Lower Bounds for Invariant Queries in Logics with Counting

Leonid Libkin*
Bell Laboratories
600 Mountain Avenue
Murray Hill, NJ 07974, USA
Email: libkin@research.bell-labs.com

Limsoon Wong†
Kent Ridge Digital Labs
21 Heng Mui Keng Terrace
Singapore 119613
Email: limsoon@krdl.org.sg

Abstract

We study the expressive power of counting logics in the presence of auxiliary relations such as orders and preorders. The simplest such logic, first-order with counting, captures the complexity class TC^0 over ordered structures. We also consider first-order logic with arbitrary unary quantifiers, and infinitary extensions.

We start by giving a simple direct proof that first-order with counting, in the presence of preorders that are almost-everywhere linear orders, cannot express the transitive closure of a binary relation. The proof is based on locality of formulae. We then show that the technique cannot be extended to linear orders, and that the result does not say anything about the power of invariant queries in first-order with counting, in the presence of those preorders, vs. the class TC^0.

In the second part of the paper we then prove a separation result showing that for all the counting logics above, a linear order is more powerful than a preorder that is a linear order almost everywhere. In fact, we prove that the expressive power of invariant queries in the presence of those preorders can be characterized by a property normally associated with first-order definability over *unordered* structures. We do this by using locality techniques from finite-model theory; however, as some standard notions of locality fail in this setting, we have to modify them to prove the main result.

1 Introduction

The development of Descriptive Complexity suggests a very close connection between proving lower bounds in complexity theory and proving inexpressibility results in logic. The latter are of the form "a property P cannot be expressed in a logic $\mathcal{L}$ over a class of finite models." Developing tools for proving such expressivity bounds is one of the central problems in Finite-Model Theory. In this paper we show how tools based on locality of logics can be applied to the complexity class TC^0 and, more generally, how they allow us to derive new expressivity bounds of counting extensions of first-order logic in the presence of complex auxiliary relations.

The class TC^0 is an important complexity class: problems such as integer multiplication and division, and sorting belong to TC^0; this class has also been studied in connection with neural nets, cf. [30]. Despite serious efforts and a number of proved lower bounds (see [1] for a survey), it is still not known

*Part of this work was done while visiting INRIA and Kent Ridge Digital Labs.
†Part of this work was done while visiting Bell Labs.

if $\mathrm{TC}^0 \subsetneq \mathrm{NP}$, and the results of [31] show that traditional approaches to circuit lower bounds are unlikely to succeed in proving this separation.

A starting point for our study is a result by Barrington, Immerman and Straubing [2] stating that:

$$\mathrm{FO}(\mathbf{C}) + < \quad = \quad \text{uniform } \mathrm{TC}^0.$$

Here, as usual, TC^0 is the class of problems solvable by polynomial-size, constant-depth threshold circuits, and uniform means DLOGTIME-uniform, see [2] for more details. From now on, whenever we write TC^0, we mean the uniform class.

By $\mathrm{FO}(\mathbf{C})$ we mean the extension of first-order logic with counting quantifiers $\exists i$, where $\exists ix.\varphi(x)$ means that there are at least i elements x that satisfy φ. We shall give a full definition later, and at this point offer an example: $\exists i, j((j + j = i) \wedge \exists! ix.\varphi(x))$ (where $\exists! i$ is a shorthand for "exists exactly i") states that the number of x satisfying φ is even — this is known not to be expressible in first-order logic alone. By $\mathrm{FO}(\mathbf{C})+<$ we mean $\mathrm{FO}(\mathbf{C})$ in the presence of a built-in order relation.

The problem of separation of uniform TC^0 from classes such as DL, NL, P, etc, is thus reduced to proving that their complete problems are inexpressible in $\mathrm{FO}(\mathbf{C})+<$. However, it appears that the presence of an order relation is a major obstacle to proving such expressivity bounds for $\mathrm{FO}(\mathbf{C})$. Several partial results (for example, [8, 22]) show that there are problems complete for DL that cannot be defined by $\mathrm{FO}(\mathbf{C})$ in the presence of auxiliary relations, whose degrees are bounded by a fixed constant k. If we talk about directed graphs, by *degrees* we mean in- and out-degrees of nodes. For example, in the graph of a successor relation, every node has in- and out-degree either 0 or 1. In contrast, in a linear order on an n-element set, all n different (in- and out-) degrees from 0 to $n-1$ are realized. Thus, in order to move closer to proving expressivity bounds in the presence of an order relation, one has to at least be able to lift the results from constant degrees to those that depend on the size of the input.

A result in this direction was proved in [22], using a definition on *moderate degree* by Fagin, Stockemeyer and Vardi [9]. We say that a class $\mathcal{C}$ of graphs (more generally, relational structures) is of moderate degree, if $\mathit{degmax}_{\mathcal{C}}(n)$, the maximal in- or out-degree of an n-element graph from $\mathcal{C}$, is at most $\log^{o(1)} n$. That is, for some function $\delta(n)$ such that $\lim_{n\to\infty} \delta(n) = 0$, we have $\mathit{degmax}_{\mathcal{C}}(n) \le \log^{\delta(n)} n$. Then [22] proved that there is a DL-complete problem which is not definable in $\mathrm{FO}(\mathbf{C})$ in the presence of auxiliary relations of moderate degree.

In [9], auxiliary relations of moderate degree were shown to be of no help for expressing connectivity of graphs in monadic Σ^1_1. Starting from their result, Schwentick extended it to degrees $n^{o(1)}$ [32] and to a linear order [33]. So one may wonder if a similar program can be carried out for $\mathrm{FO}(\mathbf{C})$.

The intuition behind the introduction of a linear order is that it allows us to simulate encodings of structures on the tape of a Turing machine (or the order of inputs of a circuit). While for order-invariant properties it does not matter in which order elements appear on the tape (indeed, properties like connectivity of graphs to do not depend on how graphs are represented), they do appear in *some* order, and one must be able to use this order in logical formulae. Even though the particular ordering does not change the truth value of an order-invariant formula, the mere presence of an order gives many logics extra power. For example, while FO+LFP and FO+PFP capture PTIME and PSPACE over ordered structures [16, 34], they possess the 0-1 law over unordered structures [20], meaning that such a simple PTIME property as parity cannot be expressed. The lower bound of Cai, Fürer and Immerman [4] shows that there are PTIME properties of unordered structures not definable even in FO+LFP extended with counting quantifiers. A similar phenomenon is observed for other logics, e.g., FO and $\mathrm{FO}(\mathbf{C})$ [3, 29].

Our main goal is to study the impact of auxiliary relations, such as orderings, on the expressive power of logics with counting. The results we prove apply to a variety of logics, starting with FO and FO($\mathbf{C}$), and ending with a logic $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$ proposed in [23]. This logic subsumes FO($\mathbf{C}$) and all other known pure counting extensions of FO. (When we speak of counting extensions of FO, we mean extensions that only add a counting mechanism, as opposed to those – extensively studied in the literature, see [28] – that add both counting and fixpoint.)

We consider a class of relations which are extremely close to linear orderings – these are preorders, with equivalence classes of size at most 2, which coincide with linear orders almost everywhere (see Section 2 for precise definition, and Figure 1 for a picture). We first prove, by a simple direct argument, that there are DL and NL-complete problems not definable in all the counting logics above, in the presence of such relations. This immediately leads to a question whether the expressivity of, say, FO($\mathbf{C}$) in the presence of such relations is the same as that of FO($\mathbf{C}$)$+ <$. In the second part of the paper, we prove a more involved result showing that this is *not* the case. In particular, logics such FO($\mathbf{C}$) and $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$, in the presence of preorders which are almost everywhere linear orders, exhibit very tame behavior, normally associated with first-order definable properties. To prove the main result, we exploit the locality techniques in finite-model theory.

Organization In Section 2, we give formal definitions of various counting extensions of FO, notions of locality, and definability with auxiliary relations. We also give an example that shows how the presence of auxiliary relations affects expressiveness.

In Section 3, we give a direct proof that the transitive closure query is not expressible in FO($\mathbf{C}$) in the presence of almost-everywhere linear orders. We also explain that the technique of the proof does not straightforwardly generalize to proving separation results in the ordered case.

In Section 4, we address the question whether it is possible to use the almost-everywhere linear orders to prove separation results in the ordered case. We give a negative answer, by showing that, for all counting logics we consider here, adding order is strictly more expressive than adding preorder, however close to a linear order it might be. We state the result and some of its corollaries, and give the proof in Section 5, where we first describe notions of weak locality, and then combine them with bijective Ehrenfeucht-Fraïssé games.

Two extended abstracts with the results of this paper appeared in the Proceedings of 15th Symposium on Theoretical Aspects of Computer Science [26], and the Proceedings of the 14th IEEE Symposium on Logic in Computer Science [24].

2 Notations

Finite Structures and Logics All structures are assumed to be *finite*. A relational signature σ is a set of relation symbols $\{R_1, ..., R_l\}$, with associated arities $p_i > 0$. For directed graphs, the signature consists of one binary predicate. A σ-structure is $\mathcal{A} = \langle A, R_1^{\mathcal{A}}, \ldots, R_l^{\mathcal{A}}\rangle$, where A is a finite set, and $R_i^{\mathcal{A}} \subseteq A^{p_i}$ interprets R_i. The class of finite σ-structures is denoted by STRUCT[σ]. When there is no confusion, we write R_i in place of $R_i^{\mathcal{A}}$. Isomorphism is denoted by $\cong$. The carrier of a structure $\mathcal{A}$ is always denoted by A.

We abbreviate first-order logic by FO, and omit the standard definitions. *FO with counting*, denoted by FO($\mathbf{C}$), is a two-sorted logic, with second sort being interpreted as an initial segment of natural numbers. That is, a structure $\mathcal{A}$ is of the form

$$\langle \{v_1, \ldots, v_n\}, \{1, \ldots, n\}, <, \mathrm{BIT}, \underline{1}, \underline{\max}, R_1^{\mathcal{A}}, \ldots, R_l^{\mathcal{A}} \rangle.$$

Here the relations $R_i^{\mathcal{A}}$ are defined on the domain $\{v_1, \ldots, v_n\}$, while on the numerical domain $\{1, \ldots, n\}$ one has constants $\underline{1}, \underline{\max}$ (interpreted as 1 and n), the linear order $<$ and the BIT predicate available ($\mathrm{BIT}(i, j)$ iff the ith bit in the binary representation of j is one). This logic also has *counting quantifiers* $\exists ix.\varphi(x)$, meaning that that are at least i elements x that satisfy $\varphi(x)$; here i refers to the numerical domain and x to the domain $\{v_1, \ldots, v_n\}$. These quantifiers bind x but not i. Ternary predicates $+$ and $*$ are definable on the numerical domain [8], as is the quantifier $\exists! ix$ meaning the existence of exactly i elements satisfying a formula. For example, $\exists i \exists j \; [(j + j) = i \wedge \exists! ix.\varphi(x)]$ tests if the number of x satisfying φ is even; this property is not definable in FO alone. We separate first-sort variables from second-sort variables by semicolon: $\varphi(\vec{x}; \vec{\jmath})$.

There are several counting extensions of FO that are more powerful than $\mathrm{FO}(\mathbf{C})$; among them $\mathrm{FO}(\mathbf{Q}_u)$, which is FO extended with *all* unary quantifiers. We refer the reader to [13, 19] for the definition of $\mathrm{FO}(\mathbf{Q}_u)$ and its properties. Here, we mostly work with an even more powerful logic, defined below.

We denote the infinitary logic by $\mathcal{L}_{\infty\omega}$; it extends FO by allowing infinite conjunctions $\bigwedge$ and disjunctions $\bigvee$. Then $\mathcal{L}_{\infty\omega}(\mathbf{C})$ is a two-sorted logic, that extends infinitary logic $\mathcal{L}_{\infty\omega}$. Its structures are of the form $(\mathcal{A}, \mathbb{N})$, where $\mathcal{A}$ is a finite relational structure, and $\mathbb{N}$ is a copy of natural numbers. Assume that every constant $n \in \mathbb{N}$ is a second-sort term. To $\mathcal{L}_{\infty\omega}$, add counting quantifiers $\exists ix$ for every $i \in \mathbb{N}$, and *counting terms:* If φ is a formula and $\vec{x}$ is a tuple of free first-sort variables in φ, then $\#\vec{x}.\varphi$ is a term of the second sort, and its free variables are those in φ except $\vec{x}$. Its interpretation is the number of tuples $\vec{a}$ over the finite first-sort universe that satisfy φ. That is, given a structure $\mathcal{A}$, a formula $\varphi(\vec{x}, \vec{y}; \vec{\jmath})$, $\vec{b} \subseteq A$, and $\vec{\jmath}_0 \subset \mathbb{N}$, the value of the term $\#\vec{x}.\varphi(\vec{x}, \vec{b}; \vec{\jmath}_0)$ is the cardinality of the (finite) set $\{\vec{a} \subseteq A \mid \mathcal{A} \models \varphi(\vec{a}, \vec{b}; \vec{\jmath}_0)\}$. For example, the interpretation of $\#x.E(x, y)$ is the in-degree of node y in a graph with the edge-relation E.

As this logic is too powerful (it expresses every property of finite structures), we restrict it by means of the *rank* of a formulae and terms, denoted by rk. It is defined as quantifier rank (that is, it is 0 for atomic formulae, $\mathsf{rk}(\bigvee_i \varphi_i) = \max_i \mathsf{rk}(\varphi_i), \mathsf{rk}(\neg\varphi) = \mathsf{rk}(\varphi), \mathsf{rk}(\exists x\varphi) = \mathsf{rk}(\exists ix\varphi) = \mathsf{rk}(\varphi) + 1$) but it does not take into account quantification over $\mathbb{N}$: $\mathsf{rk}(\exists i\varphi) = \mathsf{rk}(\varphi)$. Furthermore, $\mathsf{rk}(\#\vec{x}.\psi) = \mathsf{rk}(\psi) + |\vec{x}|$.

Definition 1 (see [23]) *The logic $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$ is defined to be the restriction of $\mathcal{L}_{\infty\omega}(\mathbf{C})$ to terms and formulae of finite rank.*

It is known [23] that $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$ formulae are closed under Boolean connectives and all quantification, and that every predicate on $\mathbb{N} \times \ldots \times \mathbb{N}$ is definable by a $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$ formula of rank 0. Thus, we assume that $+, *, -, \leq$, and in fact *every* predicate on natural numbers is available. Known counting expansions of FO are contained in $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$. That is, for every FO, $\mathrm{FO}(\mathbf{C})$, or $\mathrm{FO}(Q_{\mathbf{u}})$ formula, there exists an equivalent $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$ formula of the same rank. A counting logic of [3] can also be embedded into $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$.

Definability with auxiliary relations An *m-ary query* on σ-structures, Q, is a mapping that associates to each $\mathcal{A} \in \mathrm{STRUCT}[\sigma]$ a structure $\langle A, S \rangle$, where $S \subseteq A^m$. We write $\vec{a} \in Q(\mathcal{A})$ if $\vec{a} \in S$, where $\langle A, S \rangle = Q(\mathcal{A})$. A query Q is definable in a logic $\mathcal{L}$ if there exists an $\mathcal{L}$ formula $\varphi(x_1, \ldots, x_m)$ such that $Q(\mathcal{A}) = \varphi[\mathcal{A}] \stackrel{\text{def}}{=} \langle A, \{\vec{a} \mid \mathcal{A} \models \varphi(\vec{a})\} \rangle$.

Let σ' be a relational signature disjoint from σ. If $\mathcal{A}$ is a σ-structure on a universe A, and $\mathcal{A}'$ is a σ'-structure on A, we use the notation $(\mathcal{A}, \mathcal{A}')$ for the $\sigma \cup \sigma'$-structure on A which inherits the interpretation of σ relational symbols from $\mathcal{A}$, and the interpretation of σ' symbols from $\mathcal{A}'$.

Let $\mathcal{C}$ be a class of σ'-structures, with σ and σ' being disjoint. Let $\mathcal{A} \in \mathrm{STRUCT}[\sigma]$. A formula $\varphi(\vec{x})$ in the language of $\sigma \cup \sigma'$ is called *$\mathcal{C}$-invariant on $\mathcal{A}$* if for any two $\mathcal{C}$ structures $\mathcal{A}'$ and $\mathcal{A}''$ on A we have $\varphi[(\mathcal{A}, \mathcal{A}')] = \varphi[(\mathcal{A}, \mathcal{A}'')]$. Associated with such a formula is the following m-ary query (where $m = |\vec{x}|$):

$$Q^w_\varphi(\mathcal{A}) = \begin{cases} \varphi[(\mathcal{A}, \mathcal{A}')], & \varphi \text{ is } \mathcal{C}\text{-invariant on } \mathcal{A} \\ \emptyset, & \text{otherwise.} \end{cases}$$

where $\mathcal{A}'$ is any structure from $\mathcal{C}$ on A. We use the notation $(\mathcal{L} + \mathcal{C})_w$ to denote all queries defined in such a way when φ ranges over formulae of $\mathcal{L}$.

A formula φ is *$\mathcal{C}$-invariant* if it is $\mathcal{C}$-invariant on every structure. With such a φ, we associate a query Q_φ given by $Q_\varphi(\mathcal{A}) = \varphi[(\mathcal{A}, \mathcal{A}')]$ where $\mathcal{A}'$ is a structure from $\mathcal{C}$ on A. The class of all such queries is denoted by $\mathcal{L} + \mathcal{C}$. Clearly,

$$\mathcal{L} + \mathcal{C} \quad \subseteq \quad (\mathcal{L} + \mathcal{C})_w$$

We thus shall aim to establish expressivity bounds for $(\mathcal{L} + \mathcal{C})_w$.

When $\mathcal{C}$ is the class of order relations, we shall write $<$ instead of $\mathcal{C}$. The capture results for complexity classes deal with the classes of queries of the form $\mathcal{L}+ <$; for example, uniform TC^0 equals $\mathrm{FO}(\mathbf{C})+ <$ [2]. While queries in $\mathcal{L}+ <$ are independent of a particular order relation used, the mere presence of such a relation can have an impact on the expressivity of a logic.

We give an example for $\mathrm{FO}(\mathbf{C})$. Assume that σ has one binary and unary relation, i.e. its structures are graphs with a selected subset of nodes. Let Q_0 be the following Boolean query [3]: given such a structure $\langle A, E, X \rangle$, where $A \neq \emptyset$, $E \subseteq A^2$ and $X \subseteq A$, return true iff E is an equivalence relation, and the number of distinct sizes of E-classes equals $|X|$. It is known that Q_0 is not expressible in $\mathrm{FO}(\mathbf{C})$ [3]. However, it *is* expressible in $\mathrm{FO}(\mathbf{C})+ <$. Indeed, the equivalence relation $x\theta y$ iff the E-equivalence classes of x and y have the same cardinality is definable in $\mathrm{FO}(\mathbf{C})$. Thus, in $\mathrm{FO}(\mathbf{C})$ one defines the set of smallest (wrt $<$) elements of each such class, and then compares, in $\mathrm{FO}(\mathbf{C})$, the size of this set to X. The two are the same iff the value of Q_0 is true. Note that *any* linear order suffices to express this query.

Thus, $\mathrm{FO}(\mathbf{C}) \subset \mathrm{FO}(\mathbf{C})+ <$. Since the latter captures uniform TC^0, this means that there are problems in TC^0 not definable in $\mathrm{FO}(\mathbf{C})$ over unordered structures. It is also known that $\mathrm{FO} \subset \mathrm{FO}+ <$. We shall see later that this continues to be true for other counting logics.

Bounded number of degrees property (BNDP) If $\mathcal{A} \in \mathrm{STRUCT}[\sigma]$, and R_i is of arity p_i, then $\mathit{degree}_j(R_i^{\mathcal{A}}, a)$ for $1 \leq j \leq p_i$ is the number of tuples $\vec{a}$ in $R_i^{\mathcal{A}}$ having a in the jth position. In the case of directed graphs, this gives us the usual notions of in- and out-degree. By $\mathit{deg\_set}(\mathcal{A})$ we mean the set of all degrees realized in $\mathcal{A}$, and $\mathit{deg\_count}(\mathcal{A})$ stands for the cardinality of $\mathit{deg\_set}(\mathcal{A})$. We use the notation $\mathrm{STRUCT}_k[\sigma]$ for $\{\mathcal{A} \in \mathrm{STRUCT}[\sigma] \mid \mathit{deg\_set}(\mathcal{A}) \subseteq \{0, 1, \ldots, k\}\}$.

Definition 2 (see [25, 5, 22]) *An m-ary query Q, $m \geq 1$, is said to have the* bounded number of degrees property[1] *, or* BNDP*, if there exists a function $f_Q : \mathbb{N} \to \mathbb{N}$ such that $\mathit{deg\_count}(Q(\mathcal{A})) \leq$*

[1]This property was formerly known as the bounded degree property, or the BDP, see [5, 14, 23, 25, 26, etc]. However, many found the name confusing, as the property refers to the *number* of degrees in the output being bounded, rather than the degrees themselves. Following a suggestion by Neil Immerman, we decided to change the name from BDP to BNDP.

$f_Q(k)$ for every $\mathcal{A} \in \mathrm{STRUCT}_k[\sigma]$. □

The BNDP is very easy to use for proving expressivity bounds [25]. For example, it is very easy to verify that (deterministic) transitive closure violates the BNDP.

Locality All existing proofs of the BNDP establish first that a logic is *local.* We now define this concept. Given a structure $\mathcal{A}$, its *Gaifman graph* [7, 10, 9] $\mathcal{G}(\mathcal{A})$ is defined as $\langle A, E\rangle$ where (a,b) is in E iff there is a tuple $\vec{c} \in R_i^{\mathcal{A}}$ for some i such that both a and b are in $\vec{c}$. The distance $d(a,b)$ is defined as the length of the shortest path from a to b in $\mathcal{G}(\mathcal{A})$; we assume $d(a,a) = 0$. If $\vec{a} = (a_1,\ldots,a_n)$ and $\vec{b} = (b_1,\ldots,b_m)$, then $d(\vec{a},\vec{b}) = \min_{ij} d(a_i,b_j)$. Given $\vec{a}$ over A, its *r-sphere* $S_r^{\mathcal{A}}(\vec{a})$ is $\{b \in A \mid d(\vec{a},b) \leq r\}$. Its *$r$-neighborhood* $N_r^{\mathcal{A}}(\vec{a})$ is defined as a structure $N_r^{\mathcal{A}}(\vec{a})$

$$\langle S_r^{\mathcal{A}}(\vec{a}), R_1^{\mathcal{A}} \cap S_r^{\mathcal{A}}(\vec{a})^{p_1}, \ldots, R_k^{\mathcal{A}} \cap S_r^{\mathcal{A}}(\vec{a})^{p_l}, a_1, \ldots, a_n\rangle$$

in the signature that extends σ with n constant symbols. That is, the carrier of $N_r^{\mathcal{A}}(\vec{a})$ is $S_r^{\mathcal{A}}(\vec{a})$, the interpretation of the σ-relations is inherited from $\mathcal{A}$, and the n extra constants are the elements of $\vec{a}$. If $\mathcal{A}$ is understood, we write $S_r(\vec{a})$ and $N_r(\vec{a})$.

If $\mathcal{A},\mathcal{B} \in \mathrm{STRUCT}[\sigma]$, and there is an isomorphism $N_r^{\mathcal{A}}(\vec{a}) \to N_r^{\mathcal{B}}(\vec{b})$ (that sends $\vec{a}$ to $\vec{b}$), we write $\vec{a} \approx_r^{\mathcal{A},\mathcal{B}} \vec{b}$. If $\mathcal{A} = \mathcal{B}$, we write $\vec{a} \approx_r^{\mathcal{A}} \vec{b}$.

Given tuples $\vec{a} = (a_1,\ldots,a_n)$ and $\vec{b} = (b_1,\ldots,b_m)$, and an element c, we write $\vec{a}\vec{b}$ for the tuple $(a_1,\ldots,a_n,b_1,\ldots,b_m)$, and $\vec{a}c$ for $(a_1,\ldots,a_n,c)$.

Definition 3 (cf. [22]) *An m-ary query Q is called* local *if there exists a number $r \geq 0$ such that, for any structure $\mathcal{A}$ and any $\vec{a},\vec{b} \in A^m$*

$$\vec{a} \approx_r^{\mathcal{A}} \vec{b} \quad \textit{implies} \quad \vec{a} \in Q(\mathcal{A}) \;\; \textit{iff} \;\; \vec{b} \in Q(\mathcal{A}).$$

The minimum such r is called the locality rank *of Q, and is denoted by $\mathsf{lr}(Q)$.* □

It follows from Gaifman's theorem [10] that every FO-definable query is local; moreover, if Q is definable by a formula $\varphi(\vec{x})$, then $\mathsf{lr}(Q) \leq (7^{\mathsf{qr}(\varphi)} - 1)/2$. It was shown in [22, 23] that every $\mathrm{FO}(Q_{\mathbf{u}})$, $\mathrm{FO}(\mathbf{C})$, and $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$-definable query is local; furthermore, $\mathsf{lr}(Q) \leq 2^{\mathsf{rk}(\varphi)}$ [23].

Fact 1 (see [5]) *Every local query has the bounded number of degrees property.* □

Thus, without auxiliary relations, queries such as transitive closure cannot be expressed in $\mathrm{FO}(\mathbf{C})$ and even in $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$.

Proviso: When we deal with queries in $\mathcal{L} + \mathcal{C}$ and $(\mathcal{L} + \mathcal{C})_w$, which are defined on structures $(\mathcal{A},\mathcal{A}')$, $\mathcal{A}' \in \mathcal{C}$, all locality concepts (neighborhoods, degrees, etc) refer only to the σ-structure $\mathcal{A}$, and not to the auxiliary structure $\mathcal{A}'$ from $\mathcal{C}$.

Almost-linear-orders We next define the class of relations that we view as "almost linear orders." First, let $\precsim_k$ stand for the class preorders R (reflexive transitive relations) in which every equivalence

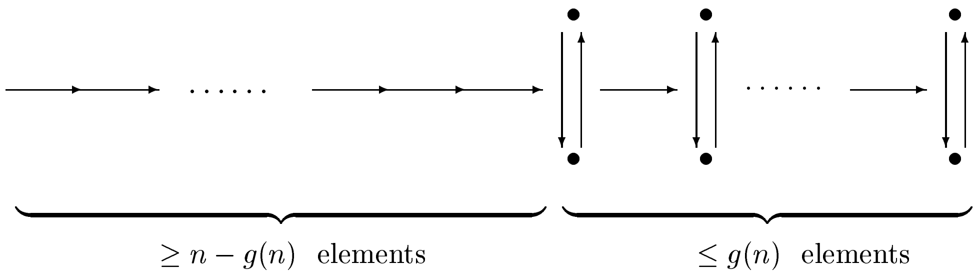

Figure 1: A relation from $<^{-}_{g}$

class of $R \cap R^{-1}$ (that is, $x \equiv y$ iff xRy and yRx) has size at most k. Note that $\precsim_1$ is the class of linear orders.

Let $g : \mathbb{N} \to \mathbb{R}$ be a nondecreasing function[2]. Define $<^{-}_{g}$ as the class of binary relations (A, R) such that there exists a partition $A = B \cup C$ with the following properties:

1. The cardinality of B is at least $n - g(n)$.
2. R restricted to B is a linear order.
3. R restricted to C is a relation from $\precsim_2$, that is, a preorder where every equivalence class has at most two elements.
4. For any $b \in B$ and $c \in C$, $(b, c) \in R$, and $(c, b) \notin R$.

See Figure 1 for a preorder from $<^{-}_{g}$. Actually, we show the associated successor relation in the Figure. A relation from $<^{-}_{g}$ is really the transitive closure of the one shown in Figure 1. Intuitively, if g is very small (e.g., $\log\log\ldots\log n$), then this can be viewed as the least possible "damage" that can be done to a linear ordering: we make a small subset at the end into a preorder, with its classes having no more than 2 elements.

3 Expressivity bounds for $\mathrm{FO}(\mathbf{C})$ and $\mathrm{FO}(Q_{\mathbf{u}})$ in the presence of relations of large degree

We start by giving a general technique for proving expressivity bounds for local logics. Then we apply it to $\mathrm{FO}(\mathbf{C})$ to prove our main result that DL-complete problems (in particular, deterministic transitive closure) cannot be expressed in it in the presence of relations that are very close to linear orderings. In particular, it will follow that $\mathrm{DL} \not\subseteq \mathrm{FO}(\mathbf{C}) + \precsim_k$ for any $k > 1$.

Proving expressivity bounds in local logics Let Q be a query that takes structures from $\mathrm{STRUCT}[\sigma]$ as inputs and returns m-ary relations (e.g, transitive closure takes graphs from $\mathrm{STRUCT}[\sigma_{\mathrm{gr}}]$ as inputs and returns graphs). Let $\mathcal{R}$ be a class of relations, and $\mathcal{L}$ a logic. Suppose we want to prove that $Q \notin (\mathcal{L} + \mathcal{R})_w$. For that purpose, we introduce two conditions.

[2]One can deal with functions $g : \mathbb{N} \to \mathbb{N}$ as well; however, as in many examples we use $\log_2$, we prefer to have $\mathbb{R}$ as the range.

Def$_{\mathcal{L}[\sigma]}[\mathcal{R},\mathcal{C}]$ Assume $\mathcal{C} \subseteq \mathrm{STRUCT}[\sigma]$. Then there exists a number n and an $\mathcal{L}$ formula φ in the vocabulary σ such that $\varphi[\mathcal{A}] \in \mathcal{R}$ for every $\mathcal{A} \in \mathcal{C}$ with $|A| > n$.

That is, relations from $\mathcal{R}$ are definable by σ-formulae of $\mathcal{L}$ on large enough structures from $\mathcal{C}$.

Sep$_{\mathcal{L}[\sigma]}[Q,\mathcal{C}]$ For any two numbers $r, n > 0$, there exists $\mathcal{A} \in \mathcal{C}$ with $|A| > n$ and two m-ary vectors $\vec{a}$, $\vec{b}$ of elements of A such that $\vec{a} \approx_r^{\mathcal{A}} \vec{b}$, $\vec{a} \in Q(\mathcal{A})$ and $\vec{b} \notin Q(\mathcal{A})$.

That is, Q separates similarly looking (in a local neighborhood) tuples on arbitrarily large structures from $\mathcal{C}$.

Theorem 1 *Assume that $\mathcal{L}$ is* FO*, or* FO(**C**)*, or* FO($Q_{\mathbf{u}}$)*, or* $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$*. Suppose for a given query Q on σ-structures, one can find $\mathcal{C} \subseteq \mathrm{STRUCT}[\sigma]$ such that both* **Def**$_{\mathcal{L}[\sigma]}[\mathcal{R},\mathcal{C}]$ *and* **Sep**$_{\mathcal{L}[\sigma]}[Q,\mathcal{C}]$ *hold. Then $Q \notin (\mathcal{L} + \mathcal{R})_w$.*

Proof: Assume that Q is in $(\mathcal{L} + \mathcal{R})_w$. Since **Sep**$_{\mathcal{L}[\sigma]}[Q,\mathcal{C}]$ holds, Q returns nonempty results for arbitrarily large structures. Thus, we assume that it is definable by a formula ψ in the vocabulary that includes σ and a symbol R for the relation from $\mathcal{R}$. Let ψ' be obtained from ψ by replacing each occurrence of $R(\cdots)$ by $\varphi(\cdots)$, where φ is given by **Def**$_{\mathcal{L}[\sigma]}[\mathcal{R},\mathcal{C}]$. Note that ψ' is a $\mathcal{L}$-formula in the vocabulary σ. By [22, 23], ψ' is local. Let $r = \mathsf{lr}(\psi')$. For an arbitrary $N > n$, we find a structure $\mathcal{A}$ of cardinality at least N and $\vec{a} \approx_r^{\mathcal{A}} \vec{b}$ such that $\vec{a} \in Q(\mathcal{A})$ and $\vec{b} \notin Q(\mathcal{A})$. Since $Q(\mathcal{A}) \neq \emptyset$, we know that ψ is invariant on $\mathcal{A}$; thus, $Q(\mathcal{A}) = \psi[(\mathcal{A}, R)]$ where R is any interpretation of a relation from $\mathcal{R}$ on A. From the invariance we now then obtain $\psi'[\mathcal{A}] = Q(\mathcal{A})$. Thus, for $\vec{a}, \vec{b}$, we have $\mathcal{A} \models \neg(\psi'(\vec{a}) \leftrightarrow \psi'(\vec{b}))$, which contradicts the locality of ψ'. This proves the theorem. $\Box$

Note that this theorem can be straightforwardly extended to the case of several built-in relations of possibly different arities, by considering $\vec{\mathcal{R}}$ instead of $\mathcal{R}$, where $\vec{\mathcal{R}}$ is a tuple of classes of auxiliary relations. Then **Def**$_{\mathcal{L}[\sigma]}[\vec{\mathcal{R}},\mathcal{C}]$ says that relations from each component of $\vec{\mathcal{R}}$ can be defined by a σ-formula of $\mathcal{L}$ on sufficiently large structures from $\mathcal{C}$.

Lower bounds for transitive closure Recall [17] that deterministic transitive closure of a graph is obtained by closing its deterministic paths, that is, if $G = \langle V, E\rangle$ is a directed graph, then $dtc(G) = \langle V, E'\rangle$ where $(a, b) \in E'$ iff either $(a, b) \in E$ or there exists a path $(a, a_1), (a_1, a_2), \ldots, (a_{n-1}, a_n), (a_n, b) \in E$ such that a and each a_i, $i = 1, \ldots, n$ have outdegree 1. That is, the edge (a, a_1) is the only outgoing edge from a, etc.

We shall use tc to denote the transitive closure of a graph. According to [17], FO + dtc+ $<$ captures DL and FO + tc+ $<$ captures NL. Note that $<$ can be replaced by a successor relation, since its (deterministic) transitive closure is a linear order.

Theorem 2 *Let $g : \mathbb{N} \to \mathbb{R}$ be a nondcreasing function that is not bounded by a constant. Then (deterministic) transitive closure is not in $(\mathcal{L}+ <^{-}_{g})_w$, where $\mathcal{L}$ is* FO(**C**)*, or* FO($Q_{\mathbf{u}}$)*, or* $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$*. In particular,* DL $\not\subseteq$ FO(**C**)+ $<^{-}_{g}$. $\Box$

This can be compared with the results of [4] where it was shown that first-order with fixpoint and counting fails to express some polynomial-time problems even in the presence of relations from $\precsim_4$ (of course first-order with fixpoint captures polynomial time in the presence of an order relation, cf. [7]).

In view of Theorem 1, to establish Theorem 2, it is enough to prove the following:

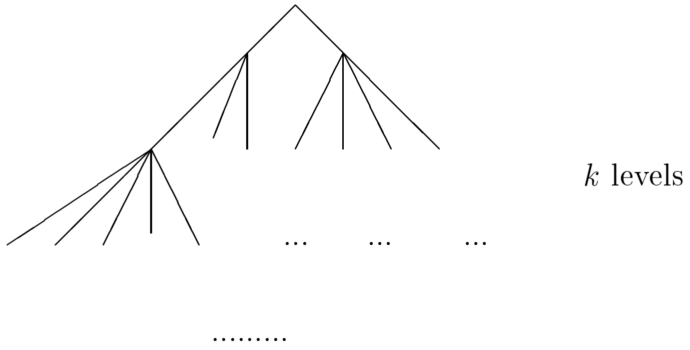

Figure 2: Canonical k-bushy tree

Proposition 1 *Let q be (deterministic) transitive closure, and $\mathcal{L}$ be* $\mathrm{FO}(\mathbf{C})$ *or* $\mathrm{FO}(Q_{\mathbf{u}})$. *Assume that $g : \mathbb{N} \to \mathbb{R}$ is a nondcreasing function that is not bounded by a constant. Then there exists a class $\mathcal{C}$ of graphs such that both* $\mathbf{Def}_{\mathcal{L}[\sigma_{\mathrm{gr}}]}(<^-_g, \mathcal{C})$ *and* $\mathbf{Sep}_{\mathcal{L}[\sigma_{\mathrm{gr}}]}(q, \mathcal{C})$ *hold.*

Bushy trees In what follows, trees are directed graphs with edges oriented from the root to the leaves.

A tree is called *bushy* if, for any two non-leaf nodes $x \neq y$, $\textit{out-deg}(x) \neq \textit{out-deg}(y)$. A *$k$-bushy* tree is a bushy tree in which every path from the root to a leaf has the same length k. A *canonical* k-bushy tree is obtained as follows. We start with the root of outdegree 2. Its first child has 3 children, the second child has 4 children. This completes level 2, and we now have 7 elements at level 3. They will have 5, 6, 7, 8, 9, 10 and 11 children, respectively. This gives us 56 nodes at level 4, which will have 12(=11+1), 13, ..., 67(=11+56) children, resp. We continue until we fully filled all k levels. See the picture in Figure 2. We use B_k to denote the canonical k-bushy tree.

Proof of Proposition 1 We start by defining a family of graphs $G^0_{d,k}$, $d, k \in \mathbb{N}_+$, $d > k+1$. Let s_k be the total number of nodes in the canonical k-bushy tree. The root of $G^0_{d,k}$ has $s_k + 1$ children. Two of them are roots of two copies of a canonical k-bushy tree, denoted here by B^1_k and B^2_k. To other $s_k - 1$ nodes at the second level, we give $s_k + 2, s_k + 3, \ldots, s_k + (s_k - 1) = 2s_k - 1$ children respectively. Now, to those nodes at the second level that do not belong to the two canonical k-bushy trees, we give $2s_k, 2s_k + 1, \ldots$ children, as before, increasing the number by one. We continue this process until we fully fill the $k + 1$st level.

Now that the $k + 1$st level is filled (i.e. we have a graph with all paths from root to leaves being of length $k + 1$), we look at the node at the level k with most children, say M of them, and start giving nodes at the $k + 1$st level $M + 1, M + 2, M + 3, \ldots$ children. We stop the process when we completely fill the dth level.

This is the graph $G^0_{d,k}$. Note that every two non-leaf nodes $x \neq y$ have different outdegrees, unless one of them is in B^1_k and the other is in B^2_k. We now define $G_{d,k}$ by adding an arbitrary linear ordering on the leaves. That is, if Z is the set of leaves of $G^0_{d,k}$, and L_Z is a linear order on Z, then the set of edges of $G_{d,k}$ is that of $G^0_{d,k}$ plus L_Z. When we speak of "leaf nodes" of $G_{d,k}$, we actually mean the leaf nodes of $G^0_{d,k}$.

Let B_1° and B_2° be the sets of non-leaf nodes in B_k^1 and B_k^2. Then, for any two distinct nodes $x, y \notin B_1^\circ \cup B_2^\circ$, it is the case that $(\textit{in-deg}(x), \textit{out-deg}(x)) \neq (\textit{in-deg}(y), \textit{out-deg}(y))$. Indeed, outdegrees are different for non-leaf nodes of $G_{d,k}^0$ (except for those in $B_1^\circ \cup B_2^\circ$), and all in-degrees are different for the leaf nodes. The only leaf node with in-degree one is the smallest in the linear order, and thus has the outdegree which exceeds that of all the internal nodes.

Define two binary relations on the set of nodes: $x \prec_0 y$ iff $\textit{in-deg}(x) < \textit{in-deg}(y)$ or $\textit{in-deg}(x) = \textit{in-deg}(y)$ and $\textit{out-deg}(x) < \textit{out-deg}(y)$. Let B° be $B_1^\circ \cup B_2^\circ$. Define:

$$x \prec y \;\text{ iff }\; \begin{cases} x \notin B^\circ, y \in B^\circ, & \text{or} \\ x, y \in B^\circ \text{ and } x \prec_0 y, & \text{or} \\ x, y \notin B^\circ \text{ and } x \prec_0 y, & \text{or} \end{cases}$$

From this description, it is clear that $\prec$ is definable in $\mathrm{FO}(\mathbf{C})$ (and thus in $\mathrm{FO}(Q_{\mathbf{u}})$ and $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$) as these logics can define the set B° (as the set of nodes for which there exists another node with the same in- and out-degree). Let $\alpha_\prec$ denote a formula defining $\prec$.

Now, given k, let d_k be the smallest number $d > k+1$ such that $2s_k < g(n)$ for all $n \geq N_{d,k}$, where $N_{d,k}$ is the total number of nodes in $G_{d,k}$. Since for every fixed k, $N_{d,k}$ grows with d, and g is nondecreasing, d_k is well-defined and indeed depends only on k. Let $\mathcal{C}_g = \{G_{d,k} \mid d, k \in \mathbb{N}_+, d > d_k\}$. From the construction above, it follows that, for a nondecreasing g, $\alpha_\prec$ defines an element of $<\bar{}_g$ on every $G_{d,k} \in \mathcal{C}_g$. Thus, $\mathbf{Def}_{\mathcal{L}[\sigma_{\mathrm{gr}}]}[<\bar{}_g, \mathcal{C}_g]$ holds.

We now show that $\mathbf{Sep}_{\mathcal{L}[\sigma_{\mathrm{gr}}]}[tc, \mathcal{C}_g]$ holds. Fix two numbers $r, n > 0$. We must show that there exist a graph $G_{d,k} \in \mathcal{C}_g$ such that, for two pairs of nodes (a, b) and (a', b') with $(a, b) \approx_r (a', b')$, there is a path from a to b, but there is no path from a' to b'.

Let $k = 2r+2$, and let $d > 4r+6$ be such that $G_{d,k} \in \mathcal{C}_g$. Let a be a node at the middle ($(r+1)$th) level of B_k^1, and a' a node in B_k^2 with the same out-degree as a. Then $a \approx_r a'$. We now consider a path from a to a leaf, say l, and choose a node b on this path such that $d(b, a) > 2r+1$ and $d(b, l) > r$; this is possible since $d > 4r+6$. It is clear that $d(b, a') > 2r+1$; hence we have $(a, b) \approx_r (a', b)$. Furthermore, $(a, b) \in tc(G_{d,k})$, but $(a', b) \notin tc(G_{d,k})$, since $G_{d,k}^0$ is a tree. This proves $\mathbf{Sep}_{\mathcal{L}[\sigma_{\mathrm{gr}}]}[tc, \mathcal{C}_g]$.

To complete the proof for deterministic transitive closure, we just reverse all the edges of $G_{d,k}$. Since all paths not involving leaves now become deterministic, the above proof works for the deterministic case. □

Corollary 1 *(Deterministic) transitive closure is not definable in* $\mathrm{FO}(\mathbf{C})$ *or* $\mathrm{FO}(Q_{\mathbf{u}})$ *in the presence of relations from* $\precsim_k$ *(preorders of width at most k) for any* $k > 1$. *In particular,* $\mathrm{DL} \not\subseteq \mathrm{FO}(\mathbf{C}) + \precsim_k$. □

Limitations of the technique To summarize what has been achieved so far, we know that $\mathrm{FO}(\mathbf{C})+ <= \mathrm{TC}^0$, and the above results show that for any $k > 1$, $\mathrm{DL} \not\subseteq \mathrm{FO}(\mathbf{C}) + \precsim_k$. Furthermore, $\mathrm{DL} \not\subseteq \mathrm{FO}(\mathbf{C}) + <\bar{}_g$ for any nondecreasing function g that is not bounded by a constant. Thus, one may ask if the techniques can be pushed further to prove expressivity bounds for $\mathrm{FO}(\mathbf{C}) + <$. A possible avenue for attacking the problem of expressivity with linear order seems to be the following: try to find a class of structures $\mathcal{C}$ so that both $\mathbf{Def}_{\mathcal{L}}[<, \mathcal{C}]$ and $\mathbf{Sep}_{\mathcal{L}}[Q, \mathcal{C}]$ would hold, where Q is tc, or dtc, or any other query we want to show to be outside of $\mathrm{FO}(\mathbf{C}) + <$. If we were able to find such a class $\mathcal{C}$, it would show that $Q \notin \mathrm{FO}(\mathbf{C}) + <$. Unfortunately, as the following proposition shows, no such class exists.

Proposition 2 *Let Q be a query invariant under isomorphisms. Let $\mathcal{L}$ be* $\mathrm{FO}(\mathbf{C})$, *or* $\mathrm{FO}(Q_\mathbf{u})$, *or* $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$. *Then there does not exist a class of structures $\mathcal{C}$ such that both* $\mathbf{Def}_\mathcal{L}[<,\mathcal{C}]$ *and* $\mathbf{Sep}_\mathcal{L}[Q,\mathcal{C}]$ *hold.*

Proof. Assume that a class of structures $\mathcal{C}$ satisfying both $\mathbf{Def}_\mathcal{L}[<,\mathcal{C}]$ and $\mathbf{Sep}_\mathcal{L}[Q,\mathcal{C}]$ exists. That is, a linear order is definable by an $\mathcal{L}$ formula $\alpha(x,y)$ on large enough structures (cardinality $> n$) in $\mathcal{C}$. Let $r = \mathsf{lr}(\alpha)$, and let $d = 3r+1$. Now apply $\mathbf{Sep}_\mathcal{L}[Q,\mathcal{C}]$ to d and n to find a structure $\mathcal{A}$ of cardinality at least n such that for some $\vec{a},\vec{b}$, we have $\vec{a} \approx_d \vec{b}$. Let a_0 be the first component of $\vec{a}$ and b_0 be the first component of $\vec{b}$. We then have $a_0 \approx_d b_0$.

Assume without loss of generality that $a_0 \prec b_0$ but $b_0 \not\prec a_0$ in the linear order $\prec$ defined by α on $\mathcal{A}$. It follows from [5, 22] that there is a permutation μ on $S_{2r+1}(a_0,b_0)$ such that $N_r(a_0,x) \cong N_r(b_0,\mu(x))$ for every $x \in S_{2r+1}(a_0,b_0)$. From the locality of α and $r = \mathsf{lr}(\alpha)$, we get that for every $x \in S_{2r+1}(a_0,b_0)$, $a_0 \prec x$ iff $b_0 \prec \mu(x)$; that is, μ maps $\{x \in S_{2r+1}(a_0,b_0) \mid a_0 \prec x\}$ into $\{x \in S_{2r+1}(a_0,b_0) \mid b_0 \prec x\}$. But we know that the latter has fewer elements than the former (since $a_0 \prec b_0$ but $b_0 \not\prec a_0$), so we have an injective map from a bigger finite set to a smaller finite set. This contradiction completes the proof. $\square$

4 Expressive power with preorders

While we showed that the technique of Theorem 2 cannot be straightforwardly extended to deal with linear orders, we have not answered the following question: Is $\mathrm{FO}(\mathbf{C})+ <^-_g$ properly contained in $\mathrm{FO}(\mathbf{C})+ <$? If the two were shown to be equal, the bounds of Theorem 2 would apply to prove that the transitive closure is not in TC^0. However, we will show that there is an enormous gap between $\mathcal{L}+ <^-_g$ and $\mathcal{L}+ <$, where $\mathcal{L}$ is one of the counting logics we consider here, and g is very small. Namely, our main result is the following.

Theorem 3 *Let $g : \mathbb{N} \to \mathbb{R}$ be a nondecreasing function that is not bounded by a constant. Then every query in $(\mathcal{L}^*_{\infty\omega}(\mathbf{C})+ <^-_g)_w$ has the bounded number of degrees property.*

That is, with auxiliary structures arbitrarily close to linear orders, the most powerful of counting logics, $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$, still exhibits the very tame behavior typical for FO queries over unordered structures.

The proof of this result is somewhat involved, and will be given in the next section. Here we state some corollaries.

Corollaries With g as above, the (deterministic) transitive closure, and, more generally, problems complete for classes DLOGSPACE and above it under first-order reductions, are not definable in any of the counting logics we consider, even in the presence of relations from $<^-_g$. That is,

Corollary 2 *Let $g : \mathbb{N} \to \mathbb{R}$ be a nondecreasing function that is not bounded by a constant. Then every query in* $(\mathrm{FO}(Q_\mathbf{u})+ <^-_g)_w$, $(\mathrm{FO}(\mathbf{C})+ <^-_g)_w$, $\mathcal{L}^*_{\infty\omega}(\mathbf{C})+ <^-_g$, $\mathrm{FO}(Q_\mathbf{u})+ <^-_g$, *or* $\mathrm{FO}(\mathbf{C})+ <^-_g$ *has the BNDP.*

The following corollaries demonstrate the enormous gain in expressiveness by going from "almost orders" to orders. By a *colored* graph we mean a structure of the signature $(E, U_1, \ldots, U_m)$ where E

is binary, and U_is are unary. That is, it is a graph with a few selected subsets of nodes. A colored graph query is a binary query Q on colored graphs; that is, it returns graphs. The *hardness* of such a query is defined as the function $\mathcal{H}_Q : \mathbb{N} \to \mathbb{N}$ where $\mathcal{H}_Q(n)$ is $\max\{deg\_count(Q(\mathcal{A}))\}$ with $\mathcal{A}$ ranging over structures with $|A| = n$ and E being a successor relation.

Recall that $deg\_count(\cdot)$ is the cardinality of the set of all degrees realized in a structure. That is, the hardness shows how complex the output might look like if the input is a successor relation with a few colored subsets. Note that $0 \leq \mathcal{H}_Q(n) \leq n+1$. Since every property of ordered structures is definable in $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$ [23], we obtain the following dichotomy result:

Corollary 3
- *Let $g : \mathbb{N} \to \mathbb{R}$ be any nondecreasing function that is not bounded by a constant. Let Q be a colored graph query in $\mathcal{L}^*_{\infty\omega}(\mathbf{C})+ <^-_g$. Then there exists a constant C such that $\mathcal{H}_Q(n) < C$ for all n.*
- *For any function $f : \mathbb{N} \to \mathbb{N}$ such that $0 \leq f(n) \leq n+1$, there exists a colored graph query Q in $\mathcal{L}^*_{\infty\omega}(\mathbf{C})+ <$ such that $\mathcal{H}_Q = f$.*

Thus, dropping a tiny portion of linear order (e.g., $\log\log\ldots\log n$ elements) accounts for the increase in hardness from constant to arbitrary one!

FO(**C**) also admits this kind of dichotomy, as there exists a colored graph query Q definable in FO(**C**)+ $<$ such that $\mathcal{H}_Q(n) \geq \log n$ [14]. We thus obtain:

Corollary 4 *There are problems in uniform* TC^0 *that cannot be expressed in* FO(**C**)+ $<^-_g$. □

Moreover, it is known that there are uniform AC^0 (that is, FO(BIT)+ $<$) queries that violate the BNDP ([11], see also [6]). Hence, we obtain:

Corollary 5 $\mathrm{AC}^0 \not\subseteq (\mathcal{L}^*_{\infty\omega}(\mathbf{C})+ <^-_g)_w$. □

We thus can finally compare the expressive power of counting logics with linear orders vs. their expressiveness with preorders:

Corollary 6 *Let $g : \mathbb{N} \to \mathbb{R}$ be as in Theorem 3, and $\mathcal{L}$ be* FO(**C**)*, or* FO($Q_\mathbf{u}$)*, or* $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$*. Then* $\mathcal{L}+ <^-_g \neq \mathcal{L}+ <$*. Furthermore,* FO(**C**) $\subset$ FO(**C**)+ $<^-_g$.

Note that the presence of some form of counting is essential in these results: it was shown recently [12] that every invariant query in FO+ $<$ has the BNDP. That is, $\mathcal{H}_Q$ is bounded by a constant for colored graph queries in FO+ $<$.

5 Proof of the Main Theorem: Failure of Locality, Weak Locality, and Bijective Games

5.1 Failure of locality

All proofs of the BNDP that are currently known derive it from locality of queries. Unfortunately, we cannot use this method as queries in $(\mathcal{L}^*_{\infty\omega}(\mathbf{C})+ <^-_g)_w$ need not be local.

Proposition 3 *Let $g(n) < \frac{\log n}{\log\log n}$ be nondecreasing, and not bounded by a constant. Then there exist nonlocal queries in $(\mathcal{L}^*_{\infty\omega}(\mathbf{C}) + <^-_g)_w$.*

Proof: We construct a query Q definable by a formula $\varphi(x)$ and a sequence of structures $\mathcal{A}_n$, $n \in \mathbb{N}$, with an n-element universe, so that for each n large enough, there are two points a, b in $\mathcal{A}_n$ with isomorphic r-neighborhoods, and $(\mathcal{A}_n, P) \models \varphi(a) \wedge \neg\varphi(b)$, where $r = O(\log\log n)$, for any $P \in <^-_g$. This will prove that $(\mathcal{L}^*_{\infty\omega}(\mathbf{C}) + <^-_g)_w$ is not local. By $\log n$ we mean $\lfloor \log_2(n+1) \rfloor$.

The signature σ consists of three unary relations U_1, U_2 and C, and one binary relation E. We use P for the auxiliary relation from $<^-_g$. Let $l(n) = \lfloor \frac{\log(n - \log n)}{g(n)+1} \rfloor$. This function is not bounded by a constant as n grows. In $\mathcal{A}_n$, whose universe is denoted by A, U_1 has cardinality $M_n = l(n)(g(n)+1) \leq \log(n - \log n)$, and U_2 is interpreted as $A - U_1^{\mathcal{A}_n}$. The unary relation C is interpreted as a two-element subset of U_2. Let E' be defined on U_1 to be a disjoint union of $g(n)+1$ successor relations of length $l(n)$ each. For each such successor relation E'_i, $i = 1, \ldots, g(n)+1$, let c_i be the node at the distance $\lfloor l(n)/3 \rfloor$ from the start node, and d_i be the node at the distance $\lfloor 2 \cdot l(n)/3 \rfloor$ from the start node. Let $C^{\mathcal{A}_n} = \{a, b\}$. We then define

$$E^{\mathcal{A}_n} \quad = \quad E' \ \cup \ \bigcup_{i=1}^{g(n)+1} \{(a, c_i), (b, d_i)\}.$$

Next, define $\alpha(x) \equiv \forall y.(P(x,y) \wedge P(y,x) \rightarrow y = x)$ saying that x is in the linear order part of P, which we shall denote by P_α. From the above, we obtain $M_n \leq \log(|P_\alpha|)$. We now show that there exists a formula $\beta(x,y)$ in FO($\mathbf{C}$) such that $\beta(x,y)$ implies $x, y \in C$ and $(\mathcal{A}_n, P) \models \beta(a,b)$ and $(\mathcal{A}_n, P) \models \neg\beta(b,a)$ for any interpretation of P as a relation from $<^-_g$ on A. This will clearly suffice to conclude the proof, since one then defines $\varphi(x) \equiv C(x) \wedge \exists y.(C(y) \wedge \beta(x,y) \wedge \neg(x = y))$ and notices $(\mathcal{A}_n, P) \models \varphi(a) \wedge \neg\varphi(b)$ while a and b have isomorphic neighborhoods of radius $O(l(n))$.

The formula $\beta(x,y)$ is defined as $C(x) \wedge C(y) \wedge \exists u, v.(E(x,u) \wedge E(y,v) \wedge \gamma(u,v))$ where $\gamma(u,v)$ holds iff there is an E-path from u to v all of whose nodes are in P_α. Since the number of successor relations in E is $g(n)+1$, at least one of them is totally contained in P_α, which shows that $(\mathcal{A}_n, P) \models \beta(a,b)$. Since there is no path between d_i and c_i for every i, we have $(\mathcal{A}_n, P) \models \neg\beta(a,b)$. Thus, we must show how to express γ in $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$ (in fact, one can express it in FO($\mathbf{C}$)).

To express γ, we follow the proof of the failure of the BNDP for FO($\mathbf{C}$)+ $<$ given in [14]. Let $P^1_\alpha = U_1 \cap P_\alpha$. Since $|P^1_\alpha| \leq |U_1| = M_n \leq \log(|P_\alpha|)$, subsets of P^1_α can be coded by the elements of P_α: a set $S \subseteq P^1_\alpha$ is coded by $c_S \in P_\alpha$ such that

$$\{x \mid \mathrm{BIT}(m_1, m_2), \text{ where } \ m_1 = |\{y \mid y < x\}|\,, \ m_2 = |\{y \mid y < c_S\}| \ \} \quad = \quad S$$

With this coding, we can simulate monadic second-order on P^1_α in FO($\mathbf{C}$), which suffices to express γ. This concludes the proof. □

Proposition 3 provides the first nontrivial example that separates the notion of locality from the BNDP. Now one needs a different technique to prove Theorem 3. We introduce this technique in two steps. In the next subsection, we consider two ways of weakening the notion of locality, and we show that one of them, weak semi-locality, implies the BNDP. In subsection 5.3, we show how the bijective games [13] can be used to prove weak semi-locality of $(\mathcal{L}^*_{\infty\omega}(\mathbf{C}) + <^-_g)_w$ queries.

5.2 Weak locality

To define locality of a query, we considered the equivalence relation $\vec{a} \approx_r^{\mathcal{A}} \vec{b}$ iff $N_r^{\mathcal{A}}(\vec{a}) \cong N_r^{\mathcal{A}}(\vec{b})$. We now consider two refinements that lead to weaker notions of locality. First, we write $\vec{a} \mathrel{\overset{\leftrightsquigarrow}{\leftrightsquigarrow}}_r^{\mathcal{A}} \vec{b}$ if $\vec{a} \approx_r^{\mathcal{A}} \vec{b}$ and $S_r^{\mathcal{A}}(\vec{a}) \cap S_r^{\mathcal{A}}(\vec{b}) = \emptyset$.

For the other refinement, consider a partition $\mathcal{I} = (I_1, I_2)$ of the set $\{1, \ldots, n\}$. Given $\vec{x} = (x_1, \ldots, x_n)$, we denote by $\vec{x}_1^{\mathcal{I}}$ and $\vec{x}_2^{\mathcal{I}}$ the subtuples of $\vec{x}$ that consist of those components whose indices belong to I_1 or I_2, respectively. For example, if $n = 4$ and $\mathcal{I} = (\{1,3\}, \{2,4\})$, then $\vec{x}_1^{\mathcal{I}} = (x_1, x_3)$ and $\vec{x}_2^{\mathcal{I}} = (x_2, x_4)$. We then write $\vec{a} \leftrightsquigarrow_r^{\mathcal{A}} \vec{b}$, for $\vec{a}, \vec{b} \in A^n$, if there exists a partition $\mathcal{I} = (I_1, I_2)$ of $\{1, \ldots, n\}$ such that

- $\vec{a}_1^{\mathcal{I}} \approx_r^{\mathcal{A}} \vec{b}_1^{\mathcal{I}}$;
- $\vec{a}_2^{\mathcal{I}} = \vec{b}_2^{\mathcal{I}}$;
- $S_r^{\mathcal{A}}(\vec{a}_1^{\mathcal{I}})$, $S_r^{\mathcal{A}}(\vec{a}_2^{\mathcal{I}})$, $S_r^{\mathcal{A}}(\vec{b}_1^{\mathcal{I}})$ are disjoint.

Clearly, $\vec{a} \mathrel{\overset{\leftrightsquigarrow}{\leftrightsquigarrow}}_r^{\mathcal{A}} \vec{b}$ implies $\vec{a} \leftrightsquigarrow_r^{\mathcal{A}} \vec{b}$ (by taking I_2 to be empty), and $\vec{a} \leftrightsquigarrow_{r+1}^{\mathcal{A}} \vec{b}$ implies $\vec{a} \approx_r^{\mathcal{A}} \vec{b}$.

Definition 4 *An m-ary query Q on σ-structures is called* weakly local *if there exists a number $r \in \mathbb{N}$ such that for any $\mathcal{A} \in \mathrm{STRUCT}[\sigma]$ and any $\vec{a}, \vec{b} \in A^m$, $\vec{a} \mathrel{\overset{\leftrightsquigarrow}{\leftrightsquigarrow}}_r^{\mathcal{A}} \vec{b}$ implies $\vec{a} \in Q(\mathcal{A})$ iff $\vec{b} \in Q(\mathcal{A})$.*

A query Q is said to be weakly semi-local *if there exists a number $r \in \mathbb{N}$ such that for any $\mathcal{A} \in \mathrm{STRUCT}[\sigma]$ and any $\vec{a}, \vec{b} \in A^m$, $\vec{a} \leftrightsquigarrow_r^{\mathcal{A}} \vec{b}$ implies $\vec{a} \in Q(\mathcal{A})$ iff $\vec{b} \in Q(\mathcal{A})$.*

Proposition 4 *Every local query is weakly semi-local, and every weakly semi-local query is weakly local. There exist queries that are weakly local but not weakly semi-local, and there exist queries that are weakly semi-local but not local. That is,*

$$\textsc{Local} \quad \subset \quad \textsc{Weakly Semi-Local} \quad \subset \quad \textsc{Weakly Local}.$$

Proof. The chain of implications local $\Rightarrow$ weakly semi-local $\Rightarrow$ weakly local is obvious. Next, consider the following query Q_0 on graphs. If for the input graph $\langle V, E \rangle$, with vertices V and edges E, $E = \{(x_1, x_2), (x_2, x_3), \ldots, (x_{n-1}, x_n)\} \cup \{(x_i, x_i), (x_j, x_j)\}$ with $i < j$, where $V = \{x_1, \ldots, x_n\}$, then the output of Q_0 is a graph $\langle V, \{(x_i, x_j)\} \rangle$. Otherwise, the output has no edges. Clearly, this query is not local: for any r, we consider a graph as above with $i > r$, $j < n - r$ and $j - i > 2r + 1$; then $N_r(x_i, x_j) \cong N_r(x_j, x_i)$, showing that $\mathsf{lr}(Q_0)$ cannot equal r. At the same time, Q_0 is weakly semi-local, with $r = 1$ witnessing weak semi-locality. Indeed, assume that in a graph G as above $(x_k, x_l) \leftrightsquigarrow_1^G (x_i, x_j)$, with $(x_k, x_l) \neq (x_i, x_j)$. Since there are only two nodes, x_i and x_j, with loops, we get that $k = j$ and $l = i$, but this contradicts $(x_k, x_l) \leftrightsquigarrow_1^G (x_i, x_j)$. Thus, whenever we have $(x_k, x_l) \leftrightsquigarrow_1^G (x_s, x_t)$, none of the pairs is (x_i, x_j) and hence $(x_k, x_l) \notin Q_0(G)$ and $(x_s, x_t) \notin Q_0(G)$, proving weak semi-locality.

To separate weak locality from weak semi-locality, consider another graph query Q_1. If its input $G = \langle V, E \rangle$ is of the form $E = \{(x_1, x_2), (x_2, x_3), \ldots, (x_{n-1}, x_n)\} \cup \{(x_i, x_i), (x_j, x_j), (x_k, x_k)\}$ with $i < j < k$, where $V = \{x_1, \ldots, x_n\}$, then the output of Q_1 is the graph $\langle V, \{(x_i, x_j)\} \rangle$. Otherwise, the output has no edges. To show that it is not weakly semi-local, let G be as above, $r > 0$, and let i, j, k be such that $i > r, j > i + 2r + 1, k > j + 2r + 1, n - k > 2r + 1$. Then $(x_i, x_j) \leftrightsquigarrow_r^G (x_i, x_k)$ but $(x_i, x_j) \in Q_1(G)$ while $(x_i, x_k) \notin Q_1(G)$, and r cannot witness weak semi-locality. To show that Q_1 is weakly local, consider again G as above, and let $(x_i, x_j) \mathrel{\overset{\leftrightsquigarrow}{\leftrightsquigarrow}}_1^G (x_s, x_t)$. Since x_s and x_t must then be distinct and have loops, it is impossible that $S_1^G(x_i, x_j) \cap S_1^G(x_s, x_t) = \emptyset$. Thus, whenever we have

$(x_s, x_t) \leftrightsquigarrow_1^G (x_p, x_q)$, none of the pairs is (x_i, x_j) and hence $(x_s, x_t) \notin Q_1(G)$ and $(x_p, x_q) \notin Q_1(G)$. Therefore, Q_1 is weakly local. $\Box$

We study these notions because they are easier to prove than the BNDP, and we will see that the BNDP can be derived from them. The notion of weak locality is particularly simple: the only difference between it and locality is the disjointness of neighborhoods. However, it only gives us a partial result:

Proposition 5 *a) Let Q be a binary weakly local query (i.e., the output is a graph). Then Q has the bounded number of degrees property.*

b) For every $m > 2$, there exists an m-ary weakly local query that does not have the bounded number of degrees property.

Proof. We first prove a). Fix the relational signature σ, and we let $F_0(d,k)$ be an upper bound on the size of a d-neighborhood of a point in $\mathcal{A}$ where degrees are bounded by k, and $F_1(d,k)$ be an upper bound on the number of isomorphism types of d-neighborhoods of points in such structures; such bounds do exist and depend on σ, d and k only [22].

Let Q be a binary weakly local query, with d witnessing weak locality. We now calculate the number of different outdegrees in the graph $Q(\mathcal{A})$, where all degrees in the structure $\mathcal{A}$ are at most k. Whenever we say *out-deg*(x), we mean out-degree in $Q(\mathcal{A})$.

Claim 1 *There is a number $M_{d,k}$ that depends on d and k only, such that*

$$|\textit{out-deg}(a) - \textit{out-deg}(b)| \ \leq \ M_{d,k}$$

whenever $a \leftrightsquigarrow_{2d+1} b$.

Proof of the claim: We call an isomorphism type τ of a d-neighborhood of a point (a,b)-*good* if there exist three points, $c_1, c_2, c_3 \in A - S_{2d+1}(a,b)$ such that c_1, c_2, c_3 realize τ and $d(c_i, c_j) > 4d, i, j = 1,2,3, i \neq j$. We call τ (a,b)-*bad* otherwise.

Let c be a point of a (a,b)-good type τ. Let c_1, c_2, c_3 witness the goodness of the τ. Then at most one of them can belong $S_{2d}(c)$ (otherwise the distance would be under $4d$). Thus, there are two points c', c'' of type τ such that $d(c,c'), d(c,c''), d(c',c'') > 2d$; that is, their d-neighborhoods are disjoint. We now obtain

$$\begin{array}{rrcl} & (a,c) & \in & Q(\mathcal{A}) \\ \Leftrightarrow & (b,c') & \in & Q(\mathcal{A}) \\ \Leftrightarrow & (a,c'') & \in & Q(\mathcal{A}) \\ \Leftrightarrow & (b,c) & \in & Q(\mathcal{A}) \end{array}$$

by weak locality.

Now, letting M_0 be the number of points realizing (a,b)-bad types, we see that the difference between *out-deg*(a) and *out-deg*(b) cannot exceed the size of $S_{2d+1}(a,b) + M_0$; that is,

$$|\textit{out-deg}(a) - \textit{out-deg}(b)| \ \leq \ M_0 + 2 \cdot F_0(2d+1, k)$$

It thus remains to show that M_0 is determined by d and k.

Fix an (a,b)-bad type τ, and let $x \notin S_{2d+1}(a,b)$ realize τ. Suppose $y \notin S_{2d+1}(a,b) \cup S_{4d}(x)$ realize τ. Then every other point z realizing τ must be either in $S_{2d+1}(a,b)$ or in $S_{4d}(x,y)$, for otherwise x, y, z

would witness (a,b)-goodness of τ. Thus, the number of points realizing τ is at most $2 \cdot F_0(2d+1,k) + 2 \cdot F_0(4d,k)$, and hence M_0 is bounded above by

$$F_1(d,k) \cdot (2 \cdot F_0(2d+1,k) + 2 \cdot F_0(4d,k))$$

finishing the proof of the claim. $\square$

Using this, we show the following.

Claim 2 *Let $m_0 = F_0(8d+4,k)+1$. Suppose $a \approx_{2d+1} b$, and suppose that there are at least m_0 realizers of the isomorphism type of the $2d+1$-neighborhood of a. Then*

$$|\textit{out-deg}(a) - \textit{out-deg}(b)| \;\le\; 2 \cdot M_{d,k}$$

Proof of the claim: If $S_{2d+1}(a) \cap S_{2d+1}(b) = \emptyset$, this follows from the previous claim. Assume then $S_{2d+1}(a) \cap S_{2d+1}(b) \neq \emptyset$. We have $S_{2d+1}(a,b) \subseteq S_{6d+3}(a) = C$, and from the assumptions, we obtain there there exists an element $c \notin S_{2d+1}(C)$ such that $a \approx_{2d+1} c$. In particular, in this case $S_{2d+1}(a) \cap S_{2d+1}(c) = \emptyset$ and $S_{2d+1}(b) \cap S_{2d+1}(c) = \emptyset$. Then by the previous claim we have $|\textit{out-deg}(a) - \textit{out-deg}(c)| \le M_{d,k}$ and $|\textit{out-deg}(b) - \textit{out-deg}(c)| \le M_{d,k}$, which proves the claim. $\square$

Let now m_0 be as in Claim 2. Suppose τ is an isomorphism type of a $2d+1$-neighborhood that has fewer than m_0 realizers. The total number of points of such types is at most $M_1 = F_1(2d+1,k) \cdot m_0$, and thus they give rise to at most M_1 different outdegrees in the output. For any point a of a type τ (of $2d+1$-neighborhood) that is realized at least m_0 times, the possible outdegrees belong to a $2(2M_{d_k})+1$ element interval, by Claim 2. Thus, the total number of outdegrees in $Q(\mathcal{A})$ is at most

$$M_1 + F_1(2d+1,k) \cdot (4M_{d,k}+1)$$

and thus depends on the signature, d and k only. The proof for a bound on the number of indegrees is identical. This completes the proof of the BNDP for weakly local graph queries.

To show b), we consider graphs as inputs, and let $m = 3$; extension to $m > 3$ is straightforward. For a graph $G = \langle V, E \rangle$, with vertices V and edges E, $(a,b,c) \in Q(G)$ iff the following two conditions hold. First, the graph G is of the special form: there is an element $v \in V$ such that $(v,v) \in E$, $(v,v'), (v',v) \notin E$ for any $v' \neq v$, and G restricted to $V - \{v\}$ is a chain (i.e., the graph of a successor relation). Second, $c = v$, $a, b \neq v$, and (a,b) is in the transitive closure of G.

Clearly, Q violates the BNDP: for every $a \in V - \{v\}$, there are k_a tuples (a,b,c) in $Q(G)$, where k_a is the number of nodes reachable from a. Thus, $\textit{deg\_count}(Q(G)) = O(|V|)$. On the other hand, Q is weakly local (in fact, $r = 1$). Indeed, suppose $(a,b,c) \in Q(G)$ and $(a',b',c') \notin Q(G)$. Then $c = v$, and since there is only one loop in G, for $(a,b,c) \approx_1 (a',b',c')$ to hold we must have $c = c' = v$. However, in this case $S_r(a,b,c) \cap S_r(a',b',c') \neq \emptyset$, which shows that $(a,b,c) \approx\!\!\approx_1 (a',b',c')$ does not hold, and thus proves the weak locality of Q. This completes the proof. $\square$

Combined with the results of Section 5.3, that would be sufficient to derive Theorem 3 for queries that return graphs. However, for arbitrary queries, we need the more involved notion of weak semi-locality:

Theorem 4 *Every weakly semi-local query has the bounded number of degrees property.*

Proof. Let $\vec{x} = (x_1, \ldots, x_n)$ and let $\mathcal{I} = \{I_1, I_2, I_3\}$ be a partition of $\{1, \ldots, n\}$. Then by $\vec{x}_j^{\mathcal{I}}$, $j = 1, 2, 3$, we denote the subtuple of $\vec{x}$ which consists of the components of $\vec{x}$ whose indices are in I_j, appearing in the same order as in $\vec{x}$.

Given a σ-structure $\mathcal{A}$, two positive integers $d, r > 0$ and $a \leftrightsquigarrow_d^{\mathcal{A}} b$, we define a binary relation $\twoheadleftarrow\!\!\twoheadrightarrow_{r,l}^{(\mathcal{A},a,b)}$ on A^l as follows. Given $\vec{x}, \vec{y} \in A^l$, $\vec{x} \twoheadleftarrow\!\!\twoheadrightarrow_{r,l}^{(\mathcal{A},a,b)} \vec{y}$ iff there exists an isomorphism $h : N_d^{\mathcal{A}}(a) \to N_d^{\mathcal{A}}(b)$ and a partition $\mathcal{I}$ of $\{1, \ldots, l\}$ such that

- $S_r^{\mathcal{A}}(\vec{x}_1^{\mathcal{I}}) \subseteq S_d^{\mathcal{A}}(a)$ and $\vec{y}_1^{\mathcal{I}} = h(\vec{x}_1^{\mathcal{I}})$;
- $S_r^{\mathcal{A}}(\vec{x}_2^{\mathcal{I}}) \subseteq S_d^{\mathcal{A}}(b)$ and $\vec{y}_2^{\mathcal{I}} = h^{-1}(\vec{x}_2^{\mathcal{I}})$;
- $\vec{x}_3^{\mathcal{I}} = \vec{y}_3^{\mathcal{I}}$ and $S_r^{\mathcal{A}}(\vec{x}_3^{\mathcal{I}}) \cap (S_r^{\mathcal{A}}(a,b) \cup S_r^{\mathcal{A}}(\vec{x}_1^{\mathcal{I}}\vec{x}_2^{\mathcal{I}}\vec{y}_1^{\mathcal{I}}\vec{y}_2^{\mathcal{I}})) = \emptyset$.

Note that these conditions imply $S_r^{\mathcal{A}}(\vec{y}_1^{\mathcal{I}}) \subseteq S_d^{\mathcal{A}}(b)$ and $S_r^{\mathcal{A}}(\vec{y}_2^{\mathcal{I}}) \subseteq S_d^{\mathcal{A}}(a)$.

We now need the following lemma.

Lemma 1 *For any positive integers r and l, there exists a positive integer d such that for any relational vocabulary σ, any σ-structure $\mathcal{A}$ and $a \leftrightsquigarrow_d^{\mathcal{A}} b$, there exists a permutation $\pi : A^l \to A^l$ such that*

$$a\vec{x} \twoheadleftarrow\!\!\twoheadrightarrow_{(r,l)}^{(\mathcal{A},a,b)} b\pi(\vec{x})$$

for every $\vec{x} \in A^l$.

Proof of the lemma: Let $d_0 = r, d_1 = 3d_0 + 1, \ldots, d_l = 3d_{l-1} + 1$. We claim that $d = d_l$. The proof is by induction on l. If $l = 1$, we define $\pi : A \to A$ as follows:

$$\pi(x) \quad = \quad \begin{cases} h(x), & \text{if } x \in S_{2r+1}^{\mathcal{A}}(a); \\ h^{-1}(x), & \text{if } x \in S_{2r+1}^{\mathcal{A}}(b); \\ x, & \text{otherwise.} \end{cases}$$

The partition of the set $\{1\}$ containing the unique index is determined as follows:

$$\mathcal{I} \quad = \quad \begin{cases} (\{1\}, \emptyset, \emptyset), & \text{if } x \in S_{2r+1}^{\mathcal{A}}(a); \\ (\emptyset, \{1\}, \emptyset), & \text{if } x \in S_{2r+1}^{\mathcal{A}}(b); \\ (\emptyset, \emptyset, \{1\}), & \text{otherwise.} \end{cases}$$

It is routine to verify that π is a permutation and $x \twoheadleftarrow\!\!\twoheadrightarrow_{(r,1)}^{(\mathcal{A},a,b)} \pi(x)$.

Now assume $\vec{a} \leftrightsquigarrow_d^{\mathcal{A}} \vec{b}$, where $d = d_l$, $l > 1$. By the hypothesis for $l - 1$, we find a permutation $\mu : A^{l-1} \to A^{l-1}$ such that for any $\vec{x} \in A^{l-1}$, $a\vec{x} \twoheadleftarrow\!\!\twoheadrightarrow_{(d_1,l-1)}^{(\mathcal{A},a,b)} b\mu(\vec{x})$; that is,

$$a\vec{x} \twoheadleftarrow\!\!\twoheadrightarrow_{(3r+1,l-1)}^{(\mathcal{A},a,b)} b\mu(\vec{x}).$$

We now show that for every $\vec{x} \in A^{l-1}$, there exists a permutation $\eta_{\vec{x}} : A \to A$ such that $a\vec{x}z \twoheadleftarrow\!\!\twoheadrightarrow_{(r,l)}^{(\mathcal{A},a,b)} b\mu(\vec{x})\eta_{\vec{x}}(z)$; this will suffice to conclude the proof as the function given by $(x_1, \ldots, x_{l-1}, x_l) \mapsto \mu((x_1, \ldots, x_{l-1}))\eta_{(x_1,\ldots,x_{l-1})}(x_l)$ is a permutation.

Fix an isomorphism $h : N_d^{\mathcal{A}}(a) \to N_d^{\mathcal{A}}(b)$ and a partition $\mathcal{I} = (I_1, I_2, I_3)$ witnessing $a\vec{x} \mathrel{\twoheadleftarrow\!\!\twoheadrightarrow}^{(\mathcal{A},a,b)}_{(3r+1,l-1)} b\mu(\vec{x})$; that is, $S_{3r+1}^{\mathcal{A}}(\vec{x}_1^{\mathcal{I}}) \subseteq S_d^{\mathcal{A}}(a)$, $S_{3r+1}^{\mathcal{A}}(\vec{x}_2^{\mathcal{I}}) \subseteq S_d^{\mathcal{A}}(b)$, $\vec{y}_1^{\mathcal{I}} = h(\vec{x}_1^{\mathcal{I}})$, $\vec{y}_1^{\mathcal{I}} = h^{-1}(\vec{x}_1^{\mathcal{I}})$, and $S_{3r+1}^{\mathcal{A}}(\vec{x}_3^{\mathcal{I}})$ does not intersect the $3r+1$-spheres of $a, b, \vec{x}_1^{\mathcal{I}}, \vec{x}_2^{\mathcal{I}}, \vec{y}_1^{\mathcal{I}}$, and $\vec{y}_2^{\mathcal{I}}$. Below we show how to define $\eta_{\vec{x}}$ and a new partition $\mathcal{I}'$ that will witness $a\vec{x}z \mathrel{\twoheadleftarrow\!\!\twoheadrightarrow}^{(\mathcal{A},a,b)}_{(r,l)} b\mu(\vec{x})\eta_{\vec{x}}(z)$; $\mathcal{I}'$ is obtained from $\mathcal{I}$ by adding the last index l to one of its blocks. The isomorphism h remains the same.

$$\eta_{\vec{x}}(z) = \begin{cases} h(x), & \text{if } x \in S_{2r+1}^{\mathcal{A}}(a) \cup S_{2r+1}(\vec{x}_1^{\mathcal{I}}) \cup S_{2r+1}^{\mathcal{A}}(\vec{y}_2^{\mathcal{I}}); \quad (I_1' := I_1 \cup \{l\}); \\ h^{-1}(x), & \text{if } x \in S_{2r+1}^{\mathcal{A}}(b) \cup S_{2r+1}(\vec{y}_1^{\mathcal{I}}) \cup S_{2r+1}^{\mathcal{A}}(\vec{x}_2^{\mathcal{I}}); \quad (I_2' := I_2 \cup \{l\}); \\ x, & \text{otherwise}; \quad (I_3' := I_3 \cup \{l\}). \end{cases}$$

Consider the first case. Let $z' = \eta_{\vec{x}}(z)$. Since $C = S_{2r+1}^{\mathcal{A}}(a) \cup S_{2r+1}(\vec{x}_1^{\mathcal{I}}) \cup S_{2r+1}^{\mathcal{A}}(\vec{y}_2^{\mathcal{I}})$ has the property that $S_r^{\mathcal{A}}(C) \subseteq S_d^{\mathcal{A}}(a)$, we have $S_r^{\mathcal{A}}(\vec{x}_1^{\mathcal{I}'}) \subseteq S_d^{\mathcal{A}}(a)$, and since $z' = h(z)$, we obtain $S_r^{\mathcal{A}}(\vec{y}_1^{\mathcal{I}'}) \subseteq S_d^{\mathcal{A}}(b)$. Furthermore, $S_r^{\mathcal{A}}(z) \subseteq S_{3r+1}^{\mathcal{A}}(a) \cup S_{3r+1}(\vec{x}_1^{\mathcal{I}}) \cup S_{3r+1}^{\mathcal{A}}(\vec{y}_2^{\mathcal{I}})$, we obtain $S_r^{\mathcal{A}}(z) \cap S_{3r+1}^{\mathcal{A}}(\vec{x}_3^{\mathcal{I}}) = \emptyset$ and thus $S_r^{\mathcal{A}}(z) \cap S_r^{\mathcal{A}}(\vec{x}_3^{\mathcal{I}}) = \emptyset$. Similarly, $S_r^{\mathcal{A}}(z') \cap S_r^{\mathcal{A}}(\vec{x}_3^{\mathcal{I}}) = \emptyset$. The proof of correctness in the second case is identical. For the third case, as $z \notin S_{2r+1}^{\mathcal{A}}(a,b) \cup S_{2r+1}^{\mathcal{A}}(\vec{x}_1^{\mathcal{I}}\vec{x}_2^{\mathcal{I}}\vec{y}_1^{\mathcal{I}}\vec{y}_2^{\mathcal{I}})$, we obtain $S_r^{\mathcal{A}}(z) \cap (S_r^{\mathcal{A}}(a,b) \cup S_r^{\mathcal{A}}(\vec{x}_1^{\mathcal{I}}\vec{x}_2^{\mathcal{I}}\vec{y}_1^{\mathcal{I}}\vec{y}_2^{\mathcal{I}})) = \emptyset$, completing the proof that $a\vec{x}z \mathrel{\twoheadleftarrow\!\!\twoheadrightarrow}^{(\mathcal{A},a,b)}_{(r,l)} b\mu(\vec{x})\eta_{\vec{x}}(z)$. The lemma is proved. □

We now prove the theorem using the lemma. Let Q be an m-ary weakly semi-local query, $m > 1$, and let r witness weak semi-locality, i.e., $\vec{a} \leftrightsquigarrow_r^{\mathcal{A}} \vec{b}$ implies $\vec{a} \in Q(\mathcal{A})$ iff $\vec{b} \in Q(\mathcal{A})$. Let d be the positive integer given by Lemma 1 for r and $l = m - 1$. We now fix $k > 0$ and let $\mathcal{A} \in \mathrm{STRUCT}_k[\sigma]$. As before, let $F_0(d,k)$ be the maximum size of a d-neighborhood around a point in a σ-structure whose degrees do not exceed k; such a bound exists and is determined by d, k and σ [22, 14]. Let τ be an isomorphism type of a d-neighborhood of a point in a σ-structure. We call it $\mathcal{A}$-*good* if there are more than $(m+1)F_0(2d+1,k)$ realizers of τ in $\mathcal{A}$, and $\mathcal{A}$-*bad* if there are at most $(m+1)F_0(2d+1,k)$ realizers.

In what follows, the structure $\mathcal{A} \in \mathrm{STRUCT}_k[\sigma]$ is fixed, and *degree* refers to the degree in the output $Q(\mathcal{A})$.

Claim 3 *Let $a \mathrel{\substack{\leftrightsquigarrow\\ \leftrightsquigarrow}}_d b$, and assume that the isomorphism type of the d-neighborhood of a is $\mathcal{A}$-good. Then $degree_1(a) = degree_1(b)$.*

Proof of the claim: Assume that $a\vec{x} \mathrel{\twoheadleftarrow\!\!\twoheadrightarrow}^{(\mathcal{A},a,b)}_{(r,m-1)} b\vec{y}$, and let $\mathcal{I} = (I_1, I_2, I_3)$ be a partition of $\{1, \ldots, m-1\}$ and $h : N_d(a) \to N_d(b)$ an isomorphism witnessing that. Assume that $a\vec{x} \in Q(\mathcal{A})$. Let $C = S_{2d+1}(a,b) \cup S_{2d+1}(\vec{x}_3^{\mathcal{I}})$. Since $\vec{x}_3^{\mathcal{I}}$ has at most $m-1$ elements, $|C| \leq (m+1)F_0(2d+1,k)$, and thus there exists $c \notin C$ such that $c \approx_d a \approx_d b$, since the type of a is $\mathcal{A}$-good. Note that $a \mathrel{\substack{\leftrightsquigarrow\\ \leftrightsquigarrow}}_d c$, $b \mathrel{\substack{\leftrightsquigarrow\\ \leftrightsquigarrow}}_d c$ and $S_d(c) \cap S_d(\vec{x}_3^{\mathcal{I}}) = \emptyset$. Let $h_a : N_d(a) \to N_d(c)$ be an isomorphism. Let $\vec{z}_1 = h_(\vec{x}_1^{\mathcal{I}})$. Note that $S_r(c\vec{z}_1) \subseteq S_d(c)$. We then obtain $a\vec{x}_1^{\mathcal{I}}\vec{x}_2^{\mathcal{I}}\vec{x}_3^{\mathcal{I}} \leftrightsquigarrow_r c\vec{z}_1\vec{x}_2\vec{x}_3^{\mathcal{I}}$. By weak semi-locality of Q, we obtain $c\vec{z}_1\vec{x}_2^{\mathcal{I}}\vec{x}_3^{\mathcal{I}} \in Q(\mathcal{A})$. Since $S_r(\vec{x}_2^{\mathcal{I}}) \cap S_r(\vec{y}_2^{\mathcal{I}}) = \emptyset$, we obtain $c\vec{z}_1\vec{x}_2^{\mathcal{I}}\vec{x}_3^{\mathcal{I}} \leftrightsquigarrow_r c\vec{z}_1\vec{y}_2^{\mathcal{I}}\vec{y}_3^{\mathcal{I}}$, and thus $c\vec{z}_1\vec{y}_2^{\mathcal{I}}\vec{y}_3^{\mathcal{I}} \in Q(\mathcal{A})$. Next, notice that $h \circ h_a^{-1}$ maps $N_r(c\vec{z}_1)$ isomorphically onto $N_r(b\vec{y}_1^{\mathcal{I}})$, which is disjoint from $S_r(\vec{y}_2^{\mathcal{I}}\vec{y}_3^{\mathcal{I}})$; we thus conclude that $c\vec{z}_1\vec{y}_2^{\mathcal{I}}\vec{y}_3^{\mathcal{I}} \leftrightsquigarrow_r b\vec{y}_1^{\mathcal{I}}\vec{y}_2^{\mathcal{I}}\vec{y}_3^{\mathcal{I}} = b\vec{y}$, and $b\vec{y} \in Q(\mathcal{A})$ by weak semi-locality of Q.

An identical proof shows that $b\vec{y} \in Q(\mathcal{A})$ implies $a\vec{x} \in Q(\mathcal{A})$. Thus, by Lemma 1, we have a permutation π on A^{m-1} such that $a\vec{x} \in Q(\mathcal{A})$ iff $b\pi(\vec{x}) \in Q(\mathcal{A})$, which proves $degree_1(a) = degree_1(b)$. □

Using this claim, we show that if $a \approx_d^{\mathcal{A}} b$, and the isomorphism type of $N_d^{\mathcal{A}}(b)$, τ, is $\mathcal{A}$-good, then $degree_1(a) = degree_1(b)$. Indeed, if $S_d^{\mathcal{A}}(a) \cap S_d^{\mathcal{A}}(b) = \emptyset$, then $a \leftrightsquigarrow_d^{\mathcal{A}} b$, and we are done by Claim 3. If $S_d^{\mathcal{A}}(a) \cap S_d^{\mathcal{A}}(b) \neq \emptyset$, then, as the cardinality of $S_{2d+1}^{\mathcal{A}}(a,b)$ is at most $2F_0(2d+1,k)$, there is a point $c \notin S_{2d+1}^{\mathcal{A}}(a,b)$ realizing τ. We have now $a \leftrightsquigarrow_d^{\mathcal{A}} c$ and $b \leftrightsquigarrow_d^{\mathcal{A}} c$, and thus $degree_1(a) = degree_1(c)$, and $degree_1(b) = degree_1(c)$; hence $degree_1(a) = degree_1(b)$.

We now calculate the number of different values of $degree_1(\cdot)$. The cardinality of $\{degree_1(a) \mid \text{type of } N_d^{\mathcal{A}}(a) \text{ is } \mathcal{A} - \text{good}\}$ is at most $F_1(d,k)$, the maximum possible number of different isomorphism types of d-neighborhoods in a structure from $\mathrm{STRUCT}_k[\sigma]$. This number depends on d, k and σ only. The cardinality of $\{degree_1(a) \mid \text{type of } N_d^{\mathcal{A}}(a) \text{ is } \mathcal{A} - \text{bad}\}$ is at most the number of points realizing $\mathcal{A}$-bad types. As each $\mathcal{A}$-bad type has at most $(m+1)F_0(2d+1,k)$ realizers, the number of realizers of $\mathcal{A}$-bad types is at most $(m+1) \cdot F_0(2d+1,k) \cdot F_1(d,k)$. We thus obtain that

$$|\{degree_1(a) \mid a \in A\}| \quad \leq \quad F_1(d,k) + (m+1) \cdot F_0(2d+1,k) \cdot F_1(d,k)$$

and thus depends only on d, k and σ. As identical proofs work for all $degree_i$, we get that $deg\_count(Q(\mathcal{A})) \leq m \cdot F_1(d,k) \cdot ((m+1)F_0(2d+1,k)+1)$, which finishes the proof of the BNDP. □

To incorporate the information about the function g, we modify the definition as follows: $\vec{a} \leftrightsquigarrow_{g,r}^{\mathcal{A}} \vec{b}$ if $\vec{a} \leftrightsquigarrow_r^{\mathcal{A}} \vec{b}$ and $\left|S_r^{\mathcal{A}}(\vec{a}) \cup S_r(\vec{b})\right| \leq g(|A|)$. Then a query Q is *g-weakly semi-local* if there exists an $r \in \mathbb{N}$ such that $\vec{a} \leftrightsquigarrow_{g,r}^{\mathcal{A}} \vec{b}$ implies $\vec{a} \in Q(\mathcal{A})$ iff $\vec{b} \in Q(\mathcal{A})$. The following is easily derived from Theorem 4.

Corollary 7 *Let $g : \mathbb{N} \to \mathbb{R}$ be nondecreasing and not bounded by a constant. Then every g-weakly semi-local query has the BNDP.*

Proof. Let Q be g-weakly local m-ary query, with d witnessing weak locality. Let $N_{d,k}$ to be the smallest number such that $g(N) > 2m \cdot F_0(d,k)$ for any $N > N_{d,k}$. Then, if $\mathcal{A} \in \mathrm{STRUCT}_k[\sigma]$ and $|A| > N_{d,k}$, for $\vec{a}, \vec{b} \in A^m$, $\vec{a} \leftrightsquigarrow_{g,r}^{\mathcal{A}} \vec{b}$ implies $\vec{a} \leftrightsquigarrow_r \vec{b}$, since $S_d(\vec{a}) \cup S_r(\vec{b})$ has fewer than $g(|A|)$ elements. Then the proof of Theorem 4 applies verbatim to show that for some function $f_0 : \mathbb{N} \to \mathbb{N}$, $deg\_count(Q(\mathcal{A})) \leq f_0(k)$. Since $deg\_count(Q(\mathcal{A})) \leq N_{d,k} + 1$ if $|A| \leq N_{d,k}$, we obtain that $deg\_count(Q(\mathcal{A})) \leq \max\{N_{d,k}^{m-1} + 1, f_0(k)\}$, thus proving the BNDP. □

5.3 Games and weak semi-locality

The goal of this section is to prove the g-weak semi-locality of queries in $(\mathcal{L}_{\infty\omega}^*(\mathbf{C}) + <_g^-)_w$. We do this by using *bijective games* of [13].

The game is played by two players, called the spoiler and the duplicator, on two structures $\mathcal{A}, \mathcal{B} \in \mathrm{STRUCT}[\sigma]$. For the n-round game, in each round $i = 1, \ldots, n$, the duplicator selects a bijection $f_i : A \to B$, where B is the carrier of $\mathcal{B}$, and the spoiler selects a point $a_i \in A$ (if $card(A) \neq card(B)$, then the spoiler immediately wins). The duplicator wins after n rounds if the relation $\{(a_i, f_i(a_i)) \mid 1 \leq i \leq n\}$ is a partial isomorphism $\mathcal{A} \to \mathcal{B}$; otherwise the spoiler wins. If the duplicator has a winning strategy in the n-move bijective game on $\mathcal{A}$ and $\mathcal{B}$, we write $\mathcal{A} \equiv_n^{bij} \mathcal{B}$. We write $(\mathcal{A}, \vec{a}) \equiv_n^{bij} (\mathcal{B}, \vec{b})$ if the duplicator has a winning strategy in the n-move bijective game that starts with the position $(\vec{a}, \vec{b})$. This condition implies that for a FO (or $\mathrm{FO}(Q_{\mathbf{u}})$) formula $\varphi(\vec{x})$ of quantifier rank n, $\mathcal{A} \models \varphi(\vec{a})$ iff $\mathcal{B} \models \varphi(\vec{b})$ [13]. We extend this to $\mathcal{L}_{\infty\omega}^*(\mathbf{C})$. Note that the lemma below follows from a slightly more general result of [15].

Lemma 2 *Let $\varphi(x_1, \ldots, x_m)$ be a $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$ formula in the language of σ, with all free variables of the first sort. Let $(\mathcal{A}, \vec{a}) \equiv^{bij}_{\mathsf{rk}(\varphi)} (\mathcal{B}, \vec{b})$, where $\vec{a} \in A^m, \vec{b} \in B^m$. Then $\mathcal{A} \models \varphi(\vec{a})$ iff $\mathcal{B} \models \varphi(\vec{b})$.* $\square$

The following is the key lemma, which is proved by a technique reminiscent of that in [33], extended to deal with bijective games.

Lemma 3 *Let $g : \mathbb{N} \to \mathbb{R}$ be nondecreasing and not bounded by a constant. For any $\mathcal{A}$, $m > 0$, $\vec{a}, \vec{b} \in A^m$, and $n > 0$, if $\vec{a} \leftrightsquigarrow^{\mathcal{A}}_{g,2^n} \vec{b}$, then there exists a preorder P on A such that $P \in <^{-}_{g}$ and*

$$(\mathcal{A}, P, \vec{a}) \quad \equiv^{bij}_{n} \quad (\mathcal{A}, P, \vec{b})$$

Proof. Let $r = 2^n$ and $\vec{a} \leftrightsquigarrow^{\mathcal{A}}_{g,r} \vec{b}$. Let $\mathcal{I} = (I_1, I_2)$ be a partition witnessing that. We assume without loss of generality that I_1 is nonempty and equals $\{1, \ldots, l\}$, $l \leq m$. Let $\vec{a}' = (a_1, \ldots, a_l)$, $\vec{b}' = (b_1, \ldots, b_l)$, and $\vec{c} = (a_{l+1}, \ldots, a_m) = (b_{l+1}, \ldots, b_m)$. Then $\vec{a}' \overset{\leftrightsquigarrow}{\leftrightsquigarrow}{}^{\mathcal{A}}_{r} \vec{b}'$, $S^{\mathcal{A}}_r(\vec{a}'\vec{b}') \cap S^{\mathcal{A}}_r(\vec{c}) = \emptyset$, and $\left|S^{\mathcal{A}}_r(\vec{a}'\vec{b}'\vec{c})\right| \leq g(|A|)$.

We now construct P. Let A_0 be $S^{\mathcal{A}}_r(\vec{a}') - \{a_1, \ldots, a_l\}$. Pick any ordering $\prec_1$ on $S^{\mathcal{A}}_r(\vec{a}')$ such that $a_1 \prec_1 a_2 \prec_1 \ldots \prec_1 a_l$ and further, for any $a \in S^{\mathcal{A}}_r(\vec{a}') - \{a_1, \ldots, a_l\}$ we have $a_i \prec_1 a$, for each $i = 1, \ldots, l$, and for any $a', a'' \in S^{\mathcal{A}}_r(\vec{a}') - \{a_1, \ldots, a_l\}$,

$$d(a', \vec{a}') < d(a'', \vec{a}') \qquad \Rightarrow \qquad a' \prec_1 a''$$

Let h be an isomorphism of $N^{\mathcal{A}}_r(\vec{a})$ onto $N^{\mathcal{A}}_r(\vec{b})$. Define, on $S^{\mathcal{A}}_r(\vec{b}')$, an ordering $\prec_2$ by letting $b' \prec_2 b''$ iff $h^{-1}(b') \prec_1 h^{-1}(b'')$. Clearly, the initial fragment of $\prec_2$ is $(b_1, \ldots, b_l)$, and it respects the distance to $\vec{b}'$: $d(b', \vec{b}') < d(b'', \vec{b}')$ implies $b' \prec_2 b''$.

Let P_0 be an arbitrary linear ordering on $A - S^{\mathcal{A}}_r(\vec{a}'\vec{b}')$. Intuitively, P is P_0 followed by a preorder obtained by putting together $\prec_1$ and $\prec_2$, and tying them by h. Formally,

$$(x, y) \in P \;\text{ iff }\; \begin{cases} x, y \notin S^{\mathcal{A}}_r(\vec{a}'\vec{b}') \text{ and } (x, y) \in P_0, \text{ or} \\ x \notin S^{\mathcal{A}}_r(\vec{a}'\vec{b}') \text{ and } y \in S^{\mathcal{A}}_r(\vec{a}'\vec{b}'), \text{ or} \\ x \in S^{\mathcal{A}}_r(\vec{a}'), y \in S^{\mathcal{A}}_r(\vec{a}') \text{ and } x \prec_1 y, \text{ or} \\ x \in S^{\mathcal{A}}_r(\vec{b}'), y \in S^{\mathcal{A}}_r(\vec{b}') \text{ and } x \prec_2 y, \text{ or} \\ x \in S^{\mathcal{A}}_r(\vec{a}'), y \in S^{\mathcal{A}}_r(\vec{b}') \text{ and } h(x) \prec_2 y, \text{ or} \\ x \in S^{\mathcal{A}}_r(\vec{b}'), y \in S^{\mathcal{A}}_r(\vec{a}') \text{ and } x \prec_2 h(y) \end{cases}$$

It easily follows from $\vec{a}' \overset{\leftrightsquigarrow}{\leftrightsquigarrow}{}^{\mathcal{A}}_{g,r} \vec{b}'$ that $P \in <^{-}_{g}$.

Our next claims give a winning strategy for the duplicator in the bijective game on $\mathcal{A}_{\vec{a}} = (\mathcal{A}, P, \vec{a})$ and $\mathcal{A}_{\vec{b}} = (\mathcal{A}, P, \vec{b})$. Note that the universe of both structures is the same, A, and in the game the spoiler selects points in A, and the duplicator select bijections $f : A \to A$.

Define a binary relation H on $S^{\mathcal{A}}_r(\vec{a}'\vec{b}')$ by letting $(x, y) \in H$ iff $x = h(y)$ or $y = h(x)$. We will show that the duplicator can play in such a way that, if $\vec{x} = (x_1, \ldots, x_n)$ and $\vec{y} = (y_1, \ldots, y_n)$ are points played on $\mathcal{A}_{\vec{a}}$ and $\mathcal{A}_{\vec{b}}$ respectively after n rounds, then there exists a set $J \subseteq \{1, \ldots, n\}$ with the following properties. (1) If $j \in J$, then $(x_j, y_j) \in H$. (2) If $j \notin J$, then $x_j = y_j$. (3) $\vec{a}'\vec{x}^J \approx^{\mathcal{A}}_0 \vec{b}'\vec{y}^J$, where $\vec{x}^J$ is the subtuple of $\vec{x}$ that consists of the component of $\vec{x}$ whose indices are in J, and likewise for $\vec{y}^J$. (4) $d_{\mathcal{A}}(\vec{a}'\vec{x}^J, \vec{x}^{\overline{J}}) > 1$, and $d_{\mathcal{A}}(\vec{b}'\vec{y}^J, \vec{x}^{\overline{J}}) > 1$, where $d_{\mathcal{A}}$ is the distance in $\mathcal{G}(\mathcal{A})$, the Gaifman graph of $\mathcal{A}$, and $\vec{x}^{\overline{J}}$ consists of the components of $\vec{x}$ whose indices are not in $\vec{x}^J$.

We first prove that this suffices to show that the duplicator wins. For this we need to establish $\vec{a}'\vec{c}\vec{x} \approx_0^{\mathcal{A}} \vec{b}'\vec{c}\vec{y}$, and furthermore, show that the mapping F induced by these two tuples preserves P. The latter is clear though as for any $v = F(u)$, either $u = v$ or $(u, v) \in H$, by construction, and thus P is preserved. To see that $\vec{a}'\vec{c}\vec{x} \approx_0^{\mathcal{A}} \vec{b}'\vec{c}\vec{y}$, notice that $\vec{a}'\vec{x}^J \approx_0 \vec{b}'\vec{y}^J$ by (3), and by (4) and the definition of $\vec{c}$, $d_{\mathcal{A}}(\vec{a}'\vec{x}^J, \vec{c}\vec{x}^{\overline{J}}) > 1$, and $d_{\mathcal{A}}(\vec{b}'\vec{y}^J, \vec{c}\vec{x}^{\overline{J}}) > 1$. Thus no σ-relation can have a tuple containing an element of $\vec{a}'\vec{x}^J$ and an element of $\vec{c}\vec{x}^{\overline{J}}$, or an element of $\vec{b}'\vec{y}^J$ and an element of $\vec{c}\vec{x}^{\overline{J}}$. This suffices to conclude that $\vec{a}'\vec{c}\vec{x} \approx_0^{\mathcal{A}} \vec{b}'\vec{c}\vec{y}$, and thus the duplicator wins the n-round game, provided (1)-(4) hold.

To prove that the duplicator can play as required, we show, by induction on the number of moves, that the duplicator can maintain four conditions. The play is somewhat similar to the one used in [33] for ordinary (not bijective) games. We shall classify all moves into two types, *type 1* moves and *type 2* moves. We use the following notation. Let $\vec{x} = (x_1, \ldots, x_i)$ and $\vec{y} = (y_1, \ldots, y_i)$ be points played on $\mathcal{A}_{\vec{a}}$ and $\mathcal{A}_{\vec{b}}$ after i rounds. That is, in the jth round, $j \leq i$, x_j is played in $\mathcal{A}_{\vec{a}}$, and $y_j = f_j(x_j)$ is the duplicator's response, where f_j is the bijection chosen by the duplicator for this round. By $\vec{x}^{\mathbf{1}}$ and $\vec{y}^{\mathbf{1}}$ we denote subtuples consisting of points played in type 1 moves, and by $\vec{x}^{\mathbf{2}}$ and $\vec{y}^{\mathbf{2}}$ we denote subtuples of points played in type 2 moves.

Let $d_i = 2^{m-i}$. The first two conditions are as follows.

1. If jth move is a type 2 move, then $x_j = y_j$; that is, $\vec{x}^{\mathbf{2}} = \vec{y}^{\mathbf{2}}$.
2. if jth move is a type 1 move, then $S_{d_i}^{\mathcal{A}}(x_j, y_j) \subseteq S_r^{\mathcal{A}}(\vec{a}'\vec{b}')$, and $(x_j, y_j) \in H$.

Let j be a type 1 move. Since $S_r^{\mathcal{A}}(\vec{a}) \cap S_r^{\mathcal{A}}(\vec{b}) = \emptyset$, if $x_j \in S_r^{\mathcal{A}}(\vec{a}')$, then $y_j \in S_r^{\mathcal{A}}(\vec{b}')$, and vice versa. We use the notation $\vec{x}_a^{\mathbf{1}}$ for the subtuple of $\vec{x}^{\mathbf{1}}$ whose components are in $S_r^{\mathcal{A}}(\vec{a}')$, and $\vec{x}_b^{\mathbf{1}}$ for the remaining components, that is, those in $S_r^{\mathcal{A}}(\vec{b}')$. We similarly define $\vec{y}_a^{\mathbf{1}}$ and $\vec{y}_b^{\mathbf{1}}$. Notice that $\vec{y}_b^{\mathbf{1}} = h(\vec{x}_a^{\mathbf{1}})$ and $\vec{x}_b^{\mathbf{1}} = h(\vec{y}_a^{\mathbf{1}})$.

The next two requirements are

3. $h : N_{d_i}^{\mathcal{A}}(\vec{a}'\vec{x}_a^{\mathbf{1}}\vec{y}_a^{\mathbf{1}}) \cong N_{d_i}^{\mathcal{A}}(\vec{b}'\vec{y}_b^{\mathbf{1}}\vec{x}_b^{\mathbf{1}})$ (that is, h is an isomorphism between these neighborhoods).
4. $d_{\mathcal{A}}(\vec{a}'\vec{x}_a^{\mathbf{1}}\vec{y}_a^{\mathbf{1}}, \vec{x}^{\mathbf{2}}) > d_i$ and $d_{\mathcal{A}}(\vec{b}'\vec{x}_b^{\mathbf{1}}\vec{y}_b^{\mathbf{1}}, \vec{x}^{\mathbf{2}}) > d_i$.

Proving these 4 conditions indeed suffices to conclude that the duplicator wins after n rounds, as then the conditions (1)-(4) are easily verified: we take J to be the set of type 1 moves. Note that permuting indices of moves has no effect on whether the resulting map is a partial isomorphism; thus we normally put subtuples of $\vec{x}$ and $\vec{y}$ in an order that keeps notation simple.

We prove, by induction on i, that the duplicator can maintain conditions 1–4. For the first move, the duplicator's bijection f is taken to be $f(x) = h(x)$ if $x \in S_{d_1}^{\mathcal{A}}(\vec{a})$, $f(x) = h^{-1}(x)$ if $x \in S_{d_1}^{\mathcal{A}}(\vec{b})$, and $f(x) = x$ in other cases. If the spoiler's move is in $S_{d_1}^{\mathcal{A}}(\vec{a}, \vec{b})$, then it is a type 1 move, otherwise it is a type 2 move. It is routine to verify that conditions 1-4 are satisfied.

For the inductive step, assume that i rounds have been played, and conditions 1–4 are satisfied. As

$d_i = 2d_{i+1}$, we define f_{i+1}, duplicator's bijection for the round $i+1$, as follows:

$$f_{i+1}(x) \quad = \quad \begin{cases} h(x) & \text{if } x \in S^{\mathcal{A}}_{d_{i+1}}(\vec{a}'\vec{x}^{\mathbf{1}}_a\vec{y}^{\mathbf{1}}_a) \\ h^{-1}(x) & \text{if } x \in S^{\mathcal{A}}_{d_{i+1}}(\vec{b}'\vec{y}^{\mathbf{1}}_b\vec{x}^{\mathbf{1}}_b) \\ x & \text{otherwise.} \end{cases}$$

If a move is made according to the first or the second clause (that is, if the spoiler plays in $S^{\mathcal{A}}_{d_{i+1}}(\vec{a}'\vec{b}'\vec{x}^{\mathbf{1}}\vec{y}^{\mathbf{1}})$), then the move is a type 1 move; otherwise it is a type 2 move. Let x_{i+1}, y_{i+1} be the $i+1$st move. It remains to verify conditions 1–4.

Condition 1 is obvious, and so is the second part of condition 2 ($(x_{i+1}, y_{i+1}) \in H$). If $x_{i+1} \in S^{\mathcal{A}}_{d_{i+1}}(\vec{a}'\vec{x}^{\mathbf{1}}_a\vec{y}^{\mathbf{1}}_a)$, then $S^{\mathcal{A}}_{d_{i+1}}(x_{i+1}) \subseteq S^{\mathcal{A}}_{d_i}(\vec{a}'\vec{x}^{\mathbf{1}}_a\vec{y}^{\mathbf{1}}_a)$, as $d_i = 2d_{i+1}$. Hence $S^{\mathcal{A}}_{d_{i+1}}(x_{i+1}) \subseteq S^{\mathcal{A}}_r(\vec{a}')$, and $S^{\mathcal{A}}_{d_{i+1}}(y_{i+1}) \subseteq S^{\mathcal{A}}_r(\vec{b}')$, proving the second part of condition 2. The proof for the case $x_{i+1} \in S^{\mathcal{A}}_{d_{i+1}}(\vec{b}'\vec{y}^{\mathbf{1}}_b\vec{x}^{\mathbf{1}}_b)$ is similar.

For a type 2 move, conditions 3 and 4 follow immediately from the definition of f_{i+1} and the hypothesis. Assume that $x_{i+1} \in S^{\mathcal{A}}_{d_{i+1}}(\vec{a}'\vec{x}^{\mathbf{1}}_a\vec{y}^{\mathbf{1}}_a)$ (the case of $x_{i+1} \in S^{\mathcal{A}}_{d_{i+1}}(\vec{b}'\vec{y}^{\mathbf{1}}_b\vec{x}^{\mathbf{1}}_b)$ is symmetric). As $h : N^{\mathcal{A}}_{2d_{i+1}}(\vec{a}'\vec{x}^{\mathbf{1}}_a\vec{y}^{\mathbf{1}}_a) \to N^{\mathcal{A}}_{2d_{i+1}}(\vec{b}'\vec{y}^{\mathbf{1}}_b\vec{x}^{\mathbf{1}}_b)$ is an isomorphism, we obtain that h maps $N^{\mathcal{A}}_{d_{i+1}}(\vec{a}'\vec{x}^{\mathbf{1}}_a\vec{y}^{\mathbf{1}}_a x_{i+1})$ isomorphically onto $N^{\mathcal{A}}_{2d_{i+1}}(\vec{b}'\vec{y}^{\mathbf{1}}_b\vec{x}^{\mathbf{1}}_b y_{i+1})$, and thus condition 3 holds. Furthermore, any component of $\vec{x}^{\mathbf{2}}$ is outside of $S^{\mathcal{A}}_{2d_{i+1}}(\vec{a}'\vec{x}^{\mathbf{1}}_a\vec{y}^{\mathbf{1}}_a)$ by the hypothesis, and hence outside $S^{\mathcal{A}}_{d_{i+1}}(x_{i+1})$. Thus, $d_{\mathcal{A}}(x_{i+1}, \vec{x}^{\mathbf{2}}) > d_{i+1}$. As $y_{i+1} = h(x_{i+1}) \in S^{\mathcal{A}}_{d_{i+1}}(\vec{b}'\vec{y}^{\mathbf{1}}_b\vec{x}^{\mathbf{1}}_b)$, and every element of $\vec{x}^{\mathbf{2}}$ is at the distance exceeding $2d_{i+1}$ from $\vec{b}'\vec{y}^{\mathbf{1}}_b\vec{x}^{\mathbf{1}}_b$, we obtain $d_{\mathcal{A}}(y_{i+1}, \vec{x}^{\mathbf{2}}) > d_{i+1}$, thus proving condition 4. This completes the proof that the duplicator can play to maintain conditions 1–4. The lemma is proved. □

We now put these two lemmas together to show

Theorem 5 *Let g be nondecreasing and not bounded by a constant, and let Q be an m-ary query in $(\mathcal{L}^*_{\infty\omega}(\mathbf{C}) + <^{-}_{g})_w$. Then Q is g-weakly semi-local.*

Proof: Let Q be definable by $\varphi(x_1, \ldots, x_m)$, where φ is a $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$ formula in the language of σ and an extra symbol S for the auxiliary preorder. Let $\mathcal{A}$ be a σ-structure, with $\vec{a}, \vec{b} \in A^m$ and $\vec{a} \leftrightsquigarrow^{\mathcal{A}}_{g,2^n} \vec{b}$, where $n = \mathsf{rk}(\varphi)$. Assume that φ is $<^{-}_{g}$-invariant on $\mathcal{A}$. Let P_0 be a preorder on A, such that $P_0 \in <^{-}_{g}$. Let $\vec{a} \in Q(\mathcal{A}) = \varphi[(\mathcal{A}, P_0)]$. Choose P to be the preorder given by Lemma 3. Due to the invariance of φ, $\vec{a} \in \varphi[(\mathcal{A}, P)]$; that is, $(\mathcal{A}, P) \models \varphi(\vec{a})$. By Lemmas 3 and 2, $(\mathcal{A}, P) \models \varphi(\vec{b})$, and again by invariance $(\mathcal{A}, P_0) \models \varphi(\vec{b})$. Thus, $\vec{b} \in \varphi[(\mathcal{A}, P_0)] = Q(\mathcal{A})$. This proves g-weak semi-locality of Q. □

If $g = id$, we obtain

Corollary 8 *Let $\precsim_2$ be the class of preorders in which every equivalence class has size at most 2. Then every query definable in $\mathcal{L}^*_{\infty\omega}(\mathbf{C}) + \precsim_2$ is weakly semi-local, and has the BNDP.* □

Proof of Theorem 3 Let Q be in $(\mathcal{L}^*_{\infty\omega}(\mathbf{C}) + <^{-}_{g})_w$. By Theorem 5, it is g-weakly semi-local. By Corollary 7, it has the BNDP. □

6 Conclusion

We have shown that queries definable in counting logics $\mathrm{FO}(\mathbf{C})$, $\mathrm{FO}(Q_{\mathbf{u}})$ and $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$, in the presence of relations from the class $<^-_g$ have the bounded number of degrees property. In other words, even extremely powerful counting logics in the presence of relations which are almost-everywhere linear orders have a very tame behavior. The situation changes drastically when $<^-_g$ is replaced by a linear order: for example, $\mathcal{L}^*_{\infty\omega}(\mathbf{C})+<$ expresses every query on ordered structures. A similar phenomenon is observed for other logics, most notably, $\mathrm{FO}(\mathbf{C})$ which captures uniform TC^0 on ordered structures.

The techniques of this paper cannot be straightforwardly extended to prove separation results in the ordered case. The logic $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$ is very powerful, as it expresses every property of natural numbers, and all other known counting extensions of FO can be embedded into it. We also relied on bijective games to prove the main result. However, bijective games characterize expressiveness of a logic which defines all queries on ordered finite structures. Thus, in the ordered case one cannot use the generic techniques from [13, 22, 23, 27] that apply to a variety of counting logics.

It was shown in [8] that if there is a proof of inexpressibility of some property in $\mathrm{FO}(\mathbf{C})+<$, then there must be a proof of that based on the counting games of [18]. The counting game is weaker than the bijective game; on the other hand, it does not have the inherent limitations of the latter in the ordered case. Thus, a possible way of proving a separation result may be to modify the locality techniques to work with the counting, rather than bijective, games.

Another approach would be to modify the ordered conjecture of [21] to include counting. Namely, such a modified conjecture would say that there is no unbounded class of ordered structures on which $\mathrm{FO}(\mathbf{C})$ captures polynomial time. One reason to consider this is that there are strong indications that for FO the conjecture holds [21]. With counting, however, one has to be careful: by considering the class of linear orders and adding unary quantifiers which test for polynomial time properties of cardinalities, one obtains a counting logic for which the conjecture fails. However, $\mathrm{FO}(\mathbf{C})$ has rather limited arithmetic, and perhaps an attempt to understand why it fails to capture polynomial time on various classes of structures may lead to a better understanding of its structural properties which are not shared by other counting logics.

Acknowledgements We thank Martin Grohe, Lauri Hella, Moshe Vardi, and Mihalis Yannakakis for their comments on earlier drafts of this paper.

References

[1] E. Allender. Circuit complexity before the dawn of the new millennium. In *Proc. 16th Conf. on Foundations of Software Technology and Theoretical Computer Science*, Springer LNCS vol. 1180, 1996, 1–18.

[2] D.A. Barrington, N. Immerman, H. Straubing. On uniformity within NC^1. *JCSS*, 41:274–306,1990.

[3] M. Benedikt, H.J. Keisler. Expressive power of unary counters. *ICDT'97*, pages 291–305.

[4] J. Cai, M. Fürer and N. Immerman. On optimal lower bound on the number of variables for graph identification. *Combinatorica*, 12 (1992), 389–410.

[5] G. Dong, L. Libkin and L. Wong. Local properties of query languages. *TCS*, to appear. Extended abstract in *ICDT'97*, pages 140–154.

[6] A. Durand, C. Lautemann and M. More. Counting results in weak formalisms. Manuscript, 1998.

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

[8] K. Etessami. Counting quantifiers, successor relations, and logarithmic space, *JCSS*, 54 (1997), 400–411.

[9] R. Fagin, L. Stockmeyer and M. Vardi, On monadic NP vs monadic co-NP, *Information and Computation*, 120 (1995), 78–92.

[10] H. Gaifman. On local and non-local properties, *Proceedings of the Herbrand Symposium, Logic Colloquium '81*, North Holland, 1982.

[11] M. Grohe and L. Hella. Personal communication, 1998.

[12] M. Grohe and T. Schwentick. Locality of order-invariant first-order formulas. In *MFCS'98*, pages 437–445.

[13] L. Hella. Logical hierarchies in PTIME. *Information and Computation*, 129 (1996), 1–19.

[14] L. Hella, L. Libkin and J. Nurmonen. Notions of locality and their logical characterizations over finite models. *Journal of Symbolic Logic*, to appear.

[15] L. Hella, L. Libkin , J. Nurmonen and L. Wong. Logics with aggregate operators. In *LICS'99*, to appear.

[16] N. Immerman. Relational queries computable in polynomial time. *Information and Control*, 68 (1986), 86–104.

[17] N. Immerman. Languages that capture complexity classes. *SIAM Journal of Computing* 16 (1987), 760–778.

[18] N. Immerman and E. Lander. Describing graphs: A first order approach to graph canonization. In *"Complexity Theory Retrospective"*, Springer Verlag, Berlin, 1990.

[19] Ph. Kolaitis and J. Väänänen. Generalized quantifiers and pebble games on finite structures. *Annals of Pure and Applied Logic*, 74 (1995), 23–75.

[20] Ph. Kolaitis, M. Vardi. Infinitary logic and 0-1 laws. *Inf.& Comput.*, 98 (1992), 258–294.

[21] Ph. Kolaitis and M. Vardi. Fixpoint logic vs. infinitary logic in finite-model theory. In *LICS'92*, pages 46–57.

[22] L. Libkin. On the forms of locality over finite models. In *LICS'97*, pages 204–215.

[23] L. Libkin. On counting logics and local properties. In *LICS'98*, to pages 501–512.

[24] L. Libkin. Logics with counting, auxiliary relations, and lower bounds for invariant queries. In *LICS'99*, to appear.

[25] L. Libkin, L. Wong. Query languages for bags and aggregate functions. *JCSS* 55 (1997), 241–272. Extended abstract in *PODS'94*.

[26] L. Libkin and L. Wong. Unary quantifiers, transitive closure, and relations of large degree. In *STACS'98*, Springer LNCS 1377, pages 183–193.

[27] J. Nurmonen. On winning strategies with unary quantifiers. *J. Logic and Computation*, 6 (1996), 779–798.

[28] M. Otto. *Bounded Variable Logics and Counting: A Study in Finite Models.* Springer Verlag, 1997.

[29] M. Otto. ϵ-Logic is more expressive than first-order on finite structures. Manuscript, 1997.

[30] I. Parberry and G. Schnitger. Parallel computation and threshold functions. *JCSS* 36 (1988), 278–302.

[31] A. Razborov and S. Rudich. Natural proofs. *JCSS* 55 (1997), 24–35.

[32] T. Schwentick. Graph connectivity, monadic NP and built-in relations of moderate degree. In *Proc. 22nd International Colloq. on Automata, Languages, and Programming*, Springer LNCS 944, 1995, pages 405–416.

[33] T. Schwentick. On winning Ehrenfeucht games and monadic NP. *Annals of Pure and Applied Logic*, 79 (1996), 61–92.

[34] M. Vardi. The complexity of relational query languages. In *STOC'82*, pages 137–146.