Counting proportions of sets:

expressive power with almost order

@ LATIN'06, Valdivia, CHILE, 24.03.06

Argimiro Arratia & Carlos Ortiz

Universidad de Valladolid, Spain
Arcadia University, USA

arratia@mac.uva.es ortiz@arcadia.edu

Motivations

As almost everyone we would like to show **NP** $\neq$ **P**.

We tackle problems in computational complexity using Descriptive Complexity

Fagin, 1974 : **NP** = $\exists$SO

Immerman, Vardi, 1982 : **P** = FO + LFP

Abiteboul-Vianu, 1991 : **PSPACE** = FO + PFP

Grädel, 1992 : **P** = $\exists$SO-Horn, **NL** = $\exists$SO-Krom

Motivations

- Logics capturing complexity classes below **NP** need built-in order
- In the **presence** of order techniques for showing inexpressibility, such as Ehrenfeucht–Fraissé games, are hard to apply and often do not give meaningful lower bounds for the interesting complexity classes
- In the **absence** of order, logics may become weak: FO + LFP with order captures **P**; without order it can not express the parity of the size of models.

Proposals to overcome difficulty of order are based mainly in:

- adding counting quantifiers to FO
- adding some weak form of order (an *almost order*)

Caveat: Libkin & Wong show that FO + COUNT with almost order is weak: it has the Bounded Number of Degrees Property (BNDP)

Almost order: Let $g : \mathbb{N} \to \mathbb{N}$, so that $g(n) < n$. An almost linear order over set A determined by g is binary relation $\leq_g$ that partitions A into B and C such that

- $|B| = n - g(n)$ and $\leq_g \restriction B$ is linear order
- $\leq_g \restriction C$ is 2-preorder (reflexive, transitive relation in which every induced equivalence class of $\leq_g \cap \leq_g^{-1}$ has size 2)
- $(\forall x \in C)(\forall y \in B)(x \leq_g y)$

Picture of an almost order

$$\underbrace{\begin{array}{c}\bullet\\ \uparrow\downarrow\\ \bullet\end{array} \to \begin{array}{c}\bullet\\ \uparrow\downarrow\\ \bullet\end{array} \to \cdots \begin{array}{c}\bullet\\ \uparrow\downarrow\\ \bullet\end{array}}_{g(n)} \underbrace{\to \bullet \to \bullet \cdots \to \bullet}_{n-g(n)}$$

BNDP: G a graph. Its degree set, $deg.set(G)$, is the set of all possible in- and out-degrees that are realised in G. A formula $\psi(x, y)$ on graphs has the Bounded Number of Degrees Property (BNDP) if there is a function $f : \mathbb{N} \to \mathbb{N}$ such that for any graph G with $deg.set(G) \subseteq \{0, \ldots, k\}$, $|deg.set(\psi[G])| \leq f(k)$, where $\psi[G]$ is the graph with same universe as G and edge relation given by ψ^G.

These notions generalise to arbitrary τ-structures and

[Libkin & Wong, 02]: every formula in $\mathcal{L}^*_{\infty\omega}(C)$ (a very strong counting logic that subsumes all known pure counting extensions of FO, including FO extended with all unary quantifiers), in the presence of almost-linear orders, has the BNDP and thus "*exhibits the very tame behaviour tipical for FO queries over unordered structures*"

Our proposal

- Work with almost order. It is a reasonable weakening of the order.
- Be relax! Count proportions of elements as opposed to exact number of elements that satisfy a formula. The intuition is: an operator that counts proportions should be less susceptible to perturbations by the change of semantics from linear orders to almost orders
- Extend your counting power: From counting any element in the universe, to count elements in specific sets of the universe

2nd. Order Proportional Quantifier Logic, $\mathcal{SOLP}$

Extend FO with quantifiers acting on fmlas. $\alpha(\overline{x}, X)$ with X 2nd. order var. of arity k, and rational $r \in (0,1)$:

$$(P(X) \geq r)\alpha(\overline{x}, X) \text{ and } (P(X) \leq r)\alpha(\overline{x}, X)$$

Semantic

Let $\mathcal{B}_m$ an appropriate structure of size m,

$$\mathcal{B}_m \models (P(X) \geq r)\alpha(\overline{a}, X) \iff$$

$$\text{exists } A \subseteq \{b_0, \ldots, b_{m-1}\}^k : \mathcal{B}_m \models \phi(\overline{a}, A) \text{ and } |A| \geq rm^k$$

Similarly for $(P(X) \leq r)$

Fragments of interest

$\mathcal{SOLP}[r_1, \ldots, r_k]$, for a given sequence $r_1, r_2, \ldots, r_k$ of distinct natural numbers, is the sublogic of $\mathcal{SOLP}$ where the proportional quantifiers can only be of the form $(P(X) \leq q/r_i)$ or $(P(X) \geq q/r_i)$, for $i = 1, \ldots, k$ and q a natural number such that $0 \leq q < r_i$.

Another fragment of $\mathcal{SOLP}$ of interest for us is the *Second Order Monadic Logic of Proportionality*, denoted $\mathcal{SOMLP}$, which is $\mathcal{SOLP}$ with the arity of the second order variables being all equal to 1.

Examples

1. Let $\tau = \{R, s, t\}$ (R 3–ary, s, t constant symbols). Let r be a rational with $0 < r < 1$.

NOT-IN-CLOS$_{\leq r}$ **:=** $\{\mathcal{A} = \langle A, R, s, t\rangle : A$ has a set containing s but not t, closed under R, and of size at most a fraction r of $|A|$ $\}$.

Let $\beta_{nclos}(X)$ be the following formula

$$\begin{aligned}\beta_{nclos}(X) \quad := \quad & \forall x \forall u \forall v \; [X(s) \wedge \neg X(t) \\ & \wedge \; (X(u) \wedge X(v) \wedge R(u, v, x) \;\longrightarrow\; X(x))]\end{aligned}$$

Then

$$\mathcal{A} \in \text{NOT-IN-CLOS}_{\leq r} \iff \mathcal{A} \models (P(X) \leq r)\beta_{nclos}(X)$$

This problem is **P**–complete under first order reductions.

2. Let $\tau_2 = \{E\}$, E binary relation. A τ–structure is seen as a graph. Let.

$$\psi := \forall x \bigvee_{i=1}^{k} \left(Z_i(x) \wedge \bigwedge_{\substack{j \neq i \\ 1 \leq j \leq k}} \neg Z_j(x) \right)$$

ψ says "*each vertex has exactly one of the possible k colors*". Let

$$\theta := \forall x \forall y \bigwedge_{i=1}^{k} (Z_i(x) \wedge Z_i(y) \longrightarrow \neg E(x,y))$$

θ says "*two vertices with the same color are not connected by an edge*". Then,

$$\Psi_k := \left(P(Z_1) \geq \frac{1}{k} \right) \left(P(Z_2) \leq \frac{k-1}{k} \right) \ldots \left(P(Z_k) \leq \frac{k-1}{k} \right) (\theta \wedge \psi)$$

is a sentence in $\mathcal{SOMLP}[k]$, and expresses that "*the graph is k-colorable*"

3. $\mathbf{NCON}_{\geq r} := \{\mathcal{A} = \langle A, E, s \rangle : \langle A, E \rangle$ is digraph and **at least** a fraction r of A is **not** connected to $s\}$

Let

$$\alpha_{ncon}(B) \quad := \quad \neg B(s) \wedge \forall x \forall y (E(x,y) \wedge B(x) \longrightarrow B(y))$$

Then

$$\mathcal{A}_n \in \mathbf{NCON}_{\geq r} \iff \mathcal{A}_n \models (P(B) \geq r)\alpha_{ncon}(B)$$

Prop.: $\mathbf{NCON}_{\geq r}$ is complete for **NL** via *fop*.

The logic with order: $\mathcal{SOLP}+\leq$

We have shown

(1) In the presence of order (at least a built–in successor), $\mathbf{P} \subseteq \mathcal{SOLP}[2]$ (in the sense that any class of structures decidable in $\mathbf{P}$ is definable by a sentence of $\mathcal{SOLP}[2]$) and, furthermore, it is captured by the fragment $\mathcal{SOLP}Horn[2]$, consisting of formulas of the form $(P(X_1) \leq 1/2) \cdots (P(X_r) \leq 1/2)\alpha$, where α is a universal Horn formula.
(2) In the presence of order, **NL** is captured by $\mathcal{SOLP}Krom[2]$, a fragment consisting of formulas of the form $(P(X_1) \geq 1/2) \cdots (P(X_r) \geq 1/2)\alpha$, where α is a universal Krom formula.

(This and the previous capturing of $\mathbf{P}$ by fragments of $\mathcal{SOLP}$ are inspired on Grädel's [TCS 1992], but taking into account the limitations in the cardinalities of second order variables imposed by our counting quantifiers.)

$\mathcal{SOLP}$ restricted to almost order

We must make sure that elements that are equivalent w.r.to almost order satisfy same predicates, and avoid inconsistencies in our model theory.

An almost linear order $\leq_g$ over a set A induces an equivalence relation $\equiv_g$ in A defined by $a \equiv_g b$ iff $a \leq_g b$ and $b \leq_g a$.

An n-ary relation R on a set A is **consistent** with $\leq_g$ if for every pair of vectors $(a_1, \ldots, a_n)$ and $(b_1, \ldots, b_n)$ of elements in A with $a_i \equiv_g b_i$ for every $i \leq n$, we have that

$$R(a_1, \ldots, a_n) \text{ holds if and only if } R(b_1, \ldots, b_n) \text{ holds.}$$

A structure $\mathcal{A} = \langle A, R_1^{\mathcal{A}}, \ldots, R_k^{\mathcal{A}}, C_1^{\mathcal{A}}, \ldots, C_s^{\mathcal{A}} \rangle$ is consistent with $\leq_g$ iff for every $i \leq k$, $R_i^{\mathcal{A}}$ is consistent with $\leq_g$.

$\mathcal{SOLP} + \leq_g$, for almost order $\leq_g$, is $\mathcal{SOLP}$ with $\leq_g$ as additional built-in relation, and where

* only consider models $\mathcal{A}$ that are **consistent** with $\leq_g$.
* for formulas $(P(X) \geq r)\phi(\overline{x}, \overline{Y}, X)$ and $(P(X) \leq r)\phi(\overline{x}, \overline{Y}, X)$, for appropriate finite model $\mathcal{A}$ consistent with $\leq_g$, for elements $\overline{a} = (a_1, \ldots, a_m)$ in A and appropriate vector of relations $\overline{B}$, consistent with $\leq_g$, we should have

  $\mathcal{A} \models (P(X) \geq r)\phi(\overline{a}, \overline{B}, X) \iff$ there exists $S \subseteq A^k$, **consistent with** $\leq_g$, such that $\mathcal{A} \models \phi(\overline{a}, \overline{B}, A)$ and $|S| \geq r \cdot n^k$

Similarly for $(P(X) \leq r)\phi(\overline{a}, \overline{Y}, X)$, substituting in the above condition $\geq$ for $\leq$.

$\mathcal{SOMLP}+ \leq_g$ passed a typical benchmark

For every almost order $\leq_g$, given by a sublinear function g, we can define in $\mathcal{SOMLP}[2]+ \leq_g$, the set of models with almost order and with universe of even cardinality.

Example: Fix an almost order $\leq_g$, and consider the sentence Θ_2:

$$\left(P(B) \geq \frac{1}{2}\right)\left(P(C) \geq \frac{1}{2}\right)[\forall x(B(x) \vee C(x)) \wedge \forall y(B(y) \longrightarrow \neg C(y))]$$

Then for every structure $\mathcal{A}$, consistent with $\leq_g$,

$$\mathcal{A} \models \Theta_2 \text{ iff } |\mathcal{A}| \text{ is even}$$

$\mathcal{SOMLP}+\leq_g$ does not have the BNDP

Consider the quantifier free formula $\theta(x, y, U)$ in $\mathcal{SOMLP}(\{E\})$ (E binary):

- $x \neq y$;
- $x \in U$ and $y \in U$;
- There is no element w of U such that $E(w, x)$ and there is no element w of U such that $E(y, w)$;
- $\exists w_1, w_2 \in U$ such that $E(x, w_1)$ and $E(w_2, y)$;
- For any element z in U different from x and y there exists unique $a, b \in U$ such that $E(a, z)$ and $E(z, b)$.

and let

$$\Psi(x, y) := \left(P(U) \geq \frac{1}{2}\right) \theta(x, y, U)$$

$\Psi(x, y)$ does not have the BNDP property for most sublinear functions g; for if we look at the models $\mathcal{A}$ consistent with $\leq_g$ and of cardinality $2n$, whose graph $E(x, y)$ is just the natural successor relation induced by $\leq_g$, i.e.

$$\underbrace{\begin{array}{ccccccc} \bullet & & \bullet & & & & \bullet \\ \uparrow\downarrow & \rightarrow & \uparrow\downarrow & \rightarrow \cdots & & & \uparrow\downarrow \\ \bullet & & \bullet & & & & \bullet \end{array}}_{g(2n)} \underbrace{\rightarrow \bullet \rightarrow \bullet \ldots \rightarrow \bullet}_{2n-g(2n)}$$

we see that E is consistent with $\leq_g$ and that $deg.set(\mathcal{A}) \subseteq \{0, 1, 2\}$. However, the structure $\psi[\mathcal{A}]$ represents, for any n, the "transitive closure of length bigger or equal to half the size of the model $\mathcal{A}$", and thus $1, 2, \ldots n \in deg.set(\psi[\mathcal{A}])$ for every g sublinear. $\square$

Separation in the presence of almost order

Using appropriate Ehrenfeucht–Fraïssé type of games, we've shown

(1) With respect to almost ordered structures there exists an infinite hierarchy within the monadic fragment $\mathcal{SOMLP}$, namely,

$$\mathcal{SOMLP}[2] \subsetneq \mathcal{SOMLP}[2,3] \subsetneq \mathcal{SOMLP}[2,3,5] \subsetneq \ldots$$

(2) With respect to almost ordered structures and unbounded arity we have that

$$\mathcal{SOLP}Horn[2] \subsetneq \mathcal{SOLP}[2,3].$$

(Recall that in the presence of order, i.e. $\mathcal{SOLP}+\leq$,

$$\mathbf{P} \subseteq \mathcal{SOLP}[2] \subseteq \mathcal{SOLP}[2,3] \subseteq \mathbf{PSPACE})$$

Sketch of proof of existence of hierarchy in $\mathcal{SOMLP}$

Using our notion of games we proved

Lemma : Let $r_1, r_2, \ldots, r_k$ be distinct non zero natural numbers. Let g be a sublinear function, $\leq_g$. For every pair of structures $\mathcal{A}$ and $\mathcal{B}$, such that $\mathcal{A}/_{\sim_g} \cong \mathcal{B}/_{\sim_g}$, $|\mathcal{A}| = m$, $|\mathcal{B}| = m+1$, $m+1 > r_i$ and $m \equiv_{r_i} -1$, for every $i \leq k$, we have that,

$$\mathcal{A} \models \varphi \text{ implies } \mathcal{B} \models \varphi$$

for all sentence φ of $\mathcal{SOMLP}(\tau)[r_1, r_2, \ldots, r_k]$ $\square$

Theorem : Let $r, r_1, r_2, \ldots, r_k$ be distinct non zero natural numbers, pairwise relatively prime. Then $\mathcal{SOMLP}_A[r_1, \ldots, r_k]$ is properly contained in $\mathcal{SOMLP}_A[r_1 \ldots r_k, r]$.

Proof : We show that the query "the size of the model is a multiple of r" is not expressible in $\mathcal{SOMLP}_A[r_1 \ldots, r_k]$.

Assume there exists a sentence ϕ in $\mathcal{SOMLP}[r_1 \ldots, r_k]$ that defines the query, for all almost ordered structure $\mathcal{A}$. Using that r is relatively prime with the r_i's together with the Generalised Chinese Remainder Theorem we can get a $b \leq r(\prod_{i=1}^{k} r_i)$ such that

$$b \equiv_r 0 \text{ and } b \equiv_{r_i} -1, \text{ for all } i = 1, \ldots, k$$

Take $m = r(\prod_{i=1}^{k} r_i)n + b$, for some $n > 1$. Observe that

$$m \equiv_r 0,\ m \equiv_{r_i} -1 \text{ and } m + 1 > r_i, \text{ for all } i = 1, \ldots, k$$

Let $g = h_t(\cdot)$ with $t = r(\prod_{i=1}^{k} r_i)n$ (here $h_t(n) = 2r$, where $r \equiv_t n$). Then

$$h_t(m) = 2b \text{ and } h_t(m+1) = 2b + 2$$

Let $\mathcal{A}$ be a structure, consistent with the almost order $\leq_g$, formed by t many 2-preorders followed by a linear order of size b. Let $\mathcal{B}$ be $\mathcal{A}$ with a new element with which we form an extra 2-preorder; that is, $\mathcal{B}$ consists of $t+1$ 2-preorders and a linear order of size $b-1$. There is a natural isomorphism between $\mathcal{A}/_{\equiv_g}$ and $\mathcal{B}/_{\equiv_g}$.

On the other hand, m satisfies the conditions of previous Lemma, and $|\mathcal{A}| = m$ and $|\mathcal{B}| = m+1$. It follows that if $\mathcal{A} \models \phi$ then $\mathcal{B} \models \phi$; therefore $m+1$ is a multiple of r, which is impossible. $\square$

Thus,

$$\mathcal{SOMLP}_A[2] \subsetneq \mathcal{SOMLP}_A[2,3] \subsetneq \mathcal{SOMLP}_A[2,3,5] \subsetneq \ldots$$