Counting Proportions of Sets: Expressive Power with Almost Order

Argimiro Arratia[1,⋆] and Carlos E. Ortiz[2,⋆⋆]

[1] Dpto. de Matemática Aplicada, Facultad de Ciencias,
Universidad de Valladolid, Valladolid 47005, Spain
arratia@mac.cie.uva.es

[2] Department of Mathematics and Computer Science,
Arcadia University, 450 S. Easton Road,
Glenside, PA 19038-3295, U.S.A
ortiz@arcadia.edu

Abstract. We present a second order logic of proportional quantifiers, $\mathcal{SOLP}$, which is essentially a first order language extended with quantifiers that act upon second order variables of a given arity r, and count the fraction of elements in a subset of r–tuples of a model that satisfy a formula. Our logic is capable of expressing proportional versions of different problems of complexity up to **NP**-hard, and fragments within our logic capture complexity classes as **NL** and **P**, with auxiliary ordering relation. When restricted to monadic second order variables our logic of proportional quantifiers admits a semantic approximation based on almost linear orders, which is not as weak as other known logics with counting quantifiers, for it does not has the *bounded number of degrees property.* Moreover, we show in this almost ordered setting the existence of an infinite hierarchy inside our monadic language. We extend our inexpressibility result to an almost ordered (not necessarily monadic) fragment of $\mathcal{SOLP}$, which in the presence of full order captures **P**. To obtain all our inexpressibility results we developed combinatorial games appropriate for these logics.

Keywords: Proportional quantifiers, almost order, expressiveness, computational complexity, **P**, **NL**.

1 Introduction

An important open problem in Descriptive Complexity is to establish the existence of a logic, with recursive syntax and semantic, for describing all polynomial time computable problems, that is, for capturing the class **P**. The bottom line is that a solution to this problem should lead to a better understanding of the role of ordering in computations.

⋆ Supported by MEC, Projects MOISES (TIC2002-04019-C03-03), and SINGACOM (MTM-00958).

⋆⋆ Supported by a Faculty Award Grant from the Christian R. & Mary F. Lindback Foundation, and a Visiting Research Fellowship from Universidad de Valladolid, Spain.

J.R. Correa, A. Hevia, and M. Kiwi (Eds.): LATIN 2006, LNCS 3887, pp. 105–117, 2006.

As of today, all known logics that capture **P** need a built–in linear order as an extra symbol, so that the capturing may take place. The main issue is that a pre–defined ordering relation added to a logic and with its interpretation invariant through the models, makes the syntax of such logic non recursive (a consequence of Trahtenbrot's Theorem [2]); and thus this logic hardly classifies as "good" programming paradigm. On the other hand the presence of a built–in linear order, as part of the structures representing instances of computational problems, makes it very difficult for inexpressibility techniques from Model Theory, such as Ehrenfeucht-Fraïssé games, to succeed in showing meaningful computational lower bounds (e.g. see [5–§ 6.6]). To overcome this difficulty, and mindful of finding a logic in the aforesaid terms for **P**, various order–free extensions of first order logic (FO) have been proposed, most notably by the addition of some form of counting. However the demonstrated insufficient power of expressiveness of counting operators alone has led to the exploration (and exploitation) of some forms of pre–defined weak order and of the local nature of first order logic. The hope is that the logics with built-in weak form of order may have non-trivial expressive power, may be easier to separate, and eventually may shed light into the problem of separation of the corresponding logics with built-in order. In this context, the paper by Libkin and Wong [6] suggests that the above mentioned program may not be feasible because it shows an inherent expressive limitations of counting logics in the presence of auxiliary relations, which they call *preorders*, and their associated *almost–linear orders*. The main result of [6] is that a very powerful extension of FO with counting, denoted $\mathcal{L}^*_{\infty\omega}(C)$, which subsumes all known "pure" counting extensions of FO (meaning that fixpoint operators are not considered), in the presence of almost–linear orders, has the *bounded number of degrees property* (BNDP). The BNDP is a semantic property that limits the expressive power of logics that have it; such logics cannot express, for example, the transitive closure of a binary relation. (We will review all concepts in italics later in this paper.)

The purpose of this paper is to introduce a second order counting logic with built-in order that contains fragments whose expressive power is meaningful for Complexity Theory, and where the replacement of the built-in order by almost order does not yield logics with trivial expressive power, and where it should not be hard to obtain separation results. Our proposal consists of enhancing FO with quantifiers of the form $(P(X) \geq r)$ and $(P(X) \leq r)$ for rational $r \in (0, 1)$ and second order variable X of, say, arity $k > 0$, and whose meaning is that the cardinality of the set X is greater than or equal to (or less than or equal to) r times the cardinality of the set of k–tuples in the model. The logic obtained by adding these quantifiers, denoted by $\mathcal{SOLP}$ for *Second Order Logic of Proportions* (or *proportional quantifiers*), extends its first order counterpart $\mathcal{LP}$, which was introduced and studied by us in [1]. The intuition driving the definition of this logic is that by counting proportions as opposed to counting exact numbers of elements, the proportional quantifiers should be less susceptible to perturbations by the change of semantics from linear orders to almost-orders than the standard counting quantifiers.

Due to the proceedings' strict page limitations we must omit most of the proofs. The reader interested in learning all the details may request the extended version from the first author.

2 Second Order Logic of Proportional Quantifiers

Throughout this paper we use standard notation and concepts of Finite Model Theory as presented in the books [2] and [5]. Our vocabularies are finite and consists of relation symbols and constant symbols. Our structures are all finite, and if $\mathcal{A}$ is a structure over vocabulary τ, or τ–structure, and A is its universe, we either use $|\mathcal{A}|$ or $|A|$ to denote its size, that is, the number of elements in A.

In [1] we studied extensions of first order logic with quantifiers that count fractions of elements in a model that satisfy a given formula, and defined approximations to their semantics by giving interpretations of the formulae on finite structures where all predicates are restricted to act subject to an integer modulo. A natural extension is to have the proportional quantifiers act upon second order variables. This as we shall see gives more expressive power.

Definition 1. *The Second Order Logic of Proportional quantifiers, $\mathcal{SOLP}$, is the set of formulas of the form*

$$Q_1 \cdots Q_u \theta(x_1, \ldots, x_s, X_1, \ldots, X_r) \tag{1}$$

where $\theta(x_1, \ldots, x_s, X_1, \ldots, X_r)$ is a first order formula over some vocabulary τ with first order variables $x_1, \ldots, x_s$ and second order variables, $X_1, \ldots, X_r$; each Q_j ($j \leq u$) is either $(P(X_i) \geq t_i)$ or $(P(X_i) \leq t_i)$, where t_i is a rational in $(0, 1)$, for $i \leq r$. Whenever we want to make the underlying vocabulary τ explicit we will write $\mathcal{SOLP}(\tau)$.

We also define $\mathcal{SOLP}(\tau)[r_1, \ldots, r_k]$, for a given vocabulary τ and sequence $r_1, r_2, \ldots, r_k$ of distinct natural numbers, as the sublogic of $\mathcal{SOLP}(\tau)$ 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 $0 < q < r_i$. Another fragment of $\mathcal{SOLP}$ which will be of interest for us is the Second Order Monadic Logic of Proportional quantifiers*, denoted $\mathcal{SOMLP}$, which is $\mathcal{SOLP}$ with the arity of the second order variables in* (1) *being all equal to 1.*

The interpretation for the proportional quantifiers is the natural one: Let X be a second order variable of arity k, $\overline{Y}$ a vector of second order variables, $\overline{x} = x_1, \ldots, x_m$ first order variables and $\phi(\overline{x}, \overline{Y}, X)$ a formula in $\mathcal{SOLP}(\tau)$ over some (finite) vocabulary τ, which does not contains X or any of the variables in $\overline{Y}$ as a relation symbol. Let r be a rational in $(0, 1)$. Then

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

have the following semantics. For appropriate finite τ–structure $\mathcal{A}$, elements $\overline{a} = (a_1, \ldots, a_m)$ in A and vector of relations $\overline{B}$ over A, we have

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

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

Example 1. Let $\tau = \{R, s, t\}$ where R is a ternary relation symbol, and s and t are constant symbols. Let r be a rational with $0 < r < 1$. We define

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 $|A|$ $\}$.

$$\text{Let } \beta_{nclos}(X) := \forall x \forall u \forall v\ [X(s) \wedge \neg X(t) \wedge (X(u) \wedge X(v) \wedge R(u, v, x) \longrightarrow X(x))]$$

Then

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

We shall see in Section 3 that for $r = 1/2$ this problem is **P**–complete under first order reductions. (This result can be generalised to $r = 1/n$.) □

For **NP** we have the following problem.

Example 2. Let $\tau = \{E\}$, let r be a rational with $0 < r < 1$. We define

CLIQUE$_{\geq r}$ **:=** $\{\mathcal{A} = \langle A, E\rangle : \langle A, E\rangle$ is a graph and at least a fraction r of the vertices form a complete graph $\}$

This problem can be defined by the sentence $(P(X) \geq r)\alpha_{cliq}(X)$, where

$$\alpha_{cliq}(X) := \forall x \forall y (X(x) \wedge X(y) \wedge x \neq y \longrightarrow E(x, y))$$

One can show that, for any rational $r \in (0, 1)$, CLIQUE$_{\geq r}$ is **NP**-complete via *logspace reducibilities.*

The following remark shows that $\mathcal{SOLP}$ extends the (classical) logic ∃SO.

Remark 1. Any formula in ∃SO is equivalent to a formula in $\mathcal{SOLP}[k]$, for any $k > 1$. Indeed, consider a formula of the form $\exists X \phi(X)$, where $\phi(X)$ is a first order formula with free second order variable X of arity $r > 0$. This can be expressed in $\mathcal{SOLP}[k]$ by the formula:

$$\left(P(X_1) \leq \frac{k-1}{k}\right)\left(P(X_2) \geq \frac{k-1}{k}\right)\phi(X_1) \vee \phi(X_2)$$

where X_1 and X_2 are variables of arity r.

3 Expressiveness of $\mathcal{SOLP}$ in the Presence of Order

By Remark 1, $\mathcal{SOLP}$ subsumes ∃SO. However, it adds extra information to the description of complexity classes, provided by the computing of bounds in the cardinality of sets in instances of problems. This we shall see in this section, where we impose constraints to the syntax of $\mathcal{SOLP}$ similar to Grädel's constraints for ∃SO in [4], and capture the classes **P** and **NL**, but as an extra information we have that **P** (and **NL**) $\subseteq \mathcal{SOLP}[2]$ and the first order part of the sentences describing this class is *universal Horn* (for **NL** it will be *universal Krom*). Furthermore, observe that all our examples of computational problems are definable in $\mathcal{SOMLP}$, the monadic fragment of $\mathcal{SOLP}$, some of them with not known expression (or non expressible) in monadic ∃SO.

Definition 2. *Let $\tau = \{R_1, \ldots, R_m, C_1, \ldots, C_s\}$ be some vocabulary with relation symbols $R_1, \ldots, R_m$, and constant symbols $C_1, \ldots, C_s$, and let $X_1, \ldots, X_r$ be second order variables of arity $k_1, \ldots, k_r$, respectively. A first order formula α over $\tau \cup \{X_1, \ldots, X_r\}$, and extra binary relation symbol $=$ (equality) and the constant $\bot$ (standing for false), is a* universal Horn formula*, if α is a universally quantified conjunction of formulas over $\tau \cup \{X_1, \ldots, X_r\}$ of the form $\psi_1 \wedge \psi_2 \wedge \ldots \wedge \psi_s \longrightarrow \varphi$, where φ is either $X_i(\overline{u}_i)$ (where $\overline{u}_i$ denotes a k_i-tuple of first order terms, $i = 1, \ldots, r$) or $\bot$, and $\psi_1, \ldots, \psi_s$ are atomic or negation of atomic $(\tau \cup \{X_1, \ldots, X_r\})$-formulas except that any occurrence of the variables X_i must be positive (there are no restrictions on the predicates in τ or $=$). The logic $\mathcal{SOLP}Horn$ is the set of formulas of the form*

$$(P(X_1) \leq t_1) \cdots (P(X_r) \leq t_r)\alpha$$

where each t_i is a rational in $(0, 1)$, and α is a universal Horn formula over some vocabulary τ and second order variables $X_1, \ldots, X_r$.

By Example 1, the problem NOT-IN-CLOS$_{\leq r}$ is definable in $\mathcal{SOLP}Horn$. We can show that to test membership for a problem definable in $\mathcal{SOLP}Horn$ can be done deterministically in polynomial time.

Lemma 1. *The set of finite structures that satisfy a sentence θ in $\mathcal{SOLP}Horn$ is in* **P**. □

Thus, according to this lemma, our problem NOT-IN-CLOS$_{\leq r}$ is in **P**. We can prove that, for $r = 1/2$, it is complete for **P** via first order reductions. The idea is to define a reduction from the problem *Path System Accessibility* to NOT-IN-CLOS$_{\leq 1/2}$ using quantifier free first order formulae. An instance of the Path System Accessibility problem, which we abbreviate from now on as PS, is a finite structure $\mathcal{A} = \langle A, R, s, t \rangle$ or a *path system*, where the universe A consists of, say, n vertices, a relation $R \subseteq A \times A \times A$ (the *rules* of the system), a *source* $s \in A$, and a *target* $t \in A$ such that $s \neq t$. A positive instance of PS is a path system $\mathcal{A}$ where the target is *accessible* from the source, where a vertex v is accessible if it is the source s or if $R(x, y, v)$ holds for some accessible vertices x and y, possibly equal. In [7] Stewart shows that PS is complete for **P** via quantifier free first order reductions that include built-in order; in fact, via *projections* (see [7] for definitions and also [5–§ 11.2]). We get the following result.

Lemma 2. *The problem* NOT-IN-CLOS$_{\leq 1/2}$ *is complete for* **P** *via quantifier free projections (qfp's), that include the use of built-in successor.* □

Corollary 1. *Over finite structures, ordered with a built-in successor, the logic* $\mathcal{SOLP}Horn$ *captures* **P**. □

For logarithmic space bounded classes we have the following examples.

Example 3. Let $\tau = \{E, s\}$ where E is a binary relation symbol and s is a constant symbol. We think of τ-structures as graphs with a specify vertex s (the source). Let r be a rational with $0 < r < 1$. We define

$\text{NCON}_{\geq r}$ **:=** $\{\mathcal{A} = \langle A, E, s\rangle : \langle A, E\rangle$ is a graph and at least a fraction r of the vertices are **not** connected to $s\}$

Let $\alpha_{ncon}(Y)$ be the following formula

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

Then $\mathcal{A} \in \text{NCON}_{\geq r} \iff \mathcal{A} \models (P(Y) \geq r)\alpha_{ncon}(Y)$.

Again, inspired on work by Grädel [4] we define:

Definition 3. *Let* τ *and* $X_1, \ldots, X_r$ *be as in Definition 2. A first order formula* α *over* $\tau \cup \{X_1, \ldots, X_r\} \cup \{=, \perp\}$ *is a* universal Krom formula*, if* α *is a universally quantified conjunction of clauses, where each clause is a disjunction of literals with at most two occurrences (positive or not) of the predicates* $X_1, \ldots, X_r$*, i.e.* α *is a 2-CNF formula with respect to the variables* $X_1, \ldots, X_r$*. The logic* $\mathcal{SOLP}Krom$ *is the set of formulas of the form*

$$(P(X_1) \geq t_1) \cdots (P(X_r) \geq t_r)\alpha$$

where each t_i *is a rational in* $(0, 1)$*, and* α *is a universal Krom formula over some vocabulary* τ *and second order variables* $X_1, \ldots, X_r$*.*

The sentence defining $\text{NCON}_{\geq r}$ is in $\mathcal{SOLP}Krom$. We can show that $\text{NCON}_{\geq r}$ is in **NL**, the class of problems decidable by non deterministic logarithmic space bounded Turing machines; and, furthermore, that for $r = 1/2$ the problem $\text{NCON}_{\geq r}$ is hard for **NL** via qfp's. Then with an argument similar to the one given for $\mathcal{SOLP}Horn$ one can show that satisfiability of sentences from $\mathcal{SOLP}Krom$ can be decided in **NL**, and conclude that over finite structures, ordered with built–in successor, $\mathcal{SOLP}Krom$ captures **NL**.

Remark 2. We can say more about the capturing of the class **P** by the logic $\mathcal{SOLP}$. The problem NOT-IN-CLOS$_{\leq 1/2}$ is complete via qfp's with order, and expressible in $\mathcal{SOLP}Horn[2]$; hence by reducing every problem K in **P** to NOT-IN-CLOS$_{\leq 1/2}$ with a quantifier free first order expressible reduction (which may include a successor relation), we get a sentence in $\mathcal{SOLP}Horn[2]$ defining K. Thus, $\mathbf{P} = \mathcal{SOLP}Horn[2]$ and obviously

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

The chain (2) motivate us to study the possibility of establishing a hierarchy in $\mathcal{SOLP}[2] \subseteq \mathcal{SOLP}[2, 3] \subseteq \mathcal{SOLP}[2, 3, 5] \subseteq \ldots$, etc. We present in this paper the separation of fragments of these logics when a weak form of order is present, namely an almost linear order.

4 $\mathcal{SOLP}$ Restricted to Almost Orders

We begin with two preliminary definitions. The first is a slight modification of the notion of almost linear order from [6]; for it we remind the reader that a function $g : \mathbb{N} \to \mathbb{N}$ is sublinear if, for all $n \in \mathbb{N}$, $g(n) < n$.

Definition 4. *For a fixed positive integer k, a k-preorder over a set A is a binary, reflexive and transitive relation P in which every induced equivalence class of $P \cap P^{-1}$ has size at most k. An almost linear order over A, determined by a sublinear function $g : \mathbb{N} \to \mathbb{N}$, is a binary relation $\leq_g$ over A with a partition of the universe A into two sets B, C, such that B has cardinality $n - g(n)$ and $\leq_g$ restricted to B is a linear order, $\leq_g$ restricted to C is a 2-preorder, and for every $x \in C$ and every $y \in B$, $x \leq_g y$.*

Note that for any function $g : \mathbb{N} \to \mathbb{N}$, the almost linear order $\leq_g$ over a set A induces an equivalence relation $\sim_g$ in A defined by $a \sim_g b$ iff $a \leq_g b$ and $b \leq_g a$.

Definition 5. *Fix a sublinear $g : \mathbb{N} \to \mathbb{N}$ and let R be an n-ary relation on a set A. Let $\leq_g$ be an almost-order determined by g in A. We say that R 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 $a_i \sim_g b_i$ for every $i \leq n$, we have that*

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

Let $\mathcal{A} = \langle A, R_1^{\mathcal{A}}, \ldots, R_k^{\mathcal{A}}, C_1^{\mathcal{A}}, \ldots, C_s^{\mathcal{A}} \rangle$ be a τ-structure. We say that $\mathcal{A}$ is consistent with $\leq_g$ if and only if for every $i \leq k$, $R_i^{\mathcal{A}}$ is consistent with $\leq_g$.

By $\mathcal{SOLP}(\tau)_{\leq_g}$, for an almost order $\leq_g$, we understand the logic $\mathcal{SOLP}(\tau)$ with the almost order $\leq_g$ as additional built-in relation, and where we only consider models $\mathcal{A}$ that are consistent with $\leq_g$. Furthermore, for the formulas of the form $(P(X) \geq r)\phi(\overline{x}, \overline{Y}, X)$ and $(P(X) \leq r)\phi(\overline{x}, \overline{Y}, X)$, we require the following modification of the semantics: For an appropriate finite τ–model $\mathcal{A}$ consistent with $\leq_g$, for elements $\overline{a} = (a_1, \ldots, a_m)$ in A and an appropriate vector of relations $\overline{B}$, consistent with $\leq_g$, we should have

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

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

The property of being consistent for $\leq_g$ holds in fact for all the formulas in $\mathcal{SOLP}(\tau)_{\leq_g}$. The proof is an easy induction in formulas.

Lemma 3. *Let $\mathcal{A}$ be a τ-structure which is consistent with $\leq_g$. Then, for every formula $\psi(\overline{x})$ in $\mathcal{SOLP}(\tau)_{\leq_g}$, the set $\psi^{\mathcal{A}} := \{\overline{a} \in \mathcal{A} : \mathcal{A} \models \psi(\overline{a})\}$ is consistent with $\leq_g$.* □

Definition 6. *We will use the expression "**almost** second order proportional quantifier logic", and denote this by A–$\mathcal{SOLP}$, to refer to the collection of languages $\mathcal{SOLP}_{\leq_g}$ for every almost order $\leq_g$ given by a sublinear function g.*

Likewise, we denote A–$\mathcal{SOLP}[r_1,\ldots,r_k]$ *the collection of all the languages* $\mathcal{SOLP}_{\leq_g}[r_1,\ldots,r_k]$*, for naturals* $r_1,\ldots,r_k$*, and* A–$\mathcal{SOMLP}$ *,* A–$\mathcal{SOMLP}[r_1,\ldots,r_k]$ *for the corresponding monadic fragments.*

For an illustration of the expressive power of the almost second order proportional quantifier logic, we shall give below a definition in A–$\mathcal{SOMLP}[2]$ of the set of models with almost order and with universe of even cardinality.

Example 4. Fix an almost order $\leq_g$, and consider the sentence

$$\Theta_2 := \left(P(B) \geq \frac{1}{2}\right)\left(P(C) \geq \frac{1}{2}\right)[\forall x(B(x) \vee C(x)) \wedge \forall y(B(y) \longrightarrow \neg C(y))]$$

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

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

The direction from left to right is clear: Θ expresses that B and C constitute a partition of $\mathcal{A}$. For the opposite direction, suppose m is even. There are $r \leq g(m)/2$ classes with two elements, say $\{a_1,b_1\},\ldots,\{a_r,b_r\}$, and $l = m-2r$ with one element, say there are $\{c_1\},\ldots,\{c_l\}$. Hence, $m = 2r+l$ and since m is even, l must be even. We proceed to construct our disjoint sets C and B. Observe that for each $i = 1,\ldots,r$, both elements a_i and b_i must go into either $B^{\mathcal{A}}$ or $C^{\mathcal{A}}$, because $\mathcal{A}$ is consistent with $\leq_g$. With this in mind we do the following: If r is even then we can construct our even partition of same cardinality without much effort. If r is odd, then $r-1 = 2k$ for some k, and so we put k classes (of two elements each) into $B^{\mathcal{A}}$, and the remaining $k+1$ many 2-elements classes into $C^{\mathcal{A}}$. To compensate we put classes $\{c_1\}$ and $\{c_2\}$ in $B^{\mathcal{A}}$, and the remaining $l-2$ 1-element classes are split evenly into $B^{\mathcal{A}}$ and $C^{\mathcal{A}}$. These sets $B^{\mathcal{A}}$ and $C^{\mathcal{A}}$ verify the formula $\alpha(B,C) := \forall x(B(x) \vee C(x)) \wedge \forall y(B(y) \longrightarrow \neg C(y))$ in $\mathcal{A}$ and have same cardinality. □

In a similar way, one can prove that for every natural $d > 2$, there exists a formula Θ_d, in the almost monadic second order proportional quantifier logic, with quantifiers of the form $P(X) \geq 1/d$ and $P(X) \geq (d-1)/d$ (i.e., contained in A–$\mathcal{SOMLP}[d]$), such that for structure $\mathcal{A}$, consistent with almost order $\leq_g$, $\mathcal{A} \models \Theta_d$ iff $|\mathcal{A}|$ is a multiple of d.

It was shown in [6] that a very powerful counting logic, $\mathcal{L}^*_{\infty\omega}(C)$, when restricted to almost orders, has the BNDP; hence, it has a very limited expressive power. The next example shows that this is not the case for A–$\mathcal{SOMLP}$.

Example 5. A–$\mathcal{SOMLP}$ **does not have the BNDP:** For a graph G, 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 it is shown in [6] that every formula in $\mathcal{L}^*_{\infty\omega}(C)$, in the presence of almost-linear orders, has the BNDP and thus "*exhibits the very tame behaviour tipical for FO queries over unordered structures*" [6]. We shall see later that A–$\mathcal{SOMLP}$ presents a tame behaviour too since we can easily show separation results; however it differs from the counting logics considered by Libkin and Wong in [6] in that it does not have the BNDP.

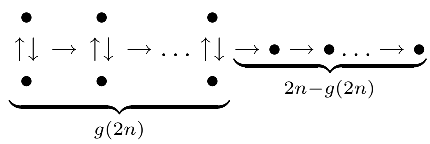
Consider the quantifier free formula $path(x, y, U)$ in A–$\mathcal{SOMLP}(\{E\})$ that states that:

- $x \neq y$, $x \in U$ 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

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

This formula 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.



we see that E is consistent with $\leq_g$ and that $deg.set(\mathcal{A}) \subseteq \{1, 2, 3, 4\}$. 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 $\lfloor n/2 \rfloor$, $\lfloor n/2 \rfloor + 1$, $\ldots \in deg.set(\psi[\mathcal{A}])$ for every g sublinear. □

5 Playing Games in $\mathcal{SOMLP}$

Definition 7. *Let τ be a vocabulary and $\mathcal{A}$ and $\mathcal{B}$ be two τ–structures, with $|B| = |A| + 1$. Let k and t be two positive integers. By $\mathcal{A} \prec_{(k,t)} \mathcal{B}$ we abbreviate the following statement:*

> *For every formula $\varphi(X_1, \ldots, X_t)$ of $FO(\tau \cup \{X_1, \ldots, X_t\})$ of (first order) quantifier rank $\leq k$ and unary second order variables $X_1, \ldots, X_t$, for all subsets $C_1, \ldots, C_t$ of A, there exist subsets $D_1, \ldots, D_t$ of B, such that*
> - *$|C_i| \leq |D_i| \leq |C_i| + 1$, for $i = 1, \ldots, t$, and*
> - *$\mathcal{A} \models \varphi(C_1, \ldots, C_t)$ implies $\mathcal{B} \models \varphi(D_1, \ldots, D_t)$*

The property $\mathcal{A} \prec_{(k,t)} \mathcal{B}$ basically states a first order elementary equivalence among the extended structures $\langle \mathcal{A}, C_1, \ldots, C_t \rangle$ and $\langle \mathcal{B}, D_1, \ldots, D_t \rangle$ with respect to first order formulas of the form $\varphi(X_1, \ldots, X_t)$, viewing $X_1, \ldots, X_t$ as extra unary relation symbols. This condition is sufficient for extending elementary equivalence to $\mathcal{A}$ and $\mathcal{B}$ with respect to sentences in $\mathcal{SOMLP}$.

Theorem 1. *Let $r_1, \ldots, r_s$ be distinct non zero natural numbers. Let τ be a vocabulary and $\mathcal{A}$ and $\mathcal{B}$ be two τ–structures, with $|A| = m$, $|B| = m+1$, $m+1 > r_i$ and $m \equiv_{r_i} -1$ for $i = 1, \ldots, s$. If $\mathcal{A} \prec_{(k,t)} \mathcal{B}$ then, for all sentence φ of $\mathcal{SOMLP}(\tau)[r_1, \ldots, r_s]$, of first order quantifier rank $\leq k$ and at most t unary second order variables (free or not), we have*

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

Our next goal is to characterise $\mathcal{A} \prec_{(k,t)} \mathcal{B}$ in terms of winning strategies for a Ehrenfeucht–Fraïssé type of games. Recall that, for a positive integer k, a k *rounds first order Ehrenfeucht–Fraïssé game* is played by two players, commonly known as *Spoiler* and *Duplicator*, and the game board consists of two structures $\mathcal{D}$ and $\mathcal{E}$ of the same vocabulary. The players alternatively select elements in the structures, doing so in the opposite structure as the one selected by his opponent and through k rounds, being Spoiler the first one to move in each round. Let $d_1, \ldots, d_k$ be the elements selected in $\mathcal{D}$, and $e_1, \ldots, e_k$ the elements selected in $\mathcal{E}$. Duplicator wins if the substructure of $\mathcal{D}$ induced by $(d_1, \ldots, d_k)$ is isomorphic to the substructure of $\mathcal{E}$ induced by $(e_1, \ldots, e_k)$, under the function that maps d_i onto e_i, for $i = 1, \ldots, k$. The fundamental link between first order elementary equivalence and the k rounds first order Ehrenfeucht–Fraïssé game is given by the following theorem (cf. [2–§1.2] and [5–§6.1]).

Theorem 2 (Ehrenfeucht–Fraïssé). *For two structures $\mathcal{A}$ and $\mathcal{B}$ over the same vocabulary, and integer $k > 0$, the following two statements are equivalent:*

(i) *$\mathcal{A} \equiv_k \mathcal{B}$ (i.e., every first order sentence of quantifier rank $\leq k$ that is true in $\mathcal{A}$ is also true in $\mathcal{B}$, and vice versa).*
(ii) *Duplicator has a winning strategy in the k rounds first order Ehrenfeucht–Fraïssé game played on $\mathcal{A}$ and $\mathcal{B}$.* □

Our combinatorial game below is the classical game for monadic existential second order logic, to which we add strong restrictions on the possible cardinalities of both the structures upon the game is played and on the sets that the players choose as witnesses for second order variables (see [3] for definitions and a thorough analysis of games for monadic second order logic).

Definition 8. *Let τ be a relational vocabulary, s and k positive integers. Let $\mathcal{A}$ and $\mathcal{B}$ be two τ–structures such that $|B| = |A| + 1$. The proportional sets $(\mathcal{A}, \mathcal{B}, s, k)$–game (or simply the $(\mathcal{A}, \mathcal{B}, s, k)$–game) is played by Duplicator and Spoiler on $\mathcal{A}$ and $\mathcal{B}$ as follows:*

1. *Spoiler selects s subsets $S_1, \ldots, S_s$ of A.*
2. *Duplicator selects s subsets $T_1, \ldots, T_s$ of B, with $|S_i| \leq |T_i| \leq |S_i| + 1$, for $i = 1, \ldots, s$.*
3. *Both players play a k rounds first order Ehrenfeucht–Fraïssé game on the extended structures $\langle \mathcal{A}, S_1, \ldots, S_s \rangle$ and $\langle \mathcal{B}, T_1, \ldots, T_s \rangle$.*

Theorem 3. *Fix $k, s \in \mathbb{N}$, τ a vocabulary, $\mathcal{A}$ and $\mathcal{B}$ τ-structures with $|\mathcal{B}| = |\mathcal{A}| + 1$. $\mathcal{A} \prec_{(k,s)} \mathcal{B}$ if and only if Duplicator has a winning strategy in the $(\mathcal{A}, \mathcal{B}, s, k)$–game.* □

Now the tool for establishing non definability in $\mathcal{SOMLP}$ reads as follows.

Theorem 4. *Let $r_1, \ldots, r_n$ be distinct non zero natural numbers. Let τ be a relational vocabulary and K be a class of τ–structures. If for all positive integers k and s, there exists τ-structures $\mathcal{A}$ and $\mathcal{B}$ (that depend on k and s) such that $\mathcal{A} \in K$ and $\mathcal{B} \notin K$, $|\mathcal{B}| = |\mathcal{A}| + 1$, $|\mathcal{A}| \equiv_{r_i} -1$, for each $i = 1, \ldots, n$, and Duplicator has a winning strategy in the $(\mathcal{A}, \mathcal{B}, s, k)$–game, then K is not definable in $\mathcal{SOMLP}[r_1, \ldots, r_n]$.* □

5.1 Limitations in Expressive Power for A–$\mathcal{SOMLP}$

Recall that for a function g, the almost order $\leq_g$ on a universe A of a τ-structure $\mathcal{A}$, induces an equivalence relation $\sim_g$ on A. Let $[a]_g$ denote the $\sim_g$–equivalence class of $a \in A$, and $[A]_g := \{[a]_g : a \in A\}$. If, in addition, we ask of $\mathcal{A}$ to be consistent with $\leq_g$, then it makes sense to define the *quotient structure* $\mathcal{A}/_{\sim_g}$, as a τ-structure consisting of $[A]_g$ as its universe, and for a k-ary relation $R \in \tau$,

$$R^{\mathcal{A}/\sim_g} := \{([a_1]_g, \ldots, [a_k]_g) : (a_1, \ldots, a_k) \in R^{\mathcal{A}}\}$$

Furthermore, for a subset $B \subseteq A$ we define its $\leq_g$-*contraction* as $[B]_g := \{[b]_g : b \in B\}$; and for a subset $B \subseteq [A]_g$, its $\leq_g$-*expansion* is $(B)^g := \{a \in A : a \in [b]_g$ for some $[b]_g \in B\}$.

Definition 9. *Fix a sublinear function g and the almost order $\leq_g$. A $\leq_g$–cluster of models $\mathbf{C}$ is a collection of finite structures over same vocabulary τ, each consistent with $\leq_g$, and for each pair of τ-structures $\mathcal{A}$ and $\mathcal{B}$ in $\mathbf{C}$, their quotient under the equivalence relation $\sim_g$ are isomorphic, that is, $\mathcal{A}/_{\sim_g} \cong \mathcal{B}/_{\sim_g}$.*

Given $\mathcal{A}$ and $\mathcal{B}$ in the $\leq_g$-cluster $\mathbf{C}$, let F be an isomorphism from $\mathcal{A}/_{\sim_g}$ to $\mathcal{B}/_{\sim_g}$. Then, for $a \in A$ and $b \in B$, we write $a \equiv_{\mathbf{C}} b$ to indicate that $F([a]_g) = [b]_g$. Furthermore, for a subset $S \subseteq A$, the $\leq_g$-closure of S in $\mathcal{B}$ *is $cl_g(S, \mathcal{B}) := (F([S]_g))^g$ where $F([S]_g) := \{[b]_g \in [B]_g : F^{-1}([b]_g) \in [S]_g\}$.*

The following example gives an infinite family of sublinear functions that define almost orders.

Example 6. Fix $k \in \mathbb{N}$. Then $h_k(n) = 2r$, where $r \equiv_k n$, is a sublinear function. E.g., take $k = 3$, then $h_3(7) = 2$ and $h_3(8) = 4$. If $\mathcal{A}_7$ and $\mathcal{A}_8$ are sets of size 7 and 8 respectively, then $\mathcal{A}_7/_{\sim_{h_3}} \cong \mathcal{A}_8/_{\sim_{h_3}}$, and hence, they belong to the same $\leq_{h_3}$–cluster. □

The following lemma shows that pairs of structures, $\mathcal{A}$ and $\mathcal{B}$, that are in the same cluster and differ in one element, have the $\mathcal{A} \prec_{(k,s)} \mathcal{B}$ property.

Lemma 4. *Let g be a sublinear function and* $\mathbf{C}$ *an $\leq_g$-cluster of τ-models. Fix $\mathcal{A}$ and $\mathcal{B}$ in* $\mathbf{C}$*, with $|\mathcal{A}| = m$ and $|\mathcal{B}| = m+1$, and say $F : A/_{\sim_g} \to \mathcal{B}/_{\sim_g}$ is the isomorphism among the quotient structures. Then:*

(*i*) *For every first order formula $\phi(x_1, \ldots, x_s, \overline{Y})$ in $\mathcal{SOMLP}(\tau)$, for every a_1, $\ldots$, a_s in A, for every b_1, $\ldots$, b_s in B such that $a_i \equiv_{\mathbf{C}} b_i$, and for every sequence of subsets $S_1, \ldots, S_t$ of A, consistent with $\leq_g$, $\mathcal{A} \models \phi(a_1, \ldots, a_s, S_1, \ldots, S_t)$ iff $\mathcal{B} \models \phi(b_1, \ldots, b_s, cl_g(S_1, \mathcal{B}), \ldots, cl_g(S_t, \mathcal{B}))$;*

(*ii*) *If $S \subseteq A$ then $|S| \leq |cl_g(S, \mathcal{B})| \leq |S| + 1$.*

Corollary 2. *Let g be a sublinear function and* $\mathbf{C}$ *an $\leq_g$-cluster of τ-models. For $\mathcal{A}, \mathcal{B} \in \mathbf{C}$, with $|\mathcal{A}| = m$, $|\mathcal{B}| = m+1$, and $k, s \in \mathbb{N}$, we have $\mathcal{A} \prec_{(k,s)} \mathcal{B}$.* □

Combining the previous corollary with Theorem 1 we get

Corollary 3. *Let r_1, $\ldots$, r_k be distinct non zero natural numbers. Let g be a sublinear function, $\leq_g$ an almost order and* $\mathbf{C}$ *an $\leq_g$-cluster of τ–structures. For every pair of structures $\mathcal{A}$, $\mathcal{B}$ in* $\mathbf{C}$*, such that $|\mathcal{A}| = m$, $|\mathcal{B}| = m+1$, $m+1 > r_i$ and $m \equiv_{r_i} -1$, for every $i \leq k$, we have that, $\mathcal{A} \models \varphi$ implies $\mathcal{B} \models \varphi$, for all sentences φ of $\mathcal{SOMLP}(\tau)[r_1, \ldots, r_k]$* □

Theorem 5. *Let $r, r_1, \ldots, r_k$ be distinct non zero natural numbers, pairwise relatively prime. Then A–$\mathcal{SOMLP}[r_1, \ldots, r_k] \subsetneq A$–$\mathcal{SOMLP}[r_1, \ldots, r_k, r]$.* □

Corollary 4. *A-$\mathcal{SOMLP}[2] \subsetneq A$-$\mathcal{SOMLP}[2, 3] \subsetneq A$-$\mathcal{SOMLP}[2, 3, 5] \subsetneq \ldots$*

5.2 Limitations in Expressive Power for *A*–$\mathcal{SOLP}$

In this section we partially extend the separation result stated in Corollary 4 to second order variables of unbounded arity, that is, to A–$\mathcal{SOLP}$. It is a partial extension because we need to restrict our proportional quantifiers to be only of the form $(P(X) \leq 1/2)$, with X of arbitrary arity $r > 0$. Nonetheless, the result is interesting because it is precisely this type of quantifiers that defines $\mathcal{SOLP}Horn[2]$, which in the presence of order, captures **P**. Our main tool is a reshaping of Theorem 1 in the context of $\mathcal{SOLP}Horn[2]$.

Theorem 6. *Let τ be a vocabulary and $\mathcal{A}$ and $\mathcal{B}$ be two τ–structures, with $|A| = m$, $|B| = m+1$, $m+1 > 2$ and $m \equiv_2 -1$. If $\mathcal{A} \prec_{(k,t)} \mathcal{B}$ then, for all sentence φ of $\mathcal{SOLP}Horn(\tau)[2]$, of first order quantifier rank $\leq k$ and at most t second order variables (free or not), we have $\mathcal{A} \models \varphi$ implies $\mathcal{B} \models \varphi$* □

Theorem 7. *Let τ be a relational vocabulary and K be a class of τ–structures. If for all positive integers k and s, there exists τ-structures $\mathcal{A}$ and $\mathcal{B}$ (that depend on k and s) such that: $\mathcal{A} \in K$ and $\mathcal{B} \notin K$, $|\mathcal{B}| = |\mathcal{A}| + 1$, $|\mathcal{A}| \equiv_2 -1$, and Duplicator has a winning strategy in the $(\mathcal{A}, \mathcal{B}, s, k)$–game. Then K is not definable in $\mathcal{SOLP}Horn(\tau)[2]$.* □

Using as benchmark query: "the size of the model is a multiple of 3", which is definable in $A\text{-}\mathcal{SOLP}[2,3]$, we obtain

Corollary 5. $A\text{–}\mathcal{SOLP}Horn[2] \subsetneq A\text{–}\mathcal{SOLP}[2,3]$. □

References

1. Arratia, A., and Ortiz, C. (2004) Approximating the Expressive Power of Logics in Finite Models, *in: LATIN'04*, Lecture Notes in Comp. Sci. 2976 (Springer), 540–556.
2. Ebbinghaus, H.D., and Flum, J., *Finite Model Theory* (Springer-Verlag, 1995).
3. Fagin, R. (1997), Comparing the power of games on graphs, *Mathematical Logic Quarterly* **43**, 431–455.
4. Grädel, E. (1992), Capturing complexity classes by fragments of second order logic, *Theoretical Comp. Sci.* **101**, 35–57.
5. Immerman, N., *Descriptive Complexity* (Springer, 1998).
6. Libkin, L., and Wong, L., (2002) Lower bounds for invariant queries in logics with counting. *Theoretical Comp. Sci.* **288**, 153-180.
7. Stewart, I. (1994), Logical description of monotone NP problems, *J. Logic Computat.* **4** (4), 337-357.