Expressive power and complexity of a logic with quantifiers that count proportions of sets

Argimiro Arratia*†
Dpto. de Matemática Aplicada
Facultad de Ciencias
Universidad de Valladolid
Valladolid 47005, Spain
arratia@mac.uva.es

Carlos E. Ortiz‡
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 as, for example, the problem of deciding if at least a fraction $1/n$ of the set of vertices of a graph form a clique; 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 (restricted to almost orders), for it does not has the *bounded number of degrees property.* Moreover, we show that in this almost ordered setting different fragments of this logic vary in their expressive power, and show the existence of an infinite hierarchy inside our monadic language. We extend our inexpressibility result over almost ordered structure to a fragment of $\mathcal{SOLP}$, that in the presence of full order captures **P**. To obtain all our inexpressibility results we developed combinatorial games appropriate for these logics, whose application could go beyond the almost ordered models and hence are interesting by themselves.

Keywords: Descriptive complexity, counting quantifiers, almost order, definability, **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** ([7], [10]). The bottom line is that

*Contacting author
†Supported by Grants MOISES (TIN2005-08832-C03-02) and SINGACOM (MTM2004-00958), MEC–Spain

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

a solution to this problem should lead to a better understanding of the role of ordering in computations.

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 appropriate for the logic, makes the set of sentences of such logic, which are true in finite models, non recursive (this is a consequence of Trahtenbrot's Theorem [3]); and thus this logic hardly classifies as "good" programming paradigm.

On the other hand (and to what matters for the model theorist), 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 [10, § 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 (e.g. [4], [8], [11] among others). 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 [11] 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 [11] 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 are 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.

The paper is organized as follows. In Sections 2 and 3, we introduce the logic $\mathcal{SOLP}$ and briefly discuss its expressive power and its relationship with existential second order logic (∃SO). We define the fragments $\mathcal{SOLP}[k]$, for every integer $k > 1$, obtained by restricting the proportional quantifiers to $(P(X) \leq q/k)$ or $(P(X) \geq q/k)$ for $0 < q < k$ natural numbers; and the monadic fragment $\mathcal{SOMLP}$ obtained from $\mathcal{SOLP}$ by restricting the proportional quantifiers to act upon unary second order variables only. We define the Horn and the Krom fragments of $\mathcal{SOLP}$, and just as their analogous fragments within ∃SO, we show that with built–in order, these fragments capture **P** and **NL** (nondeterministic log–space), respectively. Although these results are on the surface similar to Grädel's in [6], there is some work to do to keep account of the proportions counted by our quantifiers, and further, these logics give a finer description of the aforesaid classes. In Section 4 we study the behaviour of $\mathcal{SOLP}$ with almost order; its expressive power, and most importantly show that this approximation makes the logic not too tame since it does not have the BNDP. In Section 5 we build a separation tool appropriate for fragments of $\mathcal{SOLP}$ based on a natural generalisation of games for existential second order logic (see [5] for a survey). This tool is used to obtain the inexpressibility results for fragments of $\mathcal{SOLP}$ with almost order, first within the monadic sub–languages and then for the Horn part of $\mathcal{SOLP}[2]$. In Section 6, we give our conclusions and draw some lines of further research related to this work.

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 by Ebbinghaus and Flum [3] and by Immerman [10]. 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. First order logic is denoted by FO; second order logic by SO; existential second order logic by ∃SO. We use the symbol $\equiv_k$, with k a positive integer, to either denote that two structures $\mathcal{A}$ and $\mathcal{B}$ are elementary equivalent with respect to formulas of quantifier rank up to k (and write $\mathcal{A} \equiv_k \mathcal{B}$), or that an integer n is equivalent modulo k to an integer m (and write $n \equiv_k m$). There should be no confusion since the context makes our use of $\equiv_k$ clear.

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 2.1 *The Second Order Logic of Proportional quantifiers, denoted $\mathcal{SOLP}$, 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, …, 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*, ...,* 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* $0 \leq q < r_i$*. 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 contain X or any of the variables in $\overline{Y}$ as a relation symbol. Let r be a rational in $(0,1)$. Then

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

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

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

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

Example 2.2 *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* $|A|$ $\}$.

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

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

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 $r = 1/n$ *this problem is* **P***–complete under first order reductions.* $\square$

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

Example 2.3 *Let* $\tau_g = \{E\}$*,* E *a binary relation symbol.* τ_g*–structures are seen as graphs. 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*

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

For any rational $r \in (0,1)$, CLIQUE$_{\geq r}$ is **NP**-complete via *logspace reducibilities.* To prove this fact we describe a reduction from the standard CLIQUE problem, which is the problem of deciding, for a graph G and an integer $k > 0$, if G contains a complete subgraph of size at least k.

Proposition 2.4 *For any rational* $r \in (0,1)$*, we have* CLIQUE $\leq_{log}$ CLIQUE$_{\geq r}$

Proof: Let $\langle G, k\rangle$ be an instance of CLIQUE, and $n = |V(G)|$. We build an instance G' of CLIQUE$_{\geq r}$ as follows.

Let $t = \frac{k}{rn}$ and $m = \lceil \frac{rtn-k}{1-r} \rceil$. Then G' consists of t disjoint copies of G and a complete graph of m vertices, K_m. Each vertex of each of the t copies of G is connected to all the vertices in K_m. There are no other extra edges. Then $|V(G')| = tn + m$ and $\omega(G') = \omega(G) + m$, where $\omega(G)$ stands for the cardinality of the greatest possible clique in G. The correctness of this reduction will be consequence of the following claim.

Claim 2.5 $k + m = \lceil r(tn + m) \rceil$

Proof: $(\geq)$: $m = \lceil \frac{rtn-k}{1-r} \rceil$. Then $m \geq \frac{rtn-k}{1-r} \iff (1-r)m \geq rtn - k \iff k + m \geq r(tn+m)$

$$\begin{aligned}
(\leq):\ (1-r)m &= (1-r)\lceil \frac{rtn-k}{1-r} \rceil \leq (1-r)\left(\frac{rtn-k}{1-r} + 1\right) \\
&= rtn - k + 1 - r < rtn - k + 1 \\
&\iff m(1-r) + k - 1 < rtn \\
&\iff k + m - 1 < r(tn+m) \\
&\iff k + m \leq \lceil r(tn+m) \rceil \quad \Box
\end{aligned}$$

Now,

$$\begin{aligned}
\langle G, k\rangle \in \text{CLIQUE} &\iff \omega(G) \geq k \iff \omega(G') \geq k + m \\
&\iff \omega(G') \geq \lceil r(tn+m) \rceil \ \text{(by the claim)} \\
&\iff \omega(G') \geq r(tn+m) \iff G' \in \text{CLIQUE}_{\geq r} \quad \Box
\end{aligned}$$

The following example shows that we can express in $\mathcal{SOMLP}$ some well known **NP**–complete problems, such as k–colorability, in its exact form and not just a proportional version of it.

Example 2.6 *Let* $\tau_g = \{E\}$ *and, as before, our problem is a class of graphs. We present a* τ_g*–sentence in* $\mathcal{SOMLP}[k]$ *that says that a graph is* k*–colorable, where we represent each of the possible* k *colors by unary relation symbols* $Z_1, \ldots, Z_k$*. Let* ψ *says "*each vertex has exactly one of the possible k colors*". Formally,*

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

Let θ *says* "two vertices with the same color are not connected by an edge"*; that is,*

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

Then, the following sentence Ψ_k*, which is in* $\mathcal{SOMLP}[k](\tau_g)$*, expresses that* "the graph is k-colorable"*:*

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

For 2–colorability we can get by without the quantifiers of the form $(P(Z) \leq 1/2)$ *by using as colors a set and its complement. The sentence*

$$\begin{aligned}\Psi_2 \quad := \quad & \left(P(B) \geq \frac{1}{2}\right)[\forall x \neg E(x,x) \wedge \forall x \forall y((B(x) \wedge B(y) \longrightarrow \neg E(x,y)) \\ & \wedge (\neg B(x) \wedge \neg B(y) \longrightarrow \neg E(x,y)))]\end{aligned}$$

defines the problem of 2-colorability, (The above sentence can be shortened using $\oplus$ *for exclusive or. Thus*
$\Psi = \left(P(B) \geq \frac{1}{2}\right)[\forall x \neg E(x,x) \wedge \forall x \forall y (E(x,y) \longrightarrow (B(x) \oplus B(y)))]$*.)*

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

Remark 2.7 *Any formula in* $\exists SO$ *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 following 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 2.7, $\mathcal{SOLP}$ subsumes $\exists$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 $\exists$SO in [6], 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 *Horn* (for **NL** it will be *Krom*) with respect to second order variables. 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 $\exists$SO.

Definition 3.1 *Let* $\tau = \{R_1, \ldots, R_m, C_1, \ldots, C_s\}$ *be some vocabulary with relation symbols* R_1*, …,* R_m*, and constant symbols* C_1*, …,* C_s*, and let* X_1*, …,* X_r *be second order variables of arity* k_1*, …,* k_r*, respectively. We will consider first order formulae* *over $\tau \cup \{X_1, \ldots, X_r\}$, and extra binary relation symbol $=$ (equality) and the constant $\perp$ (standing for false), of the form*

$$\alpha := \forall \overline{z}[\bigwedge_{i=1}^{m}(\psi_{i1} \wedge \ldots \wedge \psi_{is} \longrightarrow \varphi_i)] \tag{2}$$

where each φ_i is either $X_j(\overline{u}_j)$ (where $\overline{u}_j$ denotes a k_j-tuple of first order terms, $j = 1, \ldots, r$) or $\perp$, and $\psi_{i1}, \ldots, \psi_{is}$ are atomic or negation of atomic $(\tau \cup \{X_1, \ldots, X_r\})$-formulas except that any occurrence of the variables X_j must be positive (there are no restrictions on the predicates in τ or $=$). So, α is a universally quantified conjunction of clauses which are Horn only with respect to the second order variables $X_1, \ldots, X_r$.

The logic $\mathcal{SOLP}Horn$ is the set of formulae 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 formula of the form (2) over some vocabulary τ and second order variables $X_1, \ldots, X_r$. (Observe that we are not including quantifiers of the form $(P(X) \geq r)$.)

Example 2.2 shows that 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 3.2 *Each set of finite structures that satisfy a sentence in $\mathcal{SOLP}Horn$ is in* **P***.*

Proof: Let θ be a sentence in $\mathcal{SOLP}_{Horn}$ and let $\mathcal{A}$ be an appropriate model of size n. We have to describe a polynomial time procedure that decides whether $\mathcal{A} \models \theta$ or not. By definition θ has the form

$$Q_1(X_1) \cdots Q_r(X_r) \forall \overline{z}[\bigwedge_{i=1}^{m}(\psi_{i1} \wedge \ldots \wedge \psi_{is} \longrightarrow \varphi_i)]$$

where $Q_i(X_i)$ is $(P(X_i) \leq t_i)$, t_i a rational in $(0, 1)$, and $\psi_{i1}, \ldots, \psi_{is}, \varphi_i, X_1, \ldots, X_r$, verify the restrictions imposed in the Definition 3.1.

The idea is to substitute the universal quantifiers $\forall z_i$ for a conjunction (of length polynomial in n) containing all possible variants of $\bigwedge_{i=1}^{m}(\psi_{i1} \wedge \ldots \wedge \psi_{is} \longrightarrow \varphi_i)$ with the variables z_i's replace by values $v_i \in \{0, \ldots, n-1\}$. Then substitute the relation symbols of the vocabulary of $\mathcal{A}$, and equalities and inequalities, by their truth values in $\mathcal{A}$. We are left with a big conjunction of clauses containing atomic expressions of the form $X_i(v_{i1}, \ldots, v_{ik_i})$. Then replace each of these atomic expressions by a Boolean variable $x_i^{v_{i1}, \ldots, v_{ik_i}}$ and thus obtain a Horn Boolean proposition Ψ with polynomial in n many clauses and variables. It is known that the satisfiability of a Horn Boolean proposition can be done in **P** [12], and hence, we have a polynomial time algorithm for constructing a truth valuation of the set of variables of Ψ that makes this proposition true.

However, we want in addition to count the number of variables $x_i^{v_{i1}, \ldots, v_{ik_i}}$ that are assigned the value *true*, for each $i = 1, \ldots, r$, so that we can estimate the size of each

X_i and test if it is *below* the given upper bound $t_i n^{k_i}$. Therefore, it is necessary to compute a satisfying assignment for Ψ which has the *minimal* number of variables set to *true*. This is not hard to do in polynomial time, and here is where the Horn form of the formulas play its role: Each clause of Ψ (putted in disjunctive form) has at most one variable $x_i^{v_{i1},\ldots,v_{ik_i}}$ appearing positive.

We then proceed as follows.

> While there are clauses with unassigned variables appearing positive, do the following: put all variables appearing positive in singleton clauses to *true* and remove their negations from other clauses. If none of the clauses gets empty (so we have a satisfying assignment), output the variables set to *true* and count. For each $i = 1, \ldots, r$, if the number of variables $x_i^{v_{i1},\ldots,v_{ik_i}}$ that were assigned the value of *true* is less than or equal to $t_i n^{k_i}$, we accept, otherwise reject.

(Observe that if no clause has positive appearance of variables as literal, then the solution is trivial: all X_i are empty and the formula is trivially satisfied. Also observe that these arguments would not work if we were to consider quantifiers of the form $(P(X) \geq r)$.) $\square$

Thus, according to this lemma, our problem NOT-IN-CLOS$_{\leq r}$ is in **P**. We shall see 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 [13] Stewart shows that PS is complete for **P** via quantifier free first order reductions that include built-in order; in fact, via *projections* (see [13] for details and also [10, § 11.2]). Using that quantifier free projections are transitive, we get the following result.

Lemma 3.3 *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.*

Proof: We exhibit a projection from the complement of the problem PS to NOT-IN-CLOS$_{\leq 1/2}$. We will use a built-in successor (denoted $+1$) and built-in constants 0 and *max* standing for the first and the last element of any instance of the problem, which according to Stewart [13], can also be considered as the source (the 0) and the target (the *max*) of our path systems.

Let $\mathcal{A} = \langle A, R, s, t \rangle$ be an instance of the complement of PS (as ordered structure $s = 0$ and $t = max$ according to our comments above). Define $\mathcal{A}' = \langle A', R', s', t' \rangle$ as follows: its universe $A' = A \times \{0\} \cup A \times \{max\}$, and set of rules

$$\begin{aligned} R' &= \{((x_1,x_2),(y_1,y_2),(z_1,z_2)) : x_2 = y_2 = z_2 = 0 \wedge R(x_1,y_1,z_1)\} \\ &\cup \{((x_1,x_2),(y_1,y_2),(z_1,z_2)) : y_2 = z_2 = max \wedge x_2 = 0 \wedge x_1 = max \\ &\quad \wedge z_1 = y_1 + 1 \wedge z_1 \neq max\} \\ s' &= (0,0) \\ t' &= (max, max) \end{aligned}$$

$\mathcal{A}'$ is a copy of $\mathcal{A}$ (elements of the form $(x,0)$) and extra elements (x, max), as many as $|A|$, which are all accessible from $(max, 0)$ (and from no other element of the form $(x,0)$). Then, if max is not accessible from 0 in $\mathcal{A}$ then $(max,0)$ is not accessible from $(0,0)$ in $\mathcal{A}'$ and hence, no (x, max) is accessible from $(0,0)$ (in particular (max, max) is not in the R–closure containing $(0,0)$), and $(0,0)$ reaches less than 1/2 of elements in $\mathcal{A}'$. Conversely, if max is accessible from 0 in $\mathcal{A}$, we have that $(max,0)$ is accessible from $(0,0)$ in $\mathcal{A}'$, and from $(max,0)$ we can access any (x, max) (for $x > 0$) by the rule $((max,0),(x-1,max),(x,max))$; hence, $(0,0)$ access more than 1/2 of elements, including (max, max).

Since PS, and its complement, are complete for **P** via projections (with successor, see Stewart [13]), the result follows. □

An argument similar to the preceding proof shows that, for integer $k > 2$, NOT-IN-CLOS$_{\leq 1/k}$ is complete for **P** via qfp: using successor relation we can define new constants 0, 1, ..., $k-1$, and from an instance $\mathcal{A} = \langle A, R, 0, max\rangle$ of the complement of PS build and instance $\mathcal{A}'$ of NOT-IN-CLOS$_{\leq 1/k}$ with universe $A' = A \times \{0\} \cup A \times \{1\} \cup \ldots \cup A \times \{k-1\}$, where on the first component $(A \times \{0\})$ the new relation R' behaves like R and on the other components their elements are all R'–accessible from $(max, 0)$ (and placing (max, max) in the last component).

Corollary 3.4 *Every problem in* **P** *is a set of finite structures, with built-in successor, that satisfy a sentence in* $\mathcal{SOLP}Horn$.

Proof: Every problem in **P** is reducible to NOT-IN-CLOS$_{\leq 1/2}$ via quantifier free projections; NOT-IN-CLOS$_{\leq 1/2}$ is definable in $\mathcal{SOLP}Horn$ and this logic is closed via qfp's. □

From Lemma 3.2 and Corollary 3.4 we obtain that the logic $\mathcal{SOLP}Horn$ (in fact, $\mathcal{SOLP}Horn[2]$) captures **P**, over finite and ordered structures. Moreover, by the comments following Lemma 3.3, we can refine this to the following result.

Corollary 3.5 *For each integer* $k > 1$, $\mathcal{SOLP}Horn[k] =$ **P**, *with respect to finite structures, ordered with a built-in successor.* □

For logarithmic space bounded classes we have the following example.

Example 3.6 *Let* $\tau = \{E, s\}$ *where* E *is a binary relation symbol and* s *is a constant symbol. We think of* τ*-structures as graphs (undirected) with a specify vertex* s *(the source). Let* r *be a rational with* $0 < r < 1$. *We define*

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*

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

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

Again, inspired on work by Grädel [6] we define:

Definition 3.7 *Let* τ *and* $X_1, \ldots, X_r$ *be as in Definition 3.1. We now consider first order formulae* α *over* $\tau \cup \{X_1, \ldots, X_r\} \cup \{=, \perp\}$ *of the following form:* 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* $X_1, \ldots, X_r$ *(there are no restrictions on the predicates in* τ *or* $=$*).*

The logic $\mathcal{SOLP}Krom$ *is the set of formulae 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 formula of the form described above, over some vocabulary* τ *and second order variables* $X_1, \ldots, X_r$*. (Observe that this time we are not including quantifiers of the form* $P(X) \leq r$*).)*

The sentence defining $\mathrm{NCON}_{\geq r}$ is in $\mathcal{SOLP}Krom$. The complement of $\mathrm{NCON}_{\geq r}$ is the problem

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

Lemma 3.8 *The problem* $\mathrm{NCON}_{\geq r}$ *is in* **NL***.*

Proof: We show $\mathrm{NCON}_{\leq r}$ is definable in the logic posTC[FO_s] (the positive fragment of the Transitive Closure logic), and use Immerman's theorem that this logic captures **NL** ([9]), and that this class is closed under complement (see also [10]). We first define an edge relation between tuples of arity 2. Let

$$\theta(x, y, x', y') := x < x' \wedge suc(y, y') \wedge \mathrm{TC}[u, v : E(u, v)](s, x')$$

Observe that a θ–edge has the form $((x, y), (x', y + 1))$ with $x < x'$ and x' connected to s. Let $k = max - \lfloor r \cdot max \rfloor$ (this constant value is definable in posTC[FO_s]). Then the sentence $\exists z \mathrm{TC}[(x, y), (x', y') : \theta(x, y, x', y')]((0, 0), (z, k))$ defines a path of $k + 1$ tuples, $(0, 0)$, $(v_1, 1)$, $\ldots$, (v_k, k) where each v_i is connected to s. Therefore, a model that satisfies this sentence has at least k vertices connected to s, which is equivalent to saying that it has at most a fraction r of its vertices not connected to s, and conversely, any graph in $\mathrm{NCON}_{\leq r}$ satisfies this sentence. $\square$

We next show that for $r = 1/2$ the problem $\mathrm{NCON}_{\geq r}$ is hard for **NL** via qfp's. Recall that the Transitive Closure problem is the following class of undirected graphs, which is complete for **NL** via qfp's [10].

TC **:=** $\{\mathcal{A} = \langle A, E, s, t\rangle : \langle A, E\rangle$ is a graph with two specified vertices, s and t, and there is a path from s to $t\}$

Lemma 3.9 NCON$_{\geq 1/2}$ *is complete for* **NL** *via quantifier free projections.*

Proof: We informally describe a reduction of the complement of TC to NCON$_{\geq 1/2}$, and leave to the reader to see that our reduction is a qfp. Given a graph $G = \langle V, E\rangle$ and two specific vertices s and t in G, build the graph D_G with vertex set $V(D_G) = V \times \{s\} \cup V \times \{t\}$, source $s' = (s, s)$ and edge set

$$\begin{array}{rcl} E(D_G) & = & \{((x_1, x_2), (y_1, y_2)) : x_2 = y_2 = s \wedge E(x_1, y_1)\} \\ & \cup & \{((x_1, x_2), (y_1, y_2)) : y_2 = t \wedge x_2 = s \wedge x_1 = t\} \end{array}$$

Then

$$(G, s, t) \notin \text{TC} \iff (D_G, s) \in \text{NCON}_{\geq 1/2}$$

(If (G, s, t) is in TC, then more than 1/2 of the nodes in D_G are reachable from (s, s) since (t, s) has an edge to every node of the form (x, t) and (s, s) is connected to (t, s); hence (D_G, s) is not in NCON$_{\geq 1/2}$. If (G, s, t) is not in TC then less than 1/2 of the nodes in D_G are reachable from (s, s) since (t, s) can not be reached from (s, s); hence (D_G, s) is in NCON$_{1/2}$.) $\square$

The above result is generalisable to $r = 1/n$ for any natural n. From the previous lemmas, we conclude that over finite structures, ordered with built–in successor, $\mathcal{SOLP}Krom$ captures **NL**. Furthermore,

Corollary 3.10 *For each integer* $k > 1$, $\mathcal{SOLP}Krom[k] =$ **NL**, *with respect to finite structures, ordered with a built-in successor.* $\square$

An undirected version of NCON$_{\geq r}$ give us an example of a qfp-complete problem for the symmetric logarithmic space class, **SL**, which is also definable in our logic.

Remark 3.11 *According to Corollary 3.5,* $\mathbf{P} = \mathcal{SOLP}Horn[2]$ *and obviously*

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

The chain (3) 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 now turn to the study of the logic $\mathcal{SOLP}$ over structures with a weak form of built–in order. We begin with two preliminary definitions. The first is the notion of almost linear order from [11], with respect to natural valued sublinear functions. (Recall that a function $g : \mathbb{N} \to \mathbb{N}$ is sublinear if, for all $n \in \mathbb{N}$, $g(n) < n$.)

Definition 4.1 *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 at least* $|A| - g(|A|)$ *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 4.2 *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*

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

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

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

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

Our first lemma shows that 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 4.3 *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 $\leq_g$.* $\Box$

Definition 4.4 *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–$\mathcal{SOMLP}$ and 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 4.5 *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 $\leq_g$,

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

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 $\{a_1, b_1\}$, ..., $\{a_r, b_r\}$, and $l = m - 2r$ with one element, say these are $\{c_1\}$, ..., $\{c_l\}$. Hence, $m = 2r + l$ and since m is even, l must be even. We proceed to construct the disjoint sets $B^{\mathcal{A}}$ and $C^{\mathcal{A}}$, interpretations of B and C in $\mathcal{A}$. 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 these interpretations need to be 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. $\Box$

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.

Example 4.6 *For another example of the power of expression of A–$\mathcal{SOMLP}$ note that, for each $k > 1$, the sentence $\Psi_k \in \mathcal{SOMLP}[k]$ described in Example 2.6, which defines the k–colorability problem hold true also for structures consistent with any almost order $\leq_g$ (the satisfaction of the property of being k–colorable is indistinct of the order); hence, each is a sentence in A–$\mathcal{SOMLP}[k]$.* $\Box$

It was shown in [11] 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 4.7 *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 (see [8], [11]), and it is shown in [11] 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 typical for FO queries over unordered structures" *[11]. 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 [11] in that it does not have the BNDP.*

Consider the quantifier free formula $path(x, y, U)$ in A–$\mathcal{SOMLP}(\{E\})$ stating 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.*

$$\underbrace{\begin{array}{ccccc} \bullet & & \bullet & & \bullet \\ \uparrow\downarrow & \to & \uparrow\downarrow & \to \dots & \uparrow\downarrow \\ \bullet & & \bullet & & \bullet \end{array}}_{g(2n)} \underbrace{\to \bullet \to \bullet \dots \to \bullet}_{2n - g(2n)}$$

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.* $\square$

5 Playing games in $\mathcal{SOMLP}$

We begin designing our tools for showing separation in fragments of $\mathcal{SOLP}$.

Definition 5.1 *Let* τ *be a vocabulary and* $\mathcal{A}$ *and* $\mathcal{B}$ *be two* τ*–structures, with* $|B| = |A| + 1$*. Let* k*,* t *and* r *be three positive integers. By* $\mathcal{A} \prec_{(k,t,r)} \mathcal{B}$ *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 second order predicates* X_1*,* $\ldots$*,* X_t*, all of arity* r*, for all subsets* C_1*,* $\ldots$*,* C_t *of* A^r*, there exist subsets* D_1*,* $\ldots$*,* D_t *of* B^r*, such that*
>
> - $|C_i| \leq |D_i|$*, for* $i = 1, \ldots, t$*, and*
> - $\mathcal{A} \models \varphi(C_1, \ldots, C_t)$ *implies* $\mathcal{B} \models \varphi(D_1, \ldots, D_t)$

Remark 5.2 *We are assuming that the relation symbols* X_1*,* $\ldots$*,* X_t *have all the same arity* r*. There is no loss in expressive power by imposing this convenient restriction. Also, note that in the particular case of* $r = 1$*, i.e. the X's are all monadic, their interpretations in* $\mathcal{A}$ *and in* $\mathcal{B}$ *can be assumed to verify that* $|C_i| \leq |D_i| \leq |C_i| + 1$*.*

The property $\mathcal{A} \prec_{(k,t,r)} \mathcal{B}$ 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 relation symbols. We shall prove next that, in the monadic case, the condition $\mathcal{A} \prec_{(k,t,1)} \mathcal{B}$ is sufficient for extending elementary equivalence to $\mathcal{A}$ and $\mathcal{B}$ with respect to sentences in $\mathcal{SOMLP}$.

Theorem 5.3 *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 $\mathcal{A} \prec_{(k,t,1)} \mathcal{B}$ then, for all sentences 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), $X_1, \ldots, X_t$, we have for all $S_1, \ldots, S_t$ subsets of A, there are $T_1, \ldots, T_t$ subsets of B, with $|S_i| \leq |T_i| \leq |S_i| + 1$, such that*

$$\mathcal{A} \models \varphi(S_1, \ldots, S_t) \textit{ implies } \mathcal{B} \models \varphi(T_1, \ldots, T_t)$$

Proof: We proceed by induction on the syntactic complexity of φ. The case of φ being a first order formula of quantifier rank $\leq k$ is given by hypothesis. Thus, we are left with the case of φ having proportional quantifiers.

Proportional Quantifiers: Our inductive hypothesis reads as follows. Given a formula $\varphi(X_1, \ldots, X_t)$ in $\mathcal{SOMLP}(\tau)[r_1, \ldots, r_s]$, with $X_1, \ldots, X_t$ monadic second order free variables, and given subsets $S_1, \ldots, S_t$ of A, there exists subsets $T_1, \ldots, T_t$ of B, with $|S_i| \leq |T_i| \leq |S_i| + 1$, such that

$$\mathcal{A} \models \varphi(S_1, \ldots, S_t) \text{ implies } \mathcal{B} \models \varphi(T_1, \ldots, T_t).$$

Now we add a proportional quantifier. We have two cases to consider.

We consider first the formula, $(P(Y) \geq q_{ij}/r_i)\varphi(X_1, \ldots, X_{t-1}, Y)$. Let $m = |A|$ such that $m + 1 > r_i$ and $m \equiv_{r_i} -1$, for every $i \leq s$. Fix a sequence of subsets $S_1, \ldots, S_{t-1}$ of A. Let n be a natural number such that $m = nr_i + r_i - 1$.

Now, if $\mathcal{A} \models (P(Y) \geq q_{ij}/r_i)\varphi(S_1, \ldots, S_{t-1}, Y)$ and since $gcd(r_i, m) = 1$, then there is $D \subseteq A$ such that $\mathcal{A} \models \varphi(S_1, \ldots, S_{t-1}, D)$ and

$$|D| > \frac{q_{ij}m}{r_i} = \frac{q_{ij}(nr_i + r_i - 1)}{r_i} = q_{ij}(n+1) - \frac{q_{ij}}{r_i}$$

Since $q_{ij} < r_i$, we obtain that $|D| \geq q_{ij}(n+1)$. By induction hypothesis $\mathcal{B} \models \varphi(T_1, \ldots, T_{t-1}, E)$, for some subsets $T_1, \ldots, T_{t-1}$ and E of B, such that $|S_i| \leq |T_i| \leq |S_i| + 1$, and

$$|E| \geq |D| \geq q_{ij}(n+1) = \frac{q_{ij}}{r_i}(n+1)(r_i) = \frac{q_{ij}}{r_i}(m+1),$$

which implies that $\mathcal{B} \models (P(Y) \geq q_{ij}/r_i)\varphi(T_1, \ldots, T_{t-1}, Y)$, which is the desired result.

Next we consider the formula $(P(Y) \leq q_{ij}/r_i)\varphi(X_1, \ldots, X_{t-1}, Y)$. Let $m = |A|$ such that $m + 1 > r_i$ and $m \equiv_{r_i} -1$ for every $i \leq s$. Fix a sequence of subsets $S_1, \ldots, S_{t-1}$ of A. Let n be a natural number such that $m = nr_i + r_i - 1$.

Now, if $\mathcal{A} \models (P(Y) \leq q_{ij}/r_i)\varphi(S_1, \ldots, S_{t-1}, Y)$ and since $gcd(r_i, m) = 1$, then there is $D \subseteq A$ such that $\mathcal{A} \models \varphi(S_1, \ldots, S_{t-1}, D)$ and

$$|D| < \frac{q_{ij}m}{r_i} = \frac{q_{ij}(nr_i + r_i - 1)}{r_i} = q_{ij}(n+1) - \frac{q_{ij}}{r_i}$$

Since $q_{ij} < r_i$, we obtain that $|D| \leq q_{ij}(n+1) - 1$. By induction hypothesis $\mathcal{B} \models \varphi(T_1, \ldots, T_{t-1}, E)$, for some subsets $T_1, \ldots, T_{t-1}$ and E of B, such that $|S_i| \leq |T_i| \leq |S_i| + 1$, and

$$|E| \leq |D| + 1 \leq q_{ij}(n+1) = \frac{q_{ij}}{r_i}(n+1)(r_i) = \frac{q_{ij}}{r_i}(m+1),$$

which implies that $\mathcal{B} \models (P(Y) \leq q_{ij}/r_i)\varphi(T_1, \ldots, T_{t-1}, Y)$, which is the desired result. □

Our next goal is to characterise $\mathcal{A} \prec_{(k,t,r)} \mathcal{B}$ 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. [3, §1.2] and [10, §6.1]).

Theorem 5.4 (Ehrenfeucht–Fraïssé) *For two structures $\mathcal{A}$ and $\mathcal{B}$ over the same vocabulary, and positive integer k, 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 [5] for definitions and a thorough analysis of games for monadic second order logic). For lack of a better name we chose to call our games "*proportional sets game*".

Definition 5.5 *Let τ be a relational vocabulary, k, t and r positive integers. Let $\mathcal{A}$ and $\mathcal{B}$ be two τ–structures such that $|B| = |A| + 1$. The proportional sets $(\mathcal{A}, \mathcal{B}, k, t, r)$–game (or simply the $(\mathcal{A}, \mathcal{B}, k, t, r)$–game) is played by Duplicator and Spoiler on $\mathcal{A}$ and $\mathcal{B}$ as follows:*

1. *Spoiler selects t subsets of r–tuples, $S_1, \ldots, S_t$ of A^r.*

2. *Duplicator selects t subsets of r–tuples, $T_1, \ldots, T_t$ of B^r, with $|S_i| \leq |T_i|$, for $i = 1, \ldots, t$.*

3. *Both players play a k rounds first order Ehrenfeucht–Fraïssé game on the extended structures $\langle \mathcal{A}, S_1, \ldots, S_t \rangle$ and $\langle \mathcal{B}, T_1, \ldots, T_t \rangle$.*

Theorem 5.6 *Fix $k, t, r \in \mathbb{N}$, τ a vocabulary, $\mathcal{A}$ and $\mathcal{B}$ τ-structures with $|\mathcal{B}| = |\mathcal{A}|+1$. $\mathcal{A} \prec_{(k,t,r)} \mathcal{B}$ if and only if Duplicator has a winning strategy in the $(\mathcal{A}, \mathcal{B}, k, t, r)$–game.*

Proof: The proof is analogous to the one for the classical Ehrenfeucht–Fraïssé Theorem (Theorem 5.4). Assume $\mathcal{A} \prec_{(k,t,r)} \mathcal{B}$, and let $S_1, \ldots, S_t$ be the sets selected by Spoiler for $\mathcal{A}$. Let

$$\Phi := \bigwedge \{\psi : qr(\psi) \leq k \ \ \& \ \ \langle \mathcal{A}, S_1, \ldots, S_t \rangle \models \psi\}$$

Φ is the conjunction of all sentences of quantifier rank $\leq k$ that are true of $\langle \mathcal{A}, S_1, \ldots, S_t \rangle$. This is a finite conjunction (up to equivalence of formulas) because there are only finitely many inequivalent formulas in $\mathrm{FO}(\tau \cup \{X_1, \ldots, X_t\})$. By hypothesis, there exists subsets $T_1, \ldots, T_t$ of B^r such that $|S_i| \leq |T_i|$ and $\mathcal{B} \models \Phi(T_1, \ldots, T_t)$. Hence, $\langle \mathcal{A}, S_1, \ldots, S_t \rangle \equiv_k \langle \mathcal{B}, T_1, \ldots, T_t \rangle$ and, thus, combining with Theorem 5.4, Duplicator has a winning strategy in the $(\mathcal{A}, \mathcal{B}, k, t, r)$–game.

Conversely, assume $\mathcal{A} \prec_{(k,t,r)} \mathcal{B}$ is not true. Then, there exists $\Phi(X_1, \ldots, X_t) \in \mathrm{FO}(\tau \cup \{X_1, \ldots, X_t\})$ of quantifier rank $\leq k$, subsets $S_1, \ldots, S_t$ of A^r, such that for all subsets $T_1, \ldots, T_t$ of B^r with $|S_i| \leq |T_i|$, we have $\mathcal{A} \models \Phi(S_1, \ldots, S_t)$ but $\mathcal{B} \models \neg\Phi(T_1, \ldots, T_t)$. Then Spoiler selects those subsets $S_1, \ldots, S_t$ of A^r witnessing the truth of Φ in $\mathcal{A}$, and whatever are the predicates of arity r chosen by Duplicator in $\mathcal{B}$, the extended structures $\langle \mathcal{A}, S_1, \ldots, S_t \rangle$ and $\langle \mathcal{B}, T_1, \ldots, T_t \rangle$ disagree in Φ and, by Theorem 5.4, Spoiler can win the k rounds first order game. $\Box$

What we are going to do is to sharpen, for some fragments of $\mathcal{SOLP}$, the usual tool for establishing non definability of any class of structures in a logic. The tool we are referring to reads as follows (we state it in terms of fragments of $\mathcal{SOLP}$).

Theorem 5.7 *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 sentences $\Phi \in \mathcal{SOLP}(\tau)[r_1, \ldots, r_n]$, there exists τ-structures $\mathcal{A}$ and $\mathcal{B}$ such that:*

(i) $\mathcal{A} \in K$ *and* $\mathcal{B} \notin K$,

(ii) $\mathcal{A} \models \Phi$ *implies* $\mathcal{B} \models \Phi$.

Then K is not definable in $\mathcal{SOLP}(\tau)[r_1, \ldots, r_n]$. $\Box$

By Theorem 5.3 and Theorem 5.6 we can refine the preceding general definability tool for the monadic case to the following result with a more combinatorial flavour.

Theorem 5.8 *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 t, there exists τ-structures $\mathcal{A}$ and $\mathcal{B}$ (that depend on k and t) such that:*

(i) $\mathcal{A} \in K$ *and* $\mathcal{B} \notin K$,

(ii) $|\mathcal{B}| = |\mathcal{A}| + 1$, $|\mathcal{A}| \equiv_{r_i} -1$, *for each* $i = 1, \ldots, n$, *and*

(iii) Duplicator has a winning strategy in the $(\mathcal{A}, \mathcal{B}, k, t, 1)$*–game.*

Then K is not definable in $\mathcal{SOMLP}(\tau)[r_1, \ldots, r_n]$. $\Box$

We apply this refine definability tool in the next section.

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 $R \in \tau$,

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

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 $(B)^g := \{a \in A : a \in [b]_g \text{ for some } [b]_g \in B\}$

Definition 5.9 *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 $F([S]_g) := \{[b]_g \in [B]_g : F^{-1}([b]_g) \in [S]_g\}$. (In words, $cl_g(S, \mathcal{B})$ consists on taking the $\leq_g$-contraction of S in $\mathcal{A}$, map the set of equivalence classes $[S]_g$ into $\mathcal{B}$ by F and take the $\leq_g$-expansion in $\mathcal{B}$.)

Conversely, if S is a subset of B, then the $\leq_g$-closure of S in $\mathcal{A}$ is $cl_g(S, \mathcal{A}) := (F^{-1}([S]_g))^g$.

Given a vector of subsets $\overline{S} = (S_1, \dots, S_s)$ of A, we understand by $cl_g(\overline{S}, \mathcal{B})$ the vector of sets $(cl_g(S_1, \mathcal{B}), \dots, cl_g(S_s, \mathcal{B}))$.

The following example gives an infinite family of sublinear functions that define almost orders that will be useful later.

Example 5.10 *Fix $k \in \mathbb{N}$. Then $h_k(n) = 2r$, where $r \equiv_k n$, is a sublinear function.*

Take $k = 3$, for example, then $h_3(7) = 2$ and $h_3(8) = 4$. If $\mathcal{A}_7$ and $\mathcal{A}_8$ are τ-structures of size 7 and 8 respectively, appropriate for $\mathcal{SOLP}(\tau) + \leq_{h_3}$ (e.g. τ is empty or the interpretations of the relation symbols in these structures are consistent with $\leq_{h_3}$), then these structures look like:

$$\mathcal{A}_7 := \begin{array}{c}\bullet\\ \uparrow\downarrow\\ \bullet\end{array} \rightarrow \bullet \rightarrow \bullet \rightarrow \bullet \rightarrow \bullet \rightarrow \bullet$$

and

$$\mathcal{A}_8 := \begin{array}{c}\bullet\\ \uparrow\downarrow\\ \bullet\end{array} \rightarrow \begin{array}{c}\bullet\\ \uparrow\downarrow\\ \bullet\end{array} \rightarrow \bullet \rightarrow \bullet \rightarrow \bullet \rightarrow \bullet$$

Clearly $\mathcal{A}_7/_{\sim_{h_3}} \cong \mathcal{A}_8/_{\sim_{h_3}}$, and hence, they belong to the same $\leq_{h_3}$–cluster. $\square$

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 $\mathcal{A} \prec_{(k,t,1)} \mathcal{B}$ property.

Lemma 5.11 *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 an 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 $\overline{S} = (S_1, \ldots, S_t)$ of A, consistent with $\leq_g$,*
$$\mathcal{A} \models \phi(a_1, \ldots, a_s, \overline{S}) \text{ iff } \mathcal{B} \models \phi(b_1, \ldots, b_s, cl_g(\overline{S}, \mathcal{B}))\ ;$$
- *(ii) If $S \subseteq A$ then $|S| \leq |cl_g(S, \mathcal{B})| \leq |S| + 1$;*
- *(iii) If $S \subseteq B$ then $|S| - 1 \leq |cl_g(S, \mathcal{A})| \leq |S|$.*

Proof: (i): For an atomic formula $R(x_1, \ldots, x_s)$, where R is a relation from the vocabulary τ, the equivalence follows from the facts that $\mathcal{A}$ and $\mathcal{B}$ are in the same cluster $\mathbf{C}$ and the witnesses $a_1, \ldots, a_s$ in A for $R^{\mathcal{A}}$, and corresponding $b_1, \ldots, b_s$ in B for $R^{\mathcal{B}}$ are such that $a_i \equiv_{\mathbf{C}} b_i$, for $i = 1, \ldots, s$.

For a formula of the form $X(z)$, where X is a second order variable, assume $S \subseteq A$ is the witness for X in $\mathcal{A}$, consistent with $\leq_g$. Then

$$\begin{aligned}\mathcal{A} \models S(a) \quad &\iff \quad \mathcal{A}/_{\sim_g} \models [S]_g([a]_g) \iff \mathcal{B}/_{\sim_g} \models F([S]_g)(F([a]_g)) \\ &\iff \quad \mathcal{B}/_{\sim_g} \models F([S]_g)([b]_g) \iff \mathcal{B} \models cl_g(S, \mathcal{B})(b)\end{aligned}$$

The conjunction, disjunction, negation and formulas with (first order) quantifiers follows by logic.

(ii) (and (iii)): The key is that our almost order $\leq_g$ is made of 2-preorder followed by a linear order, and hence the only way to have $\mathcal{B}$ bigger than $\mathcal{A}$ and their quotient structures under $\sim_g$ isomorphic, is to put the extra element of $\mathcal{B}$ on top of one of the linearly ordered elements and form a new 2-preorder. Then the contraction of a set S in A and the expansion of its image under an isomorphism from $\mathcal{A}/_{\sim_g}$ to $\mathcal{B}/_{\sim_g}$ increases the size (in at most one more element) if the operation grabs the newly formed 2-preorder; otherwise stays of same cardinality as the original set S. The reverse process may decrease the size of sets in $\mathcal{B}$. □

Corollary 5.12 *Let g be a sublinear function and $\mathbf{C}$ an $\leq_g$-cluster of τ-models. For every $\mathcal{A}$ and $\mathcal{B}$ in $\mathbf{C}$, with $|\mathcal{A}| = m$ and $|\mathcal{B}| = m+1$, for every $k, t \in \mathbb{N}$, we have $\mathcal{A} \prec_{(k,t,1)} \mathcal{B}$.* □

Combining the previous corollary with Theorem 5.3 we get

Corollary 5.13 *Let $r_1, r_2, \ldots, r_k$ 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,*

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

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

We are now ready to proceed to the separability result for A–$\mathcal{SOMLP}$.

Theorem 5.14 *Let $r, r_1, r_2, \ldots, r_k$ be distinct non zero natural numbers, pairwise relatively prime. Then A–$\mathcal{SOMLP}[r_1, \ldots, r_k]$ is properly contained in A–$\mathcal{SOMLP}[r_1, \ldots, r_k, r]$.*

Proof: It is obvious that A–$\mathcal{SOMLP}[r_1, \ldots, r_k]$ is contained in A–$\mathcal{SOMLP}[r_1, \ldots, r_k, r]$. Furthermore, we saw (Example 4.5) that the query: "the size of the model is a multiple of r" is expressible in A–$\mathcal{SOMLP}[r]$ (and hence in A–$\mathcal{SOMLP}[r_1, \ldots, r_k, r]$). We will show that the above query is not expressible in A–$\mathcal{SOMLP}[r_1 \ldots, r_k]$.

Assume that there exists a sentence ϕ in $\mathcal{SOMLP}[r_1 \ldots, r_k]$ that defines the query "the size of the model is a multiple of r", 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 $g = h_t(\cdot)$ be the function defined in Example 5.10 with $t = r(\prod_{i=1}^k r_i)n$. Then

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

Let $\mathcal{A}$ be a structure, consistent with the almost order $\leq_{h_t}$, formed by b many equivalence classes of size 2 induced by $\leq_{h_t}$ followed by a linear order of size $m - 2b$. Let $\mathcal{B}$ be $\mathcal{A}$ with a new element with which we form an extra equivalence class; that is, $\mathcal{B}$ consists of $b + 1$ equivalence classes of size 2 and a linear order of size $m - 2b - 2$. There is a natural isomorphism between $\mathcal{A}/_{\sim_g}$ and $\mathcal{B}/_{\sim_g}$.

On the other hand, m satisfies the conditions of Corollary 5.13, 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. $\square$

Corollary 5.15 A–$\mathcal{SOMLP}[2] \subsetneq A$–$\mathcal{SOMLP}[2,3] \subsetneq A$–$\mathcal{SOMLP}[2,3,5] \subsetneq \ldots$ $\square$

This corollary is interesting in view of Remark 3.11: if one can prove the same result for formulas with second order variables of unbounded arity and for ordered structures, then we have a proof of **PSPACE** being different from **P**. In the next section we attempt to free this result from the arity restriction.

5.2 Limitations in expressive power for A–$\mathcal{SOLP}$

In this section we partially extend the separation result stated in Corollary 5.15 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**. What we are going to obtain in this section is that, in the almost ordered setting, A–$\mathcal{SOLP}Horn[2]$ differ in expressive power from A–$\mathcal{SOLP}[2,3]$. Our main tool is a reshaping of Theorem 5.3 in the context of $\mathcal{SOLP}Horn[2]$.

Theorem 5.16 *Let τ be a vocabulary, and k, t and r positive integers. Let g be a sublinear function and $\mathbf{C}$ an $\leq_g$-cluster of τ-models. Let $\mathcal{A}$ and $\mathcal{B}$ be two τ–structures* *in* $\mathbf{C}$*, with* $|A| = m$*,* $|B| = m+1$ *and* $m+1 > 2$*. If* $\mathcal{A} \prec_{(k,t,r)} \mathcal{B}$ *then, for all formula* $\varphi(X_1, \dots, X_t)$ *of* $\mathcal{SOLPH}orn(\tau)[2]$*, of first order quantifier rank* $\leq k$ *and at most* t *second order variables (free or not)* $X_1, \dots, X_t$ *of arity* r*, we have, for* $S_1, \dots, S_t \subseteq A^r$ *there are* $T_1, \dots, T_t \subseteq B^r$*, with* $|S_i| \leq |T_i|$*, such that*

$$\mathcal{A} \models \varphi(S_1, \dots, S_t) \text{ implies } \mathcal{B} \models \varphi(T_1, \dots, T_t)$$

Proof: The proof proceeds as the one for Theorem 5.3. The case of φ being first order is immediate. When proportional quantifiers are involved we state a similar inductive hypothesis, but now the arity of the second order variables is $r > 1$ and, furthermore, we are only considering proportional quantifiers of the form $(P(Y) \leq 1/2)$. Thus, now consider a formula of the form $(P(Y) \leq 1/2)\varphi(X_1, \dots, X_{t-1}, Y)$ with Y of arity $r > 1$. Let $\mathcal{A}$ and $\mathcal{B}$ be as in the hypothesis, with appropriate cardinality $m = |A|$ such that $m + 1 > 2$. (The way we have to choose m will be made precise towards the end.)

Fix a sequence of predicates $S_1, \dots, S_{t-1}$ of A, and assume $\mathcal{A} \models (P(Y) \leq 1/2)\varphi(S_1, \dots, S_{t-1}, Y)$. Then there is an $S \subseteq A^r$ such that $\mathcal{A} \models \varphi(S_1, \dots, S_{t-1}, S)$ and $|S| \leq (1/2)m^r$. (Observe that if m is odd then we really have $|S| \leq (1/2)(m^r - 1)$, but the arguments that follow bellow can easily be adapted for this case.)

By induction hypothesis $\mathcal{B} \models \varphi(T_1, \dots, T_{t-1}, T)$, for some predicates $T_1, \dots, T_{t-1}$ and T of B, such that $|S_i| \leq |T_i|$, and $|T| \leq (1/2)m^r + (m+1)^{r-1}$: the single extra element in $\mathcal{B}$ can be in a 2–preorder of two elements and thus can produce up to $(m+1)^{r-1}$ equivalent r-tuples for T.

Then the proportion of this set T with respect to $|\mathcal{B}|$ is

$$\begin{aligned} P(T) &\leq \frac{1}{2}\frac{(m^r + 2(m+1)^{r-1})}{(m+1)^r} \\ &= \frac{1}{2}\left[\left(1 - \frac{1}{m+1}\right)^r + \frac{2}{m+1}\right] \end{aligned}$$

Let $x := \dfrac{1}{m+1}$ and consider the function

$$f(x) = \frac{1}{2}[(1-x)^r + 2x]$$

whose derivative, $f'(x) = (1/2)[2 - r(1-x)^{r-1}]$, is negative as x tends to 0 (i.e. as m grows to infinity). Thus, as x tends to 0, $f(x)$ is bounded above by 1/2. Therefore, there is an $N > 0$, such that for $m > N$, $P(T) \leq 1/2$. Thus, choosing $m > N$, we obtain, $\mathcal{B} \models \varphi(T_1, \dots, T_{t-1}, T)$, for $T \subseteq B^r$ and $|T| \leq (1/2)(m+1)^r$; that is, $\mathcal{B} \models (P(Y) \leq 1/2)\varphi(T_1, \dots, T_{t-1}, Y)$. $\square$

With the above theorem we get a tool analogous to Theorem 5.8 for non definability in the logic $\mathcal{SOLPH}orn(\tau)[2]$, with respect to almost ordered structures.

Theorem 5.17 *Let* τ *be a vocabulary and* K *be a class of* τ*–structures. If for all positive integers* k*,* t *and* r*, there exists a sublinear function* g *and a pair of* τ*-structures* $\mathcal{A}$ *and* $\mathcal{B}$*, consistent with the almost order* $\leq_g$ *and such that:*

$$\mathcal{A} \in K \text{ and } \mathcal{B} \notin K,\ |\mathcal{B}| = |\mathcal{A}| + 1,$$

and Duplicator has a winning strategy in the $(\mathcal{A}, \mathcal{B}, k, t, r)$*–game. Then* K *is not definable in the logic* A*–*$\mathcal{SOLPH}orn(\tau)[2]$*.* $\square$

Using as benchmark query: "the size of the model is a multiple of 3", which is definable in A-$\mathcal{SOLP}[2,3]$, together with all the machinery developed in this section, we obtain

Theorem 5.18 *A–$\mathcal{SOLP}Horn[2] \subsetneq A$–$\mathcal{SOLP}[2,3]$.*

Proof: Let K be the class of sets of cardinality a multiple of 3 (the underlying vocabulary is empty). We have seen that K is definable in A–$\mathcal{SOLP}[2,3]$, and we shall see that for every k, t and r positive integers, there is no sentence φ in $\mathcal{SOLP}Horn[2]$ of quantifier rank at most k, with t second order variables of arity r, that defines K.

Fix k, t, and r positive integers. We will choose a sublinear function g and integer n, both depending on k, t and r, and a set $\mathcal{A}$ of cardinality 3^n, with elements almost ordered by $\leq_g$, a set $\mathcal{B}$ of cardinality 3^n+1 which is like $\mathcal{A}$ but with the extra element forming an extra 2–preorder of two elements, such that Duplicator has a winning strategy in the $(\mathcal{A}, \mathcal{B}, k, t, r)$–game. By our selection of cardinalities we will have that $\mathcal{A} \in K$ but $\mathcal{B} \notin K$. To choose n appropriately we use Ron Fagin's strategy of "*playing on similar s-types*" [5, §4].

For an integer s with $0 \leq s \leq k$, the s–type with respect to structure $\mathcal{A}$, denoted $type(\mathcal{A}, s)$, is defined by induction on s. The 0–type is the conjunction of all atomic or negated atomic formulas of up to k variables that are true in $\mathcal{A}$. If Γ is a set of s-types, with $0 \leq s < k$, an $(s+1)$–type is the formula

$$\bigwedge\{\exists x_{k-s}\varphi : \varphi \in \Gamma\} \wedge \bigwedge\{\forall x_{k-s}\neg\varphi : \varphi \notin \Gamma\}$$

The fundamental fact is that if $\hat{\mathcal{A}}$ and $\hat{\mathcal{B}}$ are extensions of structures $\mathcal{A}$ and $\mathcal{B}$ respectively, where t second order variables $X_1, \ldots, X_t$ of arity r have been realised, then

$$\hat{\mathcal{A}} \equiv_k \hat{\mathcal{B}} \iff \hat{\mathcal{A}} \models type(\hat{\mathcal{B}}, k)$$

Then, the Duplicator strategy, for the l-th move, is to choose his element on the same $(k-l)$–type of the conjunction of formulas in $\mathrm{FO}(\tau \cup \{X_1, \ldots, X_t\})$ of quantifier rank $\leq k$.

For k, t and r, there are finitely many k–types of formulas over the vocabulary $\tau \cup \{X_1, \ldots, X_t\}$ of quantifier rank $\leq k$. Let $\lambda := \lambda(k,t,r)$ be the number of these k–types. (This number can be very big: By [5, §9] it is at most $f(k+1)$, where $f(0) = O(k^2)$ and $f(k+1) = 2^{f(k)}$, a tower of $k+2$ exponents.)

Choose integers n, m and p such that $3^n > 3\lambda$, $p > \lambda$ and $p \equiv_m 3^n$. Let $g := h_m(\cdot)$ be the function defined in Example 5.10. Then $g(3^n) = h_m(3^n) = 2p$, and hence, $\mathcal{A}$ has $p > \lambda$ 2–preorders of 2 elements, and $3^n - 2p > \lambda$ singletons. This gives enough room for Duplicator to select t appropriate sets of r–tuples, $T_1, \ldots, T_t$, in $\mathcal{B}$, to match exactly Spoilers selection of sets $S_1, \ldots, S_t$ in $\mathcal{A}$ (both sequences of sets being consistent with $\leq_g$), such that all s–types, for $s \leq k$, can be realised, that is $\langle \mathcal{A}, S_1, \ldots, S_t\rangle \models type(\langle \mathcal{B}, T_1, \ldots, T_t\rangle, s)$. (Recall that $|\mathcal{A}| = 3^n$ and $|\mathcal{B}| = 3^n + 1$, and that $\mathcal{A}$ and $\mathcal{B}$ are basically the same structures, but for one point of $\mathcal{B}$ which is in an 2–preorder.) Thus, Duplicator has winning strategy in the k–rounds first order game. $\square$

6 Conclusions

In [5], Ron Fagin advices that in order to tackle difficult questions about the complexity of classes of problems, restrict these classes to a weaker yet more manageable framework, with the hope that this "will serve as a training ground for attacking the problem in their full generality" [5, p. 432]. Following this recommendation we have designed a counting quantifier that acts upon second order predicates and with it define a logic, $\mathcal{SOLP}$, within which we can frame questions like, is $\mathbf{P} = \mathbf{PSPACE}$?, into a chain of logics

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

Then to deal with the difficult, and desirable, separation between classes of ordered structures described by $\mathcal{SOLP}[2]$ and by $\mathcal{SOLP}[2,3]$, we have restricted the quantifiers to act only upon unary predicate variables and restricted the semantic interpretation of order to be an almost order, and developed tools and succeeded in separating classes of problems definable in this monadic fragment $\mathcal{SOMLP}$. From this restricted context we learnt how to (and succeeded in) going up to arbitrary arity, yet again over almost ordered structures, and separate $\mathcal{SOLP}Horn[2]$ from $\mathcal{SOLP}[2,3]$. The result is of interest since $\mathbf{P} = \mathcal{SOLP}Horn[2]$ over ordered structures.

Therefore, the next natural step is to refine our separation results to the context of ordered structures (not just almost ordered); that is, to enhance our results with respect to the semantical framework. We have taken a different route and instead of enhancing the semantics we propose in [2] to weaken the syntax by allowing and additive error in the proportions computed by the proportional quantifier and define *approximate formulas* that conform an approximate logic contained in $\mathcal{SOLP}$ with built-in order, that should be complementary of the semantic approximation based on almost orders. Then we link expressibility in fragments of $\mathcal{SOLP}$ over almost-ordered structures to a stronger form of expressibility with respect to approximate formulas for the corresponding fragments over ordered structures. What we hope for is to strengthen this bridge, so that proving inexpressibility results over fragments of $\mathcal{SOLP}$ with built-in order be equivalent to proving inexpressibility over the corresponding fragments with built-in almost order, where separation proofs are easier and, as shown in this paper, we know how to get.

Acknowledgments: The reduction in the proof of Proposition 2.4 was suggested by Marcos Kiwi to the first author in a brain storming session at a café in Santiago, Chile. Moritz Müller made significant comments to a first draft of this paper helping us improve the exposition. We also thank the anonymous referee for her/his helpful comments. A preliminary short version of this work have appeared in the proceedings of the *7th. LATIN 2006: Theoretical Informatics*, March 2006, Chile.

References

[1] Arratia, A., and Ortiz, C. (2004) Approximating the Expressive Power of Logics in Finite Models, *in: Latin American Theoretical INformatics, LATIN'04*, Lecture Notes in Comp. Sci. 2976 (Springer), 540–556.

[2] Arratia, A., and Ortiz, C. (2005) On a syntactic approximation to logics that capture complexity classes. *ECCC: Electronic Coloquium on Computational Complexity*, TR06-014.

[3] Ebbinghaus, H.D., and Flum, J., *Finite Model Theory* (Springer-Verlag, 1995).

[4] Etessami, K., and Immerman, N. (1995), Reachability and the power of local ordering, *Theoretical Comp. Sci.* **148**, 2, 261–279.

[5] Fagin, R. (1997), Comparing the power of games on graphs, *Mathematical Logic Quarterly* **43**, 431–455.

[6] Grädel, E. (1992), Capturing complexity classes by fragments of second order logic, *Theoretical Comp. Sci.* **101**, 35–57.

[7] Gurevich, Y. (1988), Logic and the challenge of computer science, *in*: "Current trends in theoretical computer science" (E. Börger, Ed.), Computer Science Press, pp. 1-57.

[8] Hella, L., Libkin, L., and Nurmonen, J., (1999) Notions of locality and their logical characterizations over finite models. *J. Symbolic Logic*, **64**, 1751-1773.

[9] Immerman, N. (1987), Languages that Capture Complexity Classes, *SIAM J. Comput.* **16**(4), 760-778.

[10] Immerman, N., *Descriptive Complexity* (Springer, 1998).

[11] Libkin, L., and Wong, L., (2002) Lower bounds for invariant queries in logics with counting. *Theoretical Comp. Sci.* **288**, 153-180.

[12] Plaisted, D. (1984), Complete problems in the first-order predicate calculus, *J. of Computer and System Sciences* **29** 8-35.

[13] Stewart, I. (1994), Logical description of monotone NP problems, *J. Logic Computat.* **4** (4), 337-357.