorphism algorithm: the Weisfeiler-Lehman approach [W]. Even $k = 3$, i.e., $vc(4)$ had not been ruled out. Lander and I had shown that $vc(3)$ sufficed to characterize graphs of color class 3. However, Cai, Fürer and I showed that this approach fails badly for color class 4 [CF]:

Theorem: Graphs of color class size 4 require $vc(\Omega(n))$ for their characterization.

Since any formula in $\mathbf{FO}(\mathbf{LFP}, \mathbf{COUNT})$ has a fixed number of variables, this both killed the the Weisfeiler-Lehman approach, and showed that a very simple, polynomial-time order-independent property is not in $\mathbf{FO}(\mathbf{LFP}, \mathbf{COUNT})$.

Blass and Gurevich have studied choiceless **P**, an order-independent language that arises naturally from their work on abstract state machines. Choiceless **P** is provably stronger than $\mathbf{FO}(\mathbf{LFP}, \mathbf{COUNT})$. In fact, Dawar, Richerby and Rossman have very recently shown that the property used in the above lower bound theorem is expressible in Choiceless **P** [DR].

Size Lower Bound Game

Ehrenfeucht-Fraïssé games only prove lower bounds on quantifier depth. However, with ordering, quantifier depth $2 + \log n$ suffices to express any graph property for graphs on n vertices.

Let $G = (V, E)$ be such an ordered graph. In $\mathrm{qr}(\log n)$ we can write the properties $\#_i(x)$, $i = 1, \ldots n$, meaning that vertex x is number i in the ordering. Then a complete description, φ_G, of the ordered graph G can be written in $\mathrm{qr}(2 + \log n)$ as follows:

$$\begin{aligned}\varphi_G \quad \equiv \quad & \bigwedge_{\langle i,j\rangle \in E} \exists x, y(\#_i(x) \wedge \#_j(y) \wedge E(x,y)) \quad \wedge \\ & \bigwedge_{\langle i,j\rangle \notin E} \exists x, y(\#_i(x) \wedge \#_j(y) \wedge \neg E(x,y))\end{aligned}$$

Even worse, let S be any set of graphs on n vertices. The membership property for S is also expressible in $\mathrm{qr}(2 + \log n)$:

$$\varphi_S \equiv \bigvee_{G \in S} \varphi_G \; .$$

Thus, there is no way to prove useful lower bounds above $\mathrm{qr}(\log n)$ for ordered graphs.

Adler and I defined a game with which we can prove lower bounds on the **size** of a formula needed to express a property rather than its quantifier rank. The idea is that unlike Ehrenfeucht-Fraïssé games, we play on a pair of sets of structures, A, B. We determine the size of the smallest sentence true of all of A and none of B.

We used this game to prove an optimal succinctness result for temporal logics [AI]. Similar games were used earlier by Karchmer and Wigderson to prove lower bounds on monotone circuit depth [KW]. More recent size lower bounds have been proved by Grohe and Schweikardt [GS]. I believe that this is a difficult but important and promising direction for future work.

More Than I Have Time or Space For

There are many other recent directions that I wish I had time and space to do justice to. Here I just mention one large area that I skipped, plus two beautiful papers:

- Constraint Satisfaction Problems: there is a great deal of exciting work in this area, some of which applies methods of descriptive complexity to help characterize which classes of CSP problems are tractable. I want to thank Phokion Kolaitis for introducing me to some of this work in a series of beautiful lectures he gave at at workshop a few years ago in Barbados. To find out more, I recommend his tutorial slides: [K]. I would also recommend [G2], and – if possible – go to the following workshop: [JK].

- Grohe and Frick proved a beautiful lower bound on the model checking problem for second-order logic on words [GF]. This was known to be fixed-parameter tractable: each fixed second-order sentence is checkable in linear time. Grohe and Frick prove that if $\mathbf{P} \neq \mathbf{NP}$ then the constants in these running times – for any such sequence of polynomial-time algorithms – must be non-elementary!

- Rossman this year gave a lovely proof of a long standing open problem in finite model theory: he showed that a first-order formula is preserved under homomorphisms (on finite structures) iff it is equivalent (with respect to finite structures) to an existential positive formula [R].

In summary, I have tried to give an overview of descriptive complexity. This is an area with many elegant constructions and deep questions together with a large amount of recent progress. There remains much more to be done.

Thanks, and General Appeal

Thank you to Lane for inviting me to write this guest column. Thank you to Ron Fagin for helpful comments. I wish that I had more time to catch up and then stay up on more of the current work in this area. If I have neglected to mention your paper and/or I described it in a wrong/misleading way, please know that this was due to haste and ignorance. If you would be willing to email me a copy of your relevant papers and/or comments and corrections concerning this guest column, I would be grateful.

Further Reading

For further reading I would first suggest consulting the following books on finite model theory and descriptive complexity: [EF, I, Li]. In particular, all classical results mentioned but not specifically referenced in this column can be found there.

References

[AV] S. Abiteboul and V. Vianu, "Generic Computation And Its Complexity," *IEEE Found. of Comp. Sci. Symp.* (1991), 209-219.

[AI] M. Adler and N. Immerman, "An n! Lower Bound On Formula Size," *ACM Transactions on Computational Logic*, 4(3) (2003), 296-314.

[A] G. Asser "Das Repräsentationsproblem im Prädikatenkalkül der ersten Stufe mit Identität," *Z. Math. Logik Grundlag Math* 1 (1955), 252-263.

[BK] L. Babai and L. Kučera, "Canonical Labelling of Graphs in Linear Average Time," *IEEE Found. of Comp. Sci. Symp.* (1980), 39-46.

[CF] J. Cai, M. Fürer, N. Immerman, "An Optimal Lower Bound on the Number of Variables for Graph Identification," *Combinatorica* 12(4) (1992) 389-410.

[D] A. Dawar, "Feasible Computation Through Model Theory," PhD Dissertation, University of Pennsylvania, 1993.

[DR] A. Dawar, D. Richerby, and B. Rossman, "Choiceless Polynomial Time, Counting, and Cai-Fürer-Immerman Graphs," *Twelfth Workshop on Logic, Language, Information and Computation* (2005), 13-24.

[dR] M. de Rougemont, "Uniform Definability on Finite Structures with Successor," *STOC* (1984), 409-417.

[EF] H.-D. Ebbinghaus, J. Flum, *Finite Model Theory* 1995, Springer 1995.

[EI] K. Etessami and N. Immerman, "Reachability and the Power of Local Ordering," *Theoret. Comp. Sci.* 148(2) (1995), 261-279.

[EI2] K. Etessami and N. Immerman, "Tree Canonization and Transitive Closure," *Information and Computation* 157(1,2) (2000), 2 - 24.

[F] R. Fagin, "Generalized First-Order Spectra and Polynomial-Time Recognizable Sets," in *Complexity of Computation,* (ed. R. Karp), *SIAM-AMS Proc. 7,* 1974, 27-41.

[F2] R. Fagin, "Monadic generalized spectra," *Zeitschr. f. math. Logik und Grundlagen d. Math.* 21 (1975), 89-96.

[GM] E. Grädel and G. McColm, "Hierarchies in Transitive Closure Logic, Stratified Datalog and Infinitary Logic," *Annals of Pure and Applied Logic 77* (1996), 166–199.

[G] M. Grohe, "Arity Hierarchies," *Annals of Pure and Applied Logic* 82 (1996), 103-163.

[G2] M. Grohe, "The Complexity off Homomorphism and Constraint Satisfaction Problems Seen From the Other Side" *IEEE Found. of Comp. Sci. Symp.* (2003), 552-561.

[GF] M. Grohe and M. Frick, "The Complexity of First-Order and Monadic Second-Order Logic Revisited," *Annals of Pure and Applied Logic* 130 (2004), 3-31.

[GS] M. Grohe and N. Schweikardt, "The Succinctness of First-Order logic on Linear Orders," *IEEE Symp. Logic In Comput. Sci.* (2004), 438-447.

[GS2] M. Grohe and T. Schwentick, "Locality of order-invariant first-order formulas," *ACM Transactions on Computational Logic* 1(1) (2000), 112 - 130.

[HI] W. Hesse and N. Immerman, "Complete problems for Dynamic Complexity Classes," *IEEE Symp. Logic In Comput. Sci.* (2002), 313-322.

[I] N. Immerman, *Descriptive Complexity*, 1999, Springer Graduate Texts in Computer Science, New York.

[I2] N. Immerman, *Expressibility and Parallel Complexity, SIAM J. of Comput* 18 (1989), 625-638.

[IB] N. Immerman, J. Buss, and D.M. Barrington, "Number of Variables Is Equivalent To Space," *J. Symbolic Logic* 66(3) (2001), 1217 - 1230.

[IL] N. Immerman and E. Lander, "Describing Graphs: A First-Order Approach to Graph Canonization," in *Complexity Theory Retrospective,* A. Selman, ed., Springer-Verlag, 1990, 59-81.

[JK] P. Jeavons and A. Krokhin, workshop organizers, "Mathematics of Constraint Satisfaction: Algebra, Logic and Graph Theory," March 20 - 24, 2006, Oxford University.

[K] Ph. Kolaitis, "Tutorial on Constraint Satisfaction, Complexity, and Logic" – Logic Colloquium 2005, July 2005, http://www.cs.ucsc.edu/ kolaitis/talks/csp-lc053.ps.

[KW] M. Karchmer and A. Wigderson, "Monotone circuits for connectivity require super-logarithmic depth," *SIAM J. Discrete Math. 3:2* (1990), 255-65.

[Li] L. Libkin, *Elements of Finite Model Theory,* 2004, Springer.

[LN] L. Libkin and J. Nurmonen, "Counting and locality over finite structures: a survey," in *Generalized Quantifiers and Computation*, Springer LNCS 1754, 2000, pages 18-50.

[Ly] J. Lynch, "Complexity Classes and Theories of Finite Models," *Math. Sys. Theory* 15 (1982), 127-144.

[R] B. Rossman, "Existential Positive Types and Preservation under Homomorphisms," *IEEE Symp. Logic In Comput. Sci.* (2005), 467-476.

[S] H. Scholz, "Problem #1: Ein ungelöstes Problem in der symbolischen Logik," *J. Symbolic Logic* 17 (1952), 160.

[S2] T. Schwentick, "Graph Connectivity and Monadic NP," *IEEE Found. of Comp. Sci. Symp.* (1994), 614-622.

[W] B. Weisfeiler, ed., *On Construction and Identification of Graphs,* Lecture Notes in Mathematics 558, Springer, 1976.