Logics with Counting and Local Properties

Leonid Libkin
Bell Laboratories

The expressive power of first-order logic over finite structures is limited in two ways: it lacks a recursion mechanism, and it cannot count. Overcoming the first limitation has been a subject of extensive study. A number of fixpoint logics have been introduced, and shown to be subsumed by an infinitary logic $\mathcal{L}^{\omega}_{\infty\omega}$. This logic is easier to analyze than fixpoint logics, and it still lacks counting power, as it has a 0-1 law. On the counting side, there is no analog of $\mathcal{L}^{\omega}_{\infty\omega}$. There are a number of logics with counting power, usually introduced via generalized quantifiers. Most known expressivity bounds are based on the fact that counting extensions of first-order logic preserve the locality properties.

This paper has three main goals. First, we introduce a new logic $\mathcal{L}^{*}_{\infty\omega}(\mathbf{C})$ that plays the same role for counting as $\mathcal{L}^{\omega}_{\infty\omega}$ does for recursion – it subsumes a number of extensions of first-order logic with counting, and has nice properties that make it easy to study. Second, we give a simple direct proof that $\mathcal{L}^{*}_{\infty\omega}(\mathbf{C})$ expresses only local properties: those that depend on the properties of small neighborhoods, but cannot grasp a structure as a whole. This is a general way of saying that a logic lacks a recursion mechanism. Third, we consider a finer analysis of locality of counting logics. In particular, we address the question of how local a logic is, that is, how big are those neighborhoods that local properties depend on. We get a uniform answer for a variety of logics between first-order and $\mathcal{L}^{*}_{\infty\omega}(\mathbf{C})$. This is done by introducing a new form of locality that captures the tightest condition that the duplicator needs to maintain in order to win a game. We also use this technique to give bounds on outputs of $\mathcal{L}^{*}_{\infty\omega}(\mathbf{C})$-definable queries.

Categories and Subject Descriptors: F.4.1 [**MATHEMATICAL LOGIC AND FORMAL LANGUAGES**]: Mathematical Logic

General Terms: Languages, Theory

Additional Key Words and Phrases: First-order logic, infinitary logic, counting, locality

An earlier version of this paper appeared in the *Proceedings of the IEEE Symposium on Logic in Computer Science (LICS'98)* (June 1998) pp. pages 501–512.
Address: Bell Laboratories, 600 Mountain Avenue, Rm. 2C-407, Murray Hill, NJ 07974, USA, Email: `libkin@research.bell-labs.com`.

1. INTRODUCTION

The expressive power of first-order logic (FO) on finite structures is rather limited. Two main limitations of first-order logic are its *inability to count* and the *lack of a recursion mechanism*. Since first-order logic over finite structures plays an important role in several areas of computer science (e.g., databases and complexity), various extensions have been proposed to deal with these shortcomings.

On the recursion side, a beautiful theory has been developed over the past decade. Various fixpoint extensions of first-order logic have been introduced, including least, inflationary and partial fixpoint logics, as well as transitive closure logics, cf. [1; 8]. Fixpoint logics can all be embedded into $\mathcal{L}^{\omega}_{\infty\omega}$, infinitary logic with finitely many variables, which is much easier to analyze. In particular, $\mathcal{L}^{\omega}_{\infty\omega}$ has a 0-1 law [23], which gives a uniform derivation of the 0-1 law for all fixpoint logics. It follows that $\mathcal{L}^{\omega}_{\infty\omega}$ cannot express most counting properties, such as parity of cardinality. The theory extends nicely to the ordered setting, where transitive closure and fixpoint logics capture familiar complexity classes such as L, NL, PTIME and PSPACE (cf. [8; 20]), and $\mathcal{L}^{\omega}_{\infty\omega}$ expresses every property of finite structures [6].

On the counting side, much less is known. Various extensions of first-order logic with counting exist, usually introduced by means of generalized quantifiers [22]. Examples include Härtig (equicardinality), Rescher (majority) quantifier, and counting quantifiers $\exists ix\varphi(x,\cdot)$, that assert the existence of at least i elements x that satisfy φ, see [9; 21]. We denote the extension with counting quantifiers $\exists i$ and simple arithmetic by $\mathrm{FO}(\mathbf{C})$. Alternatively, one can add counting terms [13; 14]. In [16], $\mathrm{FO}(Q_{\mathbf{u}})$, first-order logic extended with *all* unary generalized quantifiers, is considered. Precise definitions of these logics will be given in Section 2.

There are relatively few results on the expressive power of counting logics, and most were proved very recently. For example, [9] used the games of [21] to prove that there is an L-complete problem that is not definable in $\mathrm{FO}(\mathbf{C})$; this implies that connectivity of finite graphs is not definable in $\mathrm{FO}(\mathbf{C})$. In [19], nondefinability of connectivity is shown for $\mathrm{FO}(Q_{\mathbf{u}})$. More bounds were obtained in [24], which used the results of [27] to prove an analog of Gaifman's locality theorem [11] for those logics.

Currently, most bounds for extensions of FO with various counting quantifiers can be derived from its *local properties*, as shown in [17; 24; 27]; exceptions include the bound of [5], a result in [4] on counting the sizes of equivalence classes, and the hierarchy result in [14]. Locality of a logic gives us a general statement that it *lacks a nontrivial recursion mechanism*, much in the same way as 0-1 laws tell us that a logic cannot express nontrivial counting properties[1]. One way in which locality theorems are applied is the following. First, a form of locality based on Hanf's condition (see [10; 15]) is shown for a logic; this form is closely tied to a game-characterization of the logic. Then results of [17; 24] show that the logic also

[1]Of course, local logics can express complex properties of small neighborhoods, just as FO can test if the number of witnesses of a formula equals a given number.

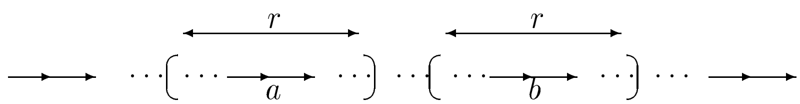

Fig. 1. A local formula cannot distinguish (a, b) from (b, a).

satisfies Gaifman's locality condition [11] and the bounded degree property [25], which are much easier to apply to prove expressivity bounds. However, no direct proofs of those conditions have been given so far for any of the extensions of FO.

The basic idea of locality is shown in Figure 1. A formula, say $\varphi(x, y)$ is local, if it can only "see" from x and y as far as a neighborhood of radius r, where r only depends on φ. In the graph of a successor relation, it means that pairs (a, b) and (b, a) are indistinguishable by $\varphi(x, y)$, if the successor relation is long enough, and a, b are far away from the endpoints and each other – this is because no point can be "seen" from both a and b, if we can only see up to the distance r. In particular, transitive closure, that distinguishes (a, b) from (b, a), is not definable in a local logic. In general, nontrivial recursive computation gives one a means of verifying global properties of structures, and most properties requiring such a form of computation can easily be shown to violate one of the forms of locality.

We now describe the three main themes of the paper, and outline the results.

$\angle$ *A General Framework for Counting*. While there are a number of counting extensions of first-order logic, we still lack a unifying framework for adding counting to FO. For example, the extension with counting quantifiers [9; 21] puts limits on available arithmetic, while the extension with unary quantifiers does not permit free numerical variables; as a result, expressing some simple properties becomes a nontrivial task, with the resulting formulae being unnecessarily awkward. It appears that we need a general framework that subsumes all these logics, and is at the same time easy to study. We introduce such a logic, denoted $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$, in Section 3. It is obtained by first adding counting terms and quantifiers to the infinitary logic $\mathcal{L}_{\infty\omega}$ over two-sorted structures (the second sort being interpreted as $\mathbb{N}$), and then restricting it to formulae of finite rank. The idea of putting the set of natural numbers "on the side" is influenced by metafinite model theory of [12]. Similar extensions exist in the literature [13; 14; 28], but they restrict the logic by means of the number of variables, which still permits fixpoint computation. In contrast, following [16; 17], we restrict the logic by requiring that the rank of a formula be finite (where the rank is defined as quantifier rank, except that it does not take into account quantifiers over $\mathbb{N}$), thus putting no limits at all on the available arithmetic. We give a simplified version of $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$, called $\mathcal{L}^\circ_{\infty\omega}(\mathbf{C})$; it is obtained by disallowing counting terms. This makes the logic easier to analyze, and we prove that no power is lost due to this restriction.

$\angle$ *Proving Locality*. How does one prove that formulae in a counting logic (e.g., $\mathrm{FO}(\mathbf{C})$, $\mathrm{FO}(Q_{\mathbf{u}})$, $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$) only express local properties, as shown in Figure 1? Currently, with the exception of FO, such results are established via Hanf's criterion [10; 15] that relates the number of isomorphism types of small neighborhoods in two structures. This criterion is closely tied to a game characterization of a logic, and may not work if such a characterization does not exist. Also, one needs to adjust the implication results for two-sorted logics. Here, we show that locality of $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$ and other counting logics can be proved *directly*, without this unnecessary detour.

∠ *Theory of Local Properties*. There are a number of issues in the theory of local properties that one has to deal with once the locality of a logic is proved. One is the question about the radius of a neighborhood that determines the truth value of a formula. For the example in Figure 1, one may ask how r depends on φ. It is known how to find FO formulae with r being $O(2^{\mathsf{qr}(\varphi)})$, where $\mathsf{qr}(\cdot)$ is the quantifier rank of a formula. Here we show that $O(2^{\mathsf{qr}(\varphi)})$ is also the upper bound for many counting logics, including $\mathrm{FO}(\mathbf{C}), \mathrm{FO}(Q_{\mathbf{u}})$ and $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$. To prove this, we introduce a new form of locality, called the *separation property*, in Section 4. It corresponds, intuitively, to the tightest condition that the duplicator needs to maintain in an Ehrenfeucht-Fraïssé game (or a bijective game [16]) in order to win. Based on the separation property, we calculate, in Section 5, the *exact* value of the maximum radius for all the formulae of rank n. We do it for two forms of locality, based on Hanf's and Gaifman's conditions, and we show that in both cases the maximum radii are the same for all the counting logics listed above.

In Section 6, we consider open local formulae as queries that map finite structures to finite structures. Extending a result from [7], we prove a bound on the number of different degrees realized in the output of a local query, and apply it to counting logics, thereby connecting this measure of "complexity" of the output with the syntactic parameters of a query. Concluding remarks are given in Section 7.

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 $p_i > 0$. We write σ_n for σ extended with n new constant symbols. 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 $\mathrm{STRUCT}[\sigma]$. 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 and the carrier of $\mathcal{B}$ is denoted by B. In case of two-sorted logics, A and B will refer to the finite universe of the non-numerical sort.

We abbreviate first-order logic by FO. *FO with counting*, denoted by $\mathrm{FO}(\mathbf{C})$, is a two-sorted logic, with the second sort being the sort 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 there 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 [9], 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 a semicolon: $\varphi(\vec{x}; \vec{\jmath})$.

We next define extensions of FO by unary generalized quantifiers. Let σ_k^{unary} be a signature of k unary relation symbols, and let $\mathcal{K}$ be a class of σ_k^{unary}-structures which is closed under isomorphisms. Then FO($Q_{\mathcal{K}}$) extends the set of formulae of FO with the following additional rule: if $\psi_1(x_1, \vec{y}_1), \ldots, \psi_k(x_k, \vec{y}_k)$ are formulae, then $Q_{\mathcal{K}} x_1 \ldots x_k.(\psi_1(x_1, \vec{y}_1), \ldots, \psi_k(x_k, \vec{y}_k))$ is a formula. Here $Q_{\mathcal{K}}$ binds x_i in the ith formula, for each $i = 1, \ldots, k$. A free occurrence of a variable y in $\psi_i(x_i, \vec{y}_i)$ remains free in this new formula unless $y = x_i$. The semantics is defined as follows: $\mathcal{A} \models Q_{\mathcal{K}} x_1 \ldots x_k.(\psi_1(x_1, \vec{a}_1), \ldots, \psi_k(x_k, \vec{a}_k))$ iff

$$(A, \psi_1[\mathcal{A}, \vec{a}_1], \ldots, \psi_k[\mathcal{A}, \vec{a}_k]) \in \mathcal{K},$$

where $\psi_i[\mathcal{A}, \vec{a}_i] = \{a \in A \mid \mathcal{A} \models \psi_i(a, \vec{a}_i)\}$. In this definition, $\vec{a}_i$ is a tuple of parameters that gives the interpretation for those free variables of $\psi_i(x_i, \vec{y}_i)$ which are not equal to x_i. Examples include the usual $\exists$ and $\forall$, as well as Rescher and Härtig quantifiers. We use the notation FO($Q_{\mathbf{u}}$) for FO extended with *all* unary quantifiers.

We denote the infinitary logic by $\mathcal{L}_{\infty\omega}$; it extends FO by allowing infinite conjunctions $\bigwedge$ and disjunctions $\bigvee$. The class of $\mathcal{L}_{\infty\omega}$ formulae that only use variables from a fixed set $\{x_1, \ldots, x_k\}$ is denoted by $\mathcal{L}^k_{\infty\omega}$, and $\mathcal{L}^\omega_{\infty\omega}$ is the infinitary logic with finitely many variables: $\mathcal{L}^\omega_{\infty\omega} = \bigcup_{k<\omega} \mathcal{L}^k_{\infty\omega}$.

The *quantifier rank* of a formula φ, denoted by $\mathsf{qr}(\varphi)$ is the depth of quantifier nesting (including $\forall, \exists, \exists i, Q_{\mathcal{K}}$) in φ. For FO($\mathbf{C}$), we do not count quantifiers over the numerical domain.

Games. We review some results on game characterizations of logics. The *n-round Ehrenfeucht-Fraïssé game* (cf. [8]), is played by two players, called the spoiler and the duplicator, on two structures $\mathcal{A}, \mathcal{B} \in \mathrm{STRUCT}[\sigma]$. In each round i, the spoiler selects either a point $a_i \in A$, or $b_i \in B$, and the duplicator responds by selecting $b_i \in B$, or $a_i \in A$, respectively. The duplicator wins after n rounds if the relation $\{(a_i, b_i) \mid 1 \le i \le n\}$ is a partial isomorphism $\mathcal{A} \to \mathcal{B}$; otherwise the spoiler wins. If the duplicator has a winning strategy in the n-round game on $\mathcal{A}$ and $\mathcal{B}$, we write $\mathcal{A} \equiv_n \mathcal{B}$. It is well known (cf. [8]) that $\mathcal{A} \equiv_n \mathcal{B}$ iff $\mathcal{A}$ and $\mathcal{B}$ agree on all FO sentences of quantifier rank up to n.

A stronger version of the game, called the *bijective Ehrenfeucht-Fraïssé game*, was introduced in [16]. Again, the spoiler and the duplicator play 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 $f_i : A \to B$, and the spoiler selects a point $a_i \in A$ (if $card(A) \neq card(B)$, then the spoiler wins). The winning condition is the same: if after the last round the relation $\{(a_i, f_i(a_i)) \mid 1 \le i \le n\}$ is a partial isomorphism

$\mathcal{A} \to \mathcal{B}$, then the duplicator wins; otherwise the spoiler wins. If the duplicator has a winning strategy in the n-round bijective game on $\mathcal{A}$ and $\mathcal{B}$, we write $\mathcal{A} \equiv_n^{bij} \mathcal{B}$. Bijective games describe elementary equivalence in $\mathrm{FO}(Q_{\mathbf{u}})$ and $\mathrm{FO}(\mathbf{C})$ as follows.

FACT 2.1 ([9; 16]). *Let $\mathcal{A}, \mathcal{B} \in \mathrm{STRUCT}[\sigma]$. If $\mathcal{A} \equiv_n^{bij} \mathcal{B}$, then $\mathcal{A}$ and $\mathcal{B}$ agree on all sentences of* $\mathrm{FO}(Q_{\mathbf{u}})$ *(or* $\mathrm{FO}(\mathbf{C})$*) of quantifier rank up to n.* □

We write $(\mathcal{A}, \vec{a}) \equiv_n (\mathcal{B}, \vec{b})$ (or $(\mathcal{A}, \vec{a}) \equiv_n^{bij} (\mathcal{B}, \vec{b})$) if the duplicator has a winning strategy in the n-round (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 $\mathcal{B} \models \varphi(\vec{b})$.

Locality. Given a structure $\mathcal{A}$, its *Gaifman graph* [8; 11; 10] $\mathcal{G}(\mathcal{A})$ is defined as $\langle A, E \rangle$ where (a, b) is in E iff $a = b$ or 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 $\{b \in A \mid d(\vec{a}, b) \le r\}$. Its *$r$-neighborhood* $N_r^{\mathcal{A}}(\vec{a})$ is defined as a σ_n structure

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

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 $N_r(\vec{a})$.

Given a tuple $\vec{a}$ of elements of A, and $d \ge 0$, by $ntp_d^{\mathcal{A}}(\vec{a})$ we denote the isomorphism type of $N_d^{\mathcal{A}}(\vec{a})$. For example, $ntp_d^{\mathcal{A}}(\vec{a}) = ntp_d^{\mathcal{B}}(\vec{b})$ means that there is an isomorphism $N_d^{\mathcal{A}}(\vec{a}) \to N_d^{\mathcal{B}}(\vec{b})$ that sends $\vec{a}$ to $\vec{b}$; in this case we will also write $\vec{a} \approx_d^{\mathcal{A},\mathcal{B}} \vec{b}$. If $\mathcal{A} = \mathcal{B}$, we write $\vec{a} \approx_d^{\mathcal{A}} \vec{b}$. 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 2.2 (GAIFMAN-LOCALITY). (cf. [24]) *A formula $\varphi(\vec{x}; \vec{\jmath})$ in a two-sorted logic is called* Gaifman-local *if there exists a number $r \ge 0$ such that, for any structure $\mathcal{A}$ and any $\vec{a}, \vec{b}$ over A,*

$$\vec{a} \approx_r^{\mathcal{A}} \vec{b} \quad \textit{implies} \quad \mathcal{A} \models \varphi(\vec{a}; \vec{\imath}) \;\; \textit{iff} \;\; \mathcal{A} \models \varphi(\vec{b}; \vec{\imath})$$

for all $\vec{\imath} \subset \mathbb{N}$. The minimum such r is called the locality rank *of φ, and is denoted by $\mathsf{lr}(\varphi)$.* □

FACT 2.3. *a)* (see [11]) *Every FO formula $\varphi(\vec{x})$ is Gaifman-local, and* $\mathsf{lr}(\varphi) \le (7^{\mathsf{qr}(\varphi)} - 1)/2$.

b) (see [24; 17]) *Every* $\mathrm{FO}(Q_{\mathbf{u}})$ *or* $\mathrm{FO}(\mathbf{C})$ *formula $\varphi(x_1, \ldots, x_n)$ (without free second-sort variables) is Gaifman-local, and* $\mathsf{lr}(\varphi) \le 3^{\mathsf{qr}(\varphi)+n} + 1$. □

Remark: [26] shows that in a), $(7^{\mathsf{qr}(\varphi)} - 1)/2$ can be replaced by $3 \cdot 4^{\mathsf{qr}(\varphi)}$.

For $\mathcal{A}, \mathcal{B} \in \mathrm{STRUCT}[\sigma]$, we write $\mathcal{A} \leftrightarrows_d \mathcal{B}$ if there exists a bijection $f : A \to B$ such that $ntp_d^{\mathcal{A}}(a) = ntp_d^{\mathcal{B}}(f(a))$ for every $a \in A$. That is, every isomorphism type of a d-neighborhood of a point has equally many elements realizing it in $\mathcal{A}$ and $\mathcal{B}$. We write $(\mathcal{A}, \vec{a}) \leftrightarrows_d (\mathcal{B}, \vec{b})$ if there is a bijection $f : A \to B$ such that $ntp_d^{\mathcal{A}}(\vec{a}c) = ntp_d^{\mathcal{B}}(\vec{b}f(c))$ for every $c \in A$.

Definition 2.4 (Hanf-locality). *a)* (see [15; 10; 24]) *A sentence* Φ *is called* Hanf-local *if there exist a number* $d \geq 0$ *such that* $\mathcal{A}$ *and* $\mathcal{B}$ *agree on* Φ *whenever* $\mathcal{A} \leftrightarrows_d \mathcal{B}$. *The minimum d for which this holds is called* Hanf locality rank *of* Φ, *and is denoted by* $\mathsf{hlr}(\Phi)$.

b) (see [17]) *A formula* $\varphi(\vec{x})$ *is called* Hanf-local *if there exist a number* $d \geq 0$ *such that for any* $\mathcal{A}, \mathcal{B}$, *and* $\vec{a}, \vec{b}$ *over* A *and* B, *resp., of the same length as* $\vec{x}$, $(\mathcal{A}, \vec{a}) \leftrightarrows_d (\mathcal{B}, \vec{b})$ *implies* $\mathcal{A} \models \varphi(\vec{a})$ *iff* $\mathcal{B} \models \varphi(\vec{b})$.

It can be seen [10] that $\mathcal{A} \leftrightarrows_d \mathcal{B}$ implies $\mathcal{A} \leftrightarrows_r \mathcal{B}$ for $r \leq d$; in particular, if $\mathcal{A} \leftrightarrows_d \mathcal{B}$, then $|A| = |B|$.

Fact 2.5. *a)* (see [10]) *If* $\mathcal{A} \leftrightarrows_{3^n} \mathcal{B}$, *then* $\mathcal{A} \equiv_n \mathcal{B}$. *In particular,* $\mathcal{A}$ *and* $\mathcal{B}$ *agree on all FO sentences of quantifier rank up to* n.

b) (see [27]; bound from [17]) *Let* $n > 0$. *Then* $\mathcal{A} \leftrightarrows_{(3^{n-1}-1)/2} \mathcal{B}$ *implies* $\mathcal{A} \equiv_n^{bij} \mathcal{B}$.

c) (see [17; 24]) *Every Hanf-local formula (without free second-sort variables, if one deals with a two-sorted logic) is Gaifman-local.* □

Next, we review results on outputs of local queries. With each formula $\varphi(x_1, \ldots, x_n)$ in the language σ, we associate a query that maps $\mathcal{A} \in \mathrm{STRUCT}[\sigma]$ to $\varphi[\mathcal{A}] = \{\vec{a} \in A^n \mid \mathcal{A} \models \varphi(\vec{a})\}$.

If $\mathcal{A} \in \mathrm{STRUCT}[\sigma]$, and R_i is of arity p_i, then $degree_j(R_i^{\mathcal{A}}, a)$ 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 $deg\_set(\mathcal{A})$ we mean the set of all degrees realized in $\mathcal{A}$, and $deg(\mathcal{A})$ stands for the cardinality of $deg\_set(\mathcal{A})$. We use the notation $\mathrm{STRUCT}_k[\sigma]$ for $\{\mathcal{A} \in \mathrm{STRUCT}[\sigma] \mid deg\_set(\mathcal{A}) \subseteq \{0, 1, \ldots, k\}\}$.

Definition 2.6 (Bounded Degree Property). (see [25; 7; 24]) *A query* q, *that is, a function that maps* $\mathcal{A} \in \mathrm{STRUCT}[\sigma]$ *to an* m*-ary relation on* A, $m \geq 1$, *is said to have the* bounded degree property, *or* BDP, *if there exists a function* $f_q : \mathbb{N} \to \mathbb{N}$ *such that* $deg(q(\mathcal{A})) \leq f_q(k)$ *for every* $\mathcal{A} \in \mathrm{STRUCT}_k[\sigma]$. □

Note that $deg(q(\mathcal{A}))$ should formally be written as $deg(\langle A, q(\mathcal{A})\rangle)$, but we shall use the abbreviated version as it will never lead to confusion.

The intuition behind this definition is that if $\mathcal{A}$ locally looks simple, then $q(\mathcal{A})$ has a simple structure as well. The BDP is very easy to use for proving expressivity bounds [25]. It is known [7] that every Gaifman-local query has the BDP. A simple proof of this can be given for formulae $\varphi(x, y)$, whose outputs $\varphi[\mathcal{A}]$ are directed graphs. Let $\#\mathsf{ntp}_d(\mathcal{A})$ stand for $|\{ntp_d^{\mathcal{A}}(a) \mid a \in A\}|$ – the number of different isomorphism types of d-neighborhoods of points realized in $\mathcal{A}$. The following result from [7] implies the BDP for queries whose outputs are graphs.

FACT 2.7. *Let $\varphi(x,y)$ be Gaifman-local and let $d = 3 \cdot \mathsf{lr}(\varphi) + 1$. Then, for any structure $\mathcal{A}$, $deg(\varphi[\mathcal{A}]) \leq 2 \cdot \#\mathsf{ntp}_d(\mathcal{A})$.* $\square$

3. $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$ AND ITS LOCALITY

The goal of this section is to define the logic $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$, which is to counting extensions of FO what $\mathcal{L}^\omega_{\infty\omega}$ is to fixpoint extensions of FO. We then define a simpler version of this logic, $\mathcal{L}^\circ_{\infty\omega}(\mathbf{C})$, and show that no expressiveness is lost.

First, define $\mathcal{L}_{\infty\omega}(\mathbf{C})$, 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 the following:

Counting terms: If φ is a formula and $\vec{x}$ is a vector 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 $\vec{a}$ over the finite first-sort universe that satisfy φ. That is, given a structure $\mathcal{A}$, a formula $\varphi(\vec{x},\vec{y};\vec{\jmath})$ and vectors $\vec{b} \subseteq A$ 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 $\{\vec{a} \subseteq A \mid \mathcal{A} \models \varphi(\vec{a},\vec{b};\vec{\jmath}_0)\}$.

Counting quantifiers: If φ is a formula and $i \in \mathbb{N}$, then $\exists ix\ \varphi$ is a formula. Its free variables are those in φ except x.

The logic $\mathcal{L}_{\infty\omega}(\mathbf{C})$ is enormously powerful: it can define *every* property of finite models, and *every* predicate or function on $\mathbb{N}$. The definition is also redundant: for example, $\exists ix\ \varphi$ can be replaced by $\#x.\varphi \geq i$. However, we need counting quantifiers separately, as will become clear soon.

Next, we restrict the logic by defining the *rank* of a formula, $\mathsf{rk}(\varphi)$. Its definition is similar to that of quantifier rank, and for FO, FO($\mathbf{C}$) and FO($Q_{\mathbf{u}}$) formulae, $\mathsf{rk}(\varphi) = \mathsf{qr}(\varphi)$. The difference is that we disregard quantification over $\mathbb{N}$, thus allowing arbitrary nesting of such quantifiers. For a variable or a constant term, the rank is 0. For other formulae and terms, it is defined as follows.

—If φ is atomic, then $\mathsf{rk}(\varphi)$ is the maximum rank of a term that occurs in φ.
—$\mathsf{rk}(\#\vec{x}.\varphi) = \mathsf{rk}(\varphi) + |\vec{x}|$.
—$\mathsf{rk}(\bigvee \varphi_i) = \mathsf{rk}(\bigwedge \varphi_i) = \sup_i \mathsf{rk}(\varphi_i)$.
—$\mathsf{rk}(\neg\varphi) = \mathsf{rk}(\varphi)$.
—$\mathsf{rk}(\forall x\ \varphi) = \mathsf{rk}(\exists x\ \varphi) = \mathsf{rk}(\exists ix\ \varphi) = \mathsf{rk}(\varphi) + 1$.
—$\mathsf{rk}(\forall i\ \varphi) = \mathsf{rk}(\exists i\ \varphi) = \mathsf{rk}(\varphi)$.

For FO, FO($\mathbf{C}$) and FO($Q_{\mathbf{u}}$), this coincides with the definition of quantifier rank.

Definition 3.1. *$\mathcal{L}^*_{\infty\omega}(\mathbf{C})$ is defined as the restriction of $\mathcal{L}_{\infty\omega}(\mathbf{C})$ to formulae and terms that have finite rank.* □

The following is immediate:

Lemma 3.2. *1) $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$ is closed under finitary Boolean connectives and all quantification.*
2) Every predicate on $\mathbb{N} \times \ldots \times \mathbb{N}$ is definable by an $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$ formula of rank 0.

Proof. 1) is obvious; for 2), observe that every predicate $P \subseteq \mathbb{N}^k$ is definable by

$$\varphi_P(i_1, \ldots, i_k) \;=\; \bigvee_{(n_1,\ldots,n_k)\in P} \; \bigwedge_{j=1}^{k} (i_j = n_j) \quad \square$$

Thus, we assume that $+, *, -, \leq$, and in fact *every* predicate on natural numbers is available. To give an example, we can express properties like: there is a node in the graph whose indegree i and outdegree j satisfy $p_i^2 > p_j$ where p_i stands for the ith prime. This is done by $\exists x \exists i \exists j.(i = \#y.E(y,x)) \wedge (j = \#y.E(x,y)) \wedge P(i,j)$ where P is the predicate on $\mathbb{N}$ for the property $p_i^2 > p_j$.

Known expansions of FO with counting properties are contained in $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$:

Proposition 3.3. *For every* FO, FO($\mathbf{C}$), *or* FO($Q_{\mathbf{u}}$) *formula, there exists an equivalent $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$ formula of the same rank.*

Proof. The proof is trivial for FO and FO($\mathbf{C}$). For FO($Q_{\mathbf{u}}$), assume we have a formula

$$\psi(\vec{y}_1, \ldots, \vec{y}_k) \;\equiv\; Q_{\mathcal{K}} x_1 \ldots x_k.(\psi_1(x_1, \vec{y}_1), \ldots, \psi_k(x_k, \vec{y}_k))$$

where each ψ_i is equivalent to a $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$ formula.

Let $\vec{y}$ be the concatenation of all $\vec{y}_i$, $i = 1, \ldots, k$. Consider 2^k mappings $f : \{1, \ldots, k\} \to \{+, -\}$, and define for each such mapping f,

$$\alpha_f(x, \vec{y}) \;\equiv\; \bigwedge_{i=1}^{k} \psi_i^{f(i)}(x, \vec{y}_i),$$

where $\beta^+ \equiv \beta$ and $\beta^- \equiv \neg\beta$ for any formula β. Note that if $(A, U_1, \ldots, U_k) \in \mathcal{K}$, and $(B, V_1, \ldots, V_k)$ is such that $|A| = |B|$, and for every Boolean combination of the U_is, the corresponding Boolean combination of the V_is has the same cardinality, then $(B, V_1, \ldots, V_k) \in \mathcal{K}$. Thus, ψ is equivalent to the formula

$$P_{\mathcal{K}}(\#z.(z = z), \#z.\alpha_{f_1}(z, \vec{y}), \ldots, \#z.\alpha_{f_{2^k}}(z, \vec{y})),$$

where $f_1, \ldots, f_{2^k}$ is some enumeration of functions $f : \{1, \ldots, k\} \to \{+, -\}$, and $P_{\mathcal{K}}(m, n_1, \ldots, n_{2^k})$ holds iff all the structures $(A, U_1, \ldots, U_k)$ with $|A| = m$ and $|U_1^{f_i(1)} \cap \ldots \cap U_k^{f_i(k)}| = m_i$, $i = 1, \ldots, 2^k$, belong to $\mathcal{K}$. Here we use U^+ to denote U and U^- to denote $A - U$. By Lemma 3.2, the formula above is a $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$ formula, and its rank is $\max\{\mathsf{rk}(\psi_1), \ldots, \mathsf{rk}(\psi_k)\} + 1 = \mathsf{rk}(\psi)$. □

It can also be shown that counting logics defined in [4] are embeddable into $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$.

Definition 3.4. *The logic $\mathcal{L}^\circ_{\infty\omega}(\mathbf{C})$ is defined as $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$ where counting terms $\#\vec{x}.\varphi$ are not allowed.* □

On the surface, $\mathcal{L}^\circ_{\infty\omega}(\mathbf{C})$ is a lot simpler than $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$, mainly because counting terms for vectors, $\#\vec{x}.\varphi$, are very convenient for defining complex counting properties. Also, $\mathcal{L}^\circ_{\infty\omega}(\mathbf{C})$ permits easier proofs by induction on the formulae; usually counting terms make such proofs much more complex. But it turns out that the power of $\mathcal{L}^\circ_{\infty\omega}(\mathbf{C})$ and $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$ is identical. This is somewhat reminiscent of a result in [22] that shows how unary generalized quantifiers can be modeled by counting quantifiers in $\mathcal{L}^k_{\infty\omega}$.

Theorem 3.5. *There is a translation $\varphi \to \varphi^\circ$ of $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$ formulae into $\mathcal{L}^\circ_{\infty\omega}(\mathbf{C})$ formulae such that φ and φ° are equivalent and $\mathsf{rk}(\varphi) = \mathsf{rk}(\varphi^\circ)$.*

Proof. Assume without loss of generality that every occurrence of $\#\vec{x}.\varphi$ is of the form $\#\vec{x}.\varphi = \#\vec{y}.\psi$ or $\#\vec{x}.\varphi = i$ (if $\#\vec{x}.\varphi$ occurs inside an arithmetic predicate P, we replace P by its definition, as in Lemma 3.2). Since $\#\vec{x}.\varphi = \#\vec{y}.\psi$ is equivalent to $\exists i.\#\vec{x}.\varphi = i \wedge \#\vec{y}.\psi = i$, whose rank is the same as the rank of $\#\vec{x}.\varphi = \#\vec{y}.\psi$, we may assume that all occurrences of #-terms are of the form $\#\vec{x}.\varphi = i$.

The proof is now by induction on the formula. The only nontrivial case is $\psi(\vec{y};\vec{\jmath}) \equiv (\#\vec{x}.\varphi(\vec{x},\vec{y};\vec{\jmath}) = i)$. Throughout this proof, we assume that i is in $\vec{\jmath}$. By the hypothesis, there exists an $\mathcal{L}^\circ_{\infty\omega}(\mathbf{C})$ formula φ° which is equivalent to φ and has the same rank. We must now produce an $\mathcal{L}^\circ_{\infty\omega}(\mathbf{C})$ formula ψ° equivalent to ψ such that $\mathsf{rk}(\psi^\circ) = \mathsf{rk}(\varphi) + |\vec{x}|$. This will follow from the lemma below.

Lemma 3.6. *Let $\varphi(\vec{x},\vec{y};\vec{\jmath})$ be a $\mathcal{L}^\circ_{\infty\omega}(\mathbf{C})$ formula. Then there exists a $\mathcal{L}^\circ_{\infty\omega}(\mathbf{C})$ formula $\gamma(\vec{y};\vec{\jmath})$ of rank $\mathsf{rk}(\varphi) + |\vec{x}|$ such that γ is equivalent to $\#\vec{x}.\varphi = i$.*

Proof of the lemma is by induction on $|\vec{x}|$. If $\vec{x}$ has a single component x, $\gamma(\vec{y};\vec{\jmath})$ is defined as

$$\exists l(l = i \wedge \exists! lx.\varphi(x,\vec{y};\vec{\jmath}))$$

which has rank $\mathsf{rk}(\varphi) + 1$. Recall that $\exists! lx.\varphi$ is an abbreviation for $(\exists lx.\varphi) \wedge \forall k(\exists kx.\varphi \to k \leq l)$.

We next assume that $\vec{x} = \vec{z}x_0$. By the hypothesis, there is an $\mathcal{L}^\circ_{\infty\omega}(\mathbf{C})$ formula $\alpha(x_0,\vec{y};\vec{\jmath},l)$ equivalent to $(l = \#\vec{z}.\varphi(\vec{z},x_0,\vec{y};\vec{\jmath}))$ such that $\mathsf{rk}(\alpha) = \mathsf{rk}(\varphi) + |\vec{z}|$. Define

$$\beta(\vec{y};\vec{\jmath},k,l) \ \equiv\ \exists! kx_0.\alpha(x_0,\vec{y};\vec{\jmath},l)$$

Then $\mathsf{rk}(\beta) = \mathsf{rk}(\alpha) + 1 = \mathsf{rk}(\varphi) + |\vec{x}|$. The formula $\beta(\cdots,k,l)$ holds iff there exists exactly k elements x_0 such that the number of vectors $\vec{x}$ with x_0 in the last position that satisfy $\varphi(\vec{x},\cdots)$ is precisely l. Note that if $\beta(\cdots,k,l)$ and $\beta(\cdots,k',l)$ hold, then k' must equal k.

Thus, to check if $\#\vec{x}.\varphi = i$, one must check if

$$\sum_{\beta(\cdots,k,l)\ \text{holds}} (k \cdot l) = i.$$

This is done as follows. Let $\gamma_p(\vec{y};\vec{\jmath})$ be defined as:

$$\exists i_1 \ldots i_p \exists j_1 \ldots j_p. \left(\begin{array}{l} \quad \bigwedge_{s=1}^{p} \beta(\vec{y};\vec{\jmath}, i_s, j_s) \\ \wedge\ \forall i, j. \beta(\vec{y};\vec{\jmath}, i, j) \rightarrow \bigvee_{s=1}^{p} (i = i_s \wedge j = j_s) \\ \wedge\ \bigwedge_{s \neq s'} (\neg(i_s = i_{s'}) \vee \neg(j_s = j_{s'})) \\ \wedge\ i_1 \cdot j_1 + \ldots + i_p \cdot j_p = i \end{array} \right)$$

That is, γ_p says that there are precisely p pairs (i_s, j_s) that satisfy $\beta(\cdots, k, l)$, and $\sum_{s=1}^{p} i_s \cdot j_s = i$. When $p = 0$, we define $\gamma_p(\vec{y};\vec{\jmath})$ as $(i = 0) \wedge \forall i', j'. \neg\beta(\vec{y};\vec{\jmath}, i', j')$. We can see that $\mathsf{rk}(\gamma_p) = \mathsf{rk}(\beta)$. We finally define

$$\gamma(\vec{y};\vec{\jmath}) \quad \equiv \quad \bigvee_{p \in \mathbb{N}} \gamma_p(\vec{y};\vec{\jmath})$$

It follows that γ is a $\mathcal{L}^{\circ}_{\infty\omega}(\mathbf{C})$ formula of rank that is equal to $\mathsf{rk}(\beta)$, and hence to $\mathsf{rk}(\varphi)+ \mid \vec{x} \mid$, and that γ is equivalent to $\#\vec{x}.\varphi = i$. This completes the proof of the lemma and the theorem. $\square$

Thus, for every FO, or FO($\mathbf{C}$), or FO($Q_{\mathbf{u}}$) formula, there exists an equivalent $\mathcal{L}^{\circ}_{\infty\omega}(\mathbf{C})$ formula of the same rank.

One can use the counting expressive power of $\mathcal{L}^{*}_{\infty\omega}(\mathbf{C})$ if one needs to show definability of some properties. At the same time, one can use $\mathcal{L}^{\circ}_{\infty\omega}(\mathbf{C})$ for proving expressivity bounds.

3.1 Definability over ordered structures

By ordered structures, we mean that one of the relations on the finite model is $<$ interpreted as a linear ordering. With $<$, one can say that a given element of A is first, second, etc., element of A. Then unlimited counting power allows us to code finite structures with numbers, and we can easily show:

PROPOSITION 3.7. *Every property of finite ordered structures is definable in* $\mathcal{L}^{*}_{\infty\omega}(\mathbf{C})$.

Proof. We sketch the proof for sentences in the language of graphs. Let $\mathcal{P}$ be a class of ordered graphs. We assume without loss of generality that the set of nodes of each such graph is a set of the form $\{0, \ldots, n\}$. Then membership in $\mathcal{P}$ is tested by the following $\mathcal{L}^{*}_{\infty\omega}(\mathbf{C})$ sentence of rank 4:

$$\bigvee_{G \in \mathcal{P}} \forall x \forall y \ \left(E(x,y) \leftrightarrow \ (\bigvee_{(k,l) \in G} (k = \#z.(z < x)) \wedge (l = \#z.(z < y))) \right).$$

$\square$

Thus, the situation is similar to $\mathcal{L}^{\omega}_{\infty\omega}$ which also expresses every property of finite ordered structures. But as with $\mathcal{L}^{\omega}_{\infty\omega}$, we will show that without an order, the power of $\mathcal{L}^{*}_{\infty\omega}(\mathbf{C})$ is severely limited.

3.2 Locality of $\mathcal{L}^{*}_{\infty\omega}(\mathbf{C})$

We give a simple and *direct* proof that $\mathcal{L}^{*}_{\infty\omega}(\mathbf{C})$ is Gaifman-local. We do not need to establish a Hanf-type locality result first, and we also improve the bound for locality rank.

THEOREM 3.8. *Every $\mathcal{L}^{*}_{\infty\omega}(\mathbf{C})$ formula is Gaifman-local; furthermore,* $\mathsf{lr}(\varphi) \leq \frac{3^{\mathsf{rk}(\varphi)}-1}{2}$.

Proof. We start with a lemma, that was proved in a rather complicated way in [24]; a simple proof is given below.

LEMMA 3.9. *Assume that in $\mathcal{A}$, $\vec{a} \approx_{3r+1} \vec{b}$. Then there exists a permutation $\pi : A \to A$ such that $\vec{a}x \approx_r \vec{b}\pi(x)$ for every $x \in A$.*

Proof of the lemma. Let τ be an isomorphism type of an r-neighborhood around a single point. Since we have an isomorphism $h : N_{3r+1}(\vec{a}) \to N_{3r+1}(\vec{b})$, which maps $S_{2r+1}(\vec{a})$ onto $S_{2r+1}(\vec{b})$, we get that the number of points c in $S_{2r+1}(\vec{a})$ and $S_{2r+1}(\vec{b})$ with $ntp^{\mathcal{A}}_r(c) = \tau$ is the same; and thus the number of points c with $ntp^{\mathcal{A}}_r(c) = \tau$ in complements of these sets is the same:

$$|\{c \in A - S^{\mathcal{A}}_{2r+1}(\vec{a}) \mid ntp^{\mathcal{A}}_r(c) = \tau\}| \;=\; |\{c \in A - S^{\mathcal{A}}_{2r+1}(\vec{b}) \mid ntp^{\mathcal{A}}_r(c) = \tau\}| \,.$$

Since $|S_{2r+1}(\vec{a})| = |S_{2r+1}(\vec{b})|$, there exists a bijection $g : A - S_{2r+1}(\vec{a}) \to A - S_{2r+1}(\vec{b})$ such that $ntp^{\mathcal{A}}_r(x) = ntp^{\mathcal{A}}_r(g(x))$ for all $x \in A - S_{2r+1}(\vec{a})$. We now define $\pi(x)$ to be $h(x)$ if $x \in S_{2r+1}(\vec{a})$ and $g(x)$ if $x \notin S_{2r+1}(\vec{a})$. Clearly, π is a permutation. If $x \in S_{2r+1}(\vec{a})$, then $S_r(\vec{a}x) \subseteq S_{3r+1}(\vec{a})$ and $S_r(\vec{b}h(x)) \subseteq S_{3r+1}(\vec{b})$; hence $\vec{a} \approx_r \vec{b}\pi(x)$, because h is an isomorphism. If $x \notin S_{2r+1}(\vec{a})$, then for $y = \pi(x)$ we have $d(y, \vec{b}) > 2r+1$ and $x \approx_r y$. Thus, $N_r(\vec{a}x)$ and $N_r(\vec{b}y)$ are disjoint unions of isomorphic r-neighborhoods, and hence isomorphic. $\square$

Now the proof of the theorem proceeds by induction on the $\mathcal{L}^{\circ}_{\infty\omega}(\mathbf{C})$ formulae. We prove the most interesting case of $\psi(\vec{y};\vec{\jmath}) = \exists ix\ \varphi(x,\vec{y};\vec{\jmath})$ here (assuming i is in $\vec{\jmath}$); other cases are simple. Let $n = \mathsf{rk}(\varphi)$; then $r = \mathsf{lr}(\varphi) \leq (3^n - 1)/2$. It suffices to show $\mathsf{lr}(\psi) \leq 3r+1$. Assume $\vec{a} \approx_{3r+1} \vec{b}$ and fix an arbitrary $\vec{\jmath}_0$. Let $\mathcal{A} \models \psi(\vec{a};\vec{\jmath}_0)$. Then for at least i distinct $c_1, \ldots, c_i$ we have $\mathcal{A} \models \varphi(c_l, \vec{a}; \vec{\jmath}_0)$, $l = 1, \ldots, i$. From Lemma 3.9, get a permutation $\pi : A \to A$ such that $\vec{a}x \approx_r \vec{b}\pi(x)$, and let $d_l = \pi(c_l)$. Since $\mathsf{lr}(\varphi) = r$ and $\vec{a}c_l \approx_r \vec{b}d_l$, we get $\mathcal{A} \models \varphi(d_l, \vec{b}; \vec{\jmath}_0)$ for $l = 1, \ldots, i$. As all d_ls are distinct, $\mathcal{A} \models \psi(\vec{b};\vec{\jmath}_0)$. The converse is identical. $\square$

COROLLARY 3.10. $\mathrm{FO}(Q_{\mathbf{u}})$, $\mathrm{FO}(\mathbf{C})$ *and* FO *formulae are Gaifman-local, and* $\mathsf{lr}(\varphi) \leq (3^{\mathsf{qr}(\varphi)} - 1)/2$. *Furthermore, $\mathcal{L}^{*}_{\infty\omega}(\mathbf{C})$ (and thus $\mathrm{FO}(Q_{\mathbf{u}})$, $\mathrm{FO}(\mathbf{C})$ and FO) formulae without free second-sort variables have the bounded degree property.* $\square$

	$\mathcal{L}^{\omega}_{\infty\omega}$	$\mathcal{L}^{*}_{\infty\omega}(\mathbf{C})$
on ordered structures	expresses every property	expresses every property
on unordered structures	cannot count: has 0-1 law	does not have recursion mechanism: is local
relationship to other logics	subsumes fixpoint logics	subsumes counting logics

Fig. 2. $\mathcal{L}^{\omega}_{\infty\omega}$ and $\mathcal{L}^{*}_{\infty\omega}(\mathbf{C})$: A comparison

Thus, $\mathcal{L}^{*}_{\infty\omega}(\mathbf{C})$ is indeed a good counterpart of $\mathcal{L}^{\omega}_{\infty\omega}$ if we want to address the issue of inexpressibility of recursive queries in counting logics: while $\mathcal{L}^{\omega}_{\infty\omega}$ subsumes fixpoint logics but cannot express nontrivial counting properties, $\mathcal{L}^{*}_{\infty\omega}(\mathbf{C})$ subsumes a number of counting logics, but lacks a recursion mechanism. Both logics express every property of ordered finite structures.

4. QUANTITATIVE ANALYSIS OF LOCALITY

We introduce a new tool for providing a finer analysis of locality of counting logics. On the surface, it is very close to Gaifman-locality. However, the new definition accounts precisely for what is happening in a bijective game played on structures $\mathcal{A} \leftrightarrows_d \mathcal{B}$, and for the increase of locality rank of a formula with the addition of a new quantifier. It will also allow us to determine *precisely* how the locality rank depends on the rank of a formula for a variety of counting logics. Currently, our estimate is $\mathsf{lr}(\varphi) \leq O(3^{\mathsf{qr}(\varphi)})$ and, intuitively, it comes from the idea that the duplicator chooses a bijection in the bijective game in such a way as to maintain *isomorphism* of neighborhoods of configurations. The radius of these neighborhoods decreases as the game progresses. The new form of locality will allow the duplicator to maintain a weaker condition, and will lead to a $O(2^{\mathsf{qr}(\varphi)})$ bound for the locality rank. We will also use this condition to improve bounds in the results of [10].

Let $\vec{x} = (x_1, \ldots, x_n)$, and let $\mathcal{I} = (I_1, \ldots, I_m)$ be a partition of $\{1, \ldots, n\}$. The subvector of $\vec{x}$ that consists of the components whose indices are in I_j is denoted by $\vec{x}^{\mathcal{I}}_j$.

Definition 4.1. *1) Let $\mathcal{I}$ be a partition $(I_1, \ldots, I_m)$ of $\{1, \ldots, n\}$. Let $r > 0$. Given two structures, $\mathcal{A}$ and $\mathcal{B}$, and $\vec{a} \in A^n, \vec{b} \in B^n$, we say that $\vec{a}$ and $\vec{b}$ are* $(\mathcal{I}, r)$-similar *if the following hold:*

—*$ntp^{\mathcal{A}}_r(\vec{a}^{\mathcal{I}}_j) = ntp^{\mathcal{B}}_r(\vec{b}^{\mathcal{I}}_j)$ for all $j = 1, \ldots, m$;*

—*$d(\vec{a}^{\mathcal{I}}_j, \vec{a}^{\mathcal{I}}_k) > r$ for all $k \neq j$;*

—*$d(\vec{b}^{\mathcal{I}}_j, \vec{b}^{\mathcal{I}}_k) > r$ for all $k \neq j$.*

We call $\vec{a}$ and $\vec{b}$ r-similar, *and write $\vec{a} \sim_r^{\mathcal{A},\mathcal{B}} \vec{b}$, if there exists a partition $\mathcal{I}$ such that $\vec{a}$ and $\vec{b}$ are $(\mathcal{I}, r)$-similar. If $\mathcal{A} = \mathcal{B}$, we write $\vec{a} \sim_r^{\mathcal{A}} \vec{b}$.*

3) A formula φ has the r-separation property *if $\vec{a} \sim_r^{\mathcal{A}} \vec{b}$ implies $\mathcal{A} \models \varphi(\vec{a}) \leftrightarrow \varphi(\vec{b})$. To extend this to two-sorted logics, we require $\varphi(\vec{a}; \vec{\jmath}_0) \leftrightarrow \varphi(\vec{b}; \vec{\jmath}_0)$ for every tuple $\vec{\jmath}_0$ over $\mathbb{N}$. A formula has the separation property iff it has the r-separation property for some finite r. We write $\mathsf{SP}(r)$ for the class of formulae that have the r-separation property.* □

PROPOSITION 4.2. *1) A formula has the separation property iff it is Gaifman-local.*
2) If $\varphi \in \mathsf{SP}(r)$, then $\mathsf{lr}(\varphi) \leq r$.
3) $\vec{a} \approx_r \vec{b}$ implies $\vec{a} \sim_r \vec{b}$ and $\vec{a} \sim_{2r+1} \vec{b}$ implies $\vec{a} \approx_r \vec{b}$.

Proof. Both 1) and 2) are immediate consequences of 3). That $\vec{a} \approx_r \vec{b}$ implies $\vec{a} \sim_r \vec{b}$ follows by considering a partition with one block $\{1, \ldots, n\}$. If $\vec{a} \sim_{2r+1} \vec{b}$, there is an isomorphism $h_j : N_{2r+1}(\vec{a}_j^{\mathcal{I}}) \to N_{2r+1}(\vec{b}_j^{\mathcal{I}})$ for each block j of a partition $\mathcal{I}$ witnessing $\vec{a} \sim_{2r+1} \vec{b}$. Let h_j' be the restriction of h_j to $N_r(\vec{a}_j^{\mathcal{I}})$, and let h be the union of all h_j's. Since there is no tuple that can contain elements of $S_r(\vec{a}_{j_1}^{\mathcal{I}})$ and $S_r(\vec{a}_{j_2}^{\mathcal{I}})$ for $j_1 \neq j_2$ (and likewise for $\vec{b}$), we obtain that h is an isomorphism between $N_r(\vec{a})$ and $N_r(\vec{b})$. □

Thus, the essence of the new notion of locality is the same as Gaifman's. However, while $\vec{a} \approx_r \vec{b}$ implies $\vec{a} \sim_r \vec{b}$, the converse is not true, and the weaker notion $\sim_r$ allows the duplicator more freedom in the game, as will be shown later. We now use the definition of the separation property to prove the following key lemma.

LEMMA 4.3. *Let $r > 0$, $\mathcal{A} \leftrightarrows_r \mathcal{B}$, and $\vec{a} \sim_{2r}^{\mathcal{A},\mathcal{B}} \vec{b}$. Then there exists a bijection $f : A \to B$ such that, for every $x \in A$, $\vec{a}x \sim_r^{\mathcal{A},\mathcal{B}} \vec{b}f(x)$.*

Proof. Assume $\mathcal{I} = (I_1, \ldots, I_m)$ is a partition of $\{1, \ldots, n\}$ that witnesses $\vec{a} \sim_{2r}^{\mathcal{A},\mathcal{B}} \vec{b}$. That is, $\vec{a}_j^{\mathcal{I}} \approx_{2r}^{\mathcal{A},\mathcal{B}} \vec{b}_j^{\mathcal{I}}$ for each $j \leq m$, and $d(\vec{a}_j^{\mathcal{I}}, a_k^{\mathcal{I}}) > 2r$ and likewise for $\vec{b}$. In particular, $S_r^{\mathcal{A}}(\vec{a}_j^{\mathcal{I}}) \cap S_r^{\mathcal{A}}(\vec{a}_k^{\mathcal{I}}) = \emptyset$ and $S_r^{\mathcal{B}}(\vec{b}_j^{\mathcal{I}}) \cap S_r^{\mathcal{B}}(\vec{b}_k^{\mathcal{I}}) = \emptyset$ for $j \neq k$. Let $h_j : N_{2r}^{\mathcal{A}}(\vec{a}_j^{\mathcal{I}}) \to N_{2r}^{\mathcal{B}}(\vec{b}_j^{\mathcal{I}})$ be an isomorphism, and let h_j^0 be its restriction to $N_r^{\mathcal{A}}(\vec{a}_j^{\mathcal{I}})$.

Let τ be an isomorphism type of an r-neighborhood of a single point. If $x \in S_r^{\mathcal{A}}(\vec{a}_j^{\mathcal{I}})$ realizes τ, then so does $h_j^0(x) \in S_r^{\mathcal{B}}(\vec{b}_j^{\mathcal{I}})$, because $S_r^{\mathcal{A}}(x) \subseteq S_{2r}^{\mathcal{A}}(\vec{a}_j^{\mathcal{I}})$, and h_j is an isomorphism. Thus, if we let $m_j^{\mathcal{A}}(\tau, \vec{a})$ be the number of elements of $S_r^{\mathcal{A}}(\vec{a}_j^{\mathcal{I}})$ that realize τ, and likewise for $m_j^{\mathcal{B}}(\tau, \vec{b})$, we obtain $m_j^{\mathcal{A}}(\tau, \vec{a}) = m_j^{\mathcal{B}}(\tau, \vec{b})$.

Next, define $m_0^{\mathcal{A}}(\tau, \vec{a})$ to be the number of elements in $A - S_r^{\mathcal{A}}(\vec{a})$ that realize τ, and likewise for $m_0^{\mathcal{B}}(\tau, \vec{a})$. Since $\mathcal{A} \leftrightarrows_r \mathcal{B}$, the number of elements realizing τ in $\mathcal{A}$ and $\mathcal{B}$ is the same, and hence $m_0^{\mathcal{A}}(\tau, \vec{a}) = m_0^{\mathcal{B}}(\tau, \vec{a})$. Thus, since $| A |=| B |$, there exists a bijection $g : A - S_r^{\mathcal{A}}(\vec{a}) \to B - S_r^{\mathcal{B}}(\vec{b})$ such that $x \approx_r^{\mathcal{A},\mathcal{B}} g(x)$ for every x.

We now define f to be h_j^0 on $S_r^{\mathcal{A}}$, $j = 1, \ldots, m$, and g on $A - S_r^{\mathcal{A}}(\vec{a}) = A -$

$\bigcup_{j=1}^{m} S_r^{\mathcal{A}}(\vec{a}_j^{\mathcal{I}})$. Clearly, f is a bijection. We now must show $\vec{a}x \sim_r^{\mathcal{A},\mathcal{B}} \vec{b}f(x)$. Let $x \in S_r^{\mathcal{A}}(\vec{a}_j^{\mathcal{I}})$. Then we define a partition $\mathcal{I}'$ of $\{1,\ldots,n+1\}$ by adding $n+1$ to the jth block of $\mathcal{I}$. Since $\vec{a}_j^{\mathcal{I}}x \approx_r^{\mathcal{A},\mathcal{B}} \vec{b}_j^{\mathcal{I}}h_j^0(x)$ and $d(x,\vec{a}_k^{\mathcal{I}}), d(h_j^0(x),\vec{b}_k^{\mathcal{I}}) > r$ for any $k \neq j$, we see that $\vec{a}x$ and $\vec{b}f(x)$ are $(\mathcal{I}',r)$-similar. Finally, if $x \notin S_r^{\mathcal{A}}(\vec{a})$, we define $\mathcal{I}'$ by extending $\mathcal{I}$ with one new block $\{n+1\}$. It follows from the properties of g that in this case $\vec{a}x$ and $\vec{b}f(x)$ are $(\mathcal{I}',r)$-similar, too. This completes the proof. □

4.1 Applications of the separation property

Using Lemma 4.3, one can modify the inductive proof of Theorem 3.8, to show the following.

THEOREM 4.4. *Let $\varphi(\vec{x};\vec{\jmath})$ be a $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$ formula. Then $\varphi \in \mathsf{SP}(2^{\mathsf{rk}(\varphi)})$.*

Proof is by induction on the formula. We assume, in view of Theorem 3.5, that φ is a $\mathcal{L}^\circ_{\infty\omega}(\mathbf{C})$ formula. For the base case of atomic formulae of rank 0, observe that $\vec{a} \sim_1 \vec{b}$ implies $\vec{a} \approx_0 \vec{b}$, which means that $(\vec{a};\vec{\jmath})$ and $(\vec{b};\vec{\jmath})$ satisfy all the same atomic $\mathcal{L}^\circ_{\infty\omega}(\mathbf{C})$ formulae. The cases of the Boolean connectives and second-sort quantification are trivial. The remaining case is $\varphi(\vec{x};\vec{\jmath}) \equiv \exists iz.\psi(\vec{x},z;\vec{\jmath})$. Let $m = \mathsf{rk}(\psi)$ and $r = 2^m$; then $\psi \in \mathsf{SP}(r)$. We must show $\varphi \in \mathsf{SP}(2r)$. Assume that in a structure $\mathcal{A}$, $\vec{a} \sim_{2r}^{\mathcal{A}} \vec{b}$. Since $\mathcal{A} \leftrightarrows_r \mathcal{A}$, by Lemma 4.3, we find a permutation f on A such that $\vec{a}x \sim_r \vec{b}f(x)$. Now assume $\mathcal{A} \models \varphi(\vec{a};\vec{\jmath})$. Then there exists i distinct elements $c_1,\ldots,c_i \in A$ such that $\mathcal{A} \models \psi(\vec{a},c_l,\vec{\jmath})$ for each $c_l, l = 1,\ldots,i$. Let $d_l = f(c_l)$. Then all d_ls are distinct, we obtain $\vec{a}c_l \sim_r^{\mathcal{A}} \vec{b}d_l$, and then $\mathcal{A} \models \psi(\vec{b},d_l,\vec{\jmath})$ for each d_l since $\psi \in \mathsf{SP}(r)$. Hence, $\mathcal{A} \models \varphi(\vec{b},\vec{\jmath})$. The same proof shows $\mathcal{A} \models \varphi(\vec{b},\vec{\jmath})$ implies $\mathcal{A} \models \varphi(\vec{a},\vec{\jmath})$. Thus, $\varphi \in \mathsf{SP}(2r) = \mathsf{SP}(2^{\mathsf{rk}(\varphi)})$. □

COROLLARY 4.5. *Let φ be an* FO*, or* FO($\mathbf{C}$)*, or* FO($Q_{\mathbf{u}}$)*, or $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$ formula. Then $\varphi \in \mathsf{SP}(2^{\mathsf{rk}(\varphi)})$ and $\mathsf{lr}(\varphi) \leq 2^{\mathsf{rk}(\varphi)}$.* □

The result below also follows from Lemma 4.3.

THEOREM 4.6. *Assume that $\mathcal{A}$, $\mathcal{B}$ are two structures of the same relational signature and $\mathcal{A} \leftrightarrows_{2^n-1} \mathcal{B}$. Then the duplicator has a strategy in the n-round bijective game that guarantees that after $i \leq n$ rounds, if $\vec{a} = (a_1,\ldots,a_i)$ and $\vec{b} = (b_1,\ldots,b_i)$ are points played on $\mathcal{A}$ and $\mathcal{B}$ respectively, then $\vec{a} \sim_{2^{n-i}}^{\mathcal{A},\mathcal{B}} \vec{b}$.*

Proof. For the first move, the duplicator chooses a bijection $f : A \to B$ that guarantees $x \approx_{2^n-1}^{\mathcal{A}} f(x)$ for all x, and hence $x \sim_{2^n-1}^{\mathcal{A}} f(x)$. This is possible since $\mathcal{A} \leftrightarrows_{2^n-1} \mathcal{B}$. For each following move, the bijection for the duplicator is provided by Lemma 4.3. □

Since $(a_1,\ldots,a_n) \sim_1^{\mathcal{A},\mathcal{B}} (b_1,\ldots,b_n)$ implies that $\{(a_i,b_i) \mid i = 1,\ldots,n\}$ is a partial isomorphism $\mathcal{A} \to \mathcal{B}$, we obtain the following.

Corollary 4.7. *Let $\mathcal{A}, \mathcal{B} \in$ STRUCT$[\sigma]$, and $\mathcal{A} \leftrightarrows_{2^n-1} \mathcal{B}$. Then $\mathcal{A} \equiv_n^{bij} \mathcal{B}$. Consequently, $\mathcal{A}$ and $\mathcal{B}$ agree on all* FO, FO$(\mathbf{C})$, *and* FO$(Q_{\mathbf{u}})$ *sentences of quantifier rank up to n.* □

Previously it was only known how to obtain bounds of the order $O(3^n)$, with the exception of [20] where it was shown that for $\mathcal{A} \leftrightarrows_{2^n-1} \mathcal{B}$, the duplicator wins an n-round Ehrenfeucht-Fraïssé game on $\mathcal{A}$ and $\mathcal{B}$.

We can also sharpen the bound in a theorem of Fagin, Stockmeyer and Vardi [10]. Recall [10] that structures $\mathcal{A}$ and $\mathcal{B}$ are (d, m)-*equivalent*, if for every isomorphism type τ of a d-neighborhood of a point, and for $n_{\mathcal{A}} =\mid \{a \in A \mid ntp_d^{\mathcal{A}}(a) = \tau\} \mid$ and $n_{\mathcal{B}} =\mid \{b \in B \mid ntp_d^{\mathcal{B}}(b) = \tau\} \mid$, either $n_{\mathcal{A}} = n_{\mathcal{B}} < m$, or $n_{\mathcal{A}}, n_{\mathcal{B}} \geq m$. The result below uses the separation property to improve $(3^{k-1}, m)$-equivalence (from [10]) to $(2^{k-1}, m)$-equivalence.

Proposition 4.8. *For every relational signature σ and all positive integers k and c, there exists a positive integer m such that $\mathcal{A} \equiv_k \mathcal{B}$, whenever $\mathcal{A}, \mathcal{B} \in$ STRUCT$_c[\sigma]$ are $(2^{k-1}, m)$-equivalent.*

Proof. Let $m = k(2c+1)^{2^{k-1}} + 1$. Note that if all degrees in $\mathcal{A}$ and $\mathcal{B}$ are bounded by c, then m exceeds the size of any r-neighborhood of k points for $r \leq 2^{k-1}$. The proof now follows along the lines of Theorem 4.6, as we show that after $i \leq k$ rounds, if $(a_1, \ldots, a_i)$ and$(b_1, \ldots, b_i)$ are points played in $\mathcal{A}$ and $\mathcal{B}$ respectively, then $(a_1, \ldots, a_i) \sim_{2^{k-i}}^{\mathcal{A},\mathcal{B}} (b_1, \ldots, b_i)$. Since $\sim_1$ implies $\approx_0$, this suffices to conclude that the duplicator wins.

For the first move, the duplicator can always respond to $a \in A$ with $b \in B$ such that $a \approx_{2^{k-1}}^{\mathcal{A},\mathcal{B}} b$ (and vice versa) – this follows from (d, m)-equivalence. Assume that after i rounds we have $\vec{a}$ and $\vec{b}$ played on $\mathcal{A}$ and $\mathcal{B}$, and $\vec{a} \sim_{2r}^{\mathcal{A},\mathcal{B}} \vec{b}$, where $r = 2^{k-i-1}$. It suffices to show that for every $a_0 \in A$, there exists $b_0 \in B$ such that $\vec{a}a_0 \sim_r^{\mathcal{A},\mathcal{B}} \vec{b}b_0$, and or every $b_0 \in B$, there exists $a_0 \in A$ such that the above holds. We show the 'forth' part; the 'back' part is identical.

Let $\mathcal{I}$ be a partition of $\{1, \ldots, i\}$ that witnesses $(\mathcal{I}, 2r)$-equivalence of $\vec{a}$ and $\vec{b}$. Assume that $a_0 \in S_r^{\mathcal{A}}(\vec{a}_j^{\mathcal{I}})$ for some j. Let h be an isomorphism $h : N_{2r}^{\mathcal{A}}(\vec{a}_j^{\mathcal{I}}) \to N_{2r}^{\mathcal{A}}(\vec{b}_j^{\mathcal{I}})$. We then let b_0 be $h(a_0)$; the proof that $\vec{a}a_0 \sim_r^{\mathcal{A},\mathcal{B}} \vec{b}b_0$ is the same as in the proof of Lemma 4.3.

Let $a_0 \notin S_r^{\mathcal{A}}(\vec{a})$. Let τ be the isomorphism type of the r-neighborhood of a_0. Let s be the number of elements of $S_r^{\mathcal{A}}(\vec{a})$ that realize τ. From $\vec{a} \sim_{2r} \vec{b}$ we see that the number of elements of $S_r^{\mathcal{B}}(\vec{b})$ that realize τ is also s. Our choice of m guarantees that $s < m$; hence, (d, m)-equivalence of $\mathcal{A}$ and $\mathcal{B}$ implies that there exists $b_0 \in B - S_r^{\mathcal{B}}(\vec{b})$ such that $a_0 \approx_r^{\mathcal{A},\mathcal{B}} b_0$. Now the same argument as in the proof of Lemma 4.3 shows that $\vec{a}a_0 \sim_r^{\mathcal{A},\mathcal{B}} \vec{b}b_0$, which concludes the proof. □

5. HOW FAR DO LOCAL QUERIES SEE?

Here we use the techniques from the previous section to find precise bounds on locality rank, and on Hanf locality rank for various counting logics.

Assume that we deal with graphs, i.e, the signature has one binary relation symbol E. Consider the family of FO formulae: $d_0(x,y) = E(x,y) \vee E(y,x) \vee y = x$ and $d_{k+1}(x,y) = \exists z. d_k(x,z) \wedge d_k(z,y)$. Note that $\mathsf{rk}(d_k) = \mathsf{qr}(d_k) = k$. In a graph G, $G \models d_k(a,b)$ iff the distance from a to b in G is at most 2^k. This implies $\mathsf{lr}(d_k) \geq 2^{k-1}$.

Thus, the locality rank is necessarily exponential in the quantifier-rank of a first-order formula. For a logic $\mathcal{L}$, we define

$$\text{Loc\_rank}_{\mathcal{L}}(n) = \max\{\mathsf{lr}(\varphi) \mid \varphi \in \mathcal{L}, \mathsf{rk}(\varphi) = n\}.$$

From Corollary 4.5, we obtain $2^{n-1} \leq \text{Loc\_rank}_{\mathcal{L}}(n) \leq 2^n$, if $\mathcal{L}$ is FO, or FO($\mathbf{C}$), or FO($Q_{\mathbf{u}}$), or $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$. Can the *precise* value of the function Loc_rank be calculated? We do it below, by modifying slightly both the family $\{d_k(x,y)\}$ and the separation property.

THEOREM 5.1. *Let $\mathcal{L}$ be* FO, *or* FO($\mathbf{C}$), *or* FO($Q_{\mathbf{u}}$), *or* $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$. *Then, for any* $n > 0$,

$$\text{Loc\_rank}_{\mathcal{L}}(n) \;=\; 2^n - 1.$$

Proof. To see the lower bound, we need a family of FO formulae slightly different from $\{d_k(x,y)\}$. We assume a signature that consists of one binary relation E and one unary relation U and define $\alpha_0(x) = U(x)$ and $\alpha_{k+1}(x) = \exists z. d_k(x,z) \wedge \alpha_k(z)$. This formula says that there exists an element of U at the distance at most $2^k - 1$ from x. It is easy to see that $\mathsf{lr}(\alpha_k) \geq 2^k - 1$ and $\mathsf{rk}(\alpha_k) = k$; hence $\text{Loc\_rank}_{\mathcal{L}}(n) \geq 2^n - 1$.

To prove the upper bound, define, for two structures $\mathcal{A}, \mathcal{B} \in \text{STRUCT}[\sigma]$ and $d > 0$, a binary relation $\theta_d^{\mathcal{A},\mathcal{B}}$ on tuples of the same length from A and B, as follows. Let $\vec{a} = (a_1, \ldots, a_n) \in A^n$ and $\vec{b} = (b_1, \ldots, b_n) \in B^n$. Then $\vec{a}\theta_d^{\mathcal{A},\mathcal{B}}\vec{b}$ if there exists a partition $\mathcal{I} = (I_1, \ldots, I_m)$ of $\{1, \ldots, n\}$ such that

$\vec{a}_j^{\mathcal{I}} \approx_{d-1}^{\mathcal{A},\mathcal{B}} \vec{b}_j^{\mathcal{I}}$ for every $j = 1, \ldots, m$, and
$d(\vec{a}_j^{\mathcal{I}}, \vec{a}_k^{\mathcal{I}}), d(\vec{b}_j^{\mathcal{I}}, \vec{b}_k^{\mathcal{I}}) > d$ for every $j \neq k$.

Then, using an argument similar to the proof of Lemma 4.3, we show:

LEMMA 5.2. *If $r > 0$, $\mathcal{A} \leftrightarrows_{r-1} \mathcal{B}$ and $\vec{a}\theta_{2r}^{\mathcal{A},\mathcal{B}}\vec{b}$, then there exists a bijection $f : A \to B$ such that $\vec{a}x\theta_r^{\mathcal{A},\mathcal{B}}\vec{b}f(x)$ for every $x \in A$.*

Proof of the lemma: Let $\mathcal{I}$ be a partition witnessing $\vec{a}\theta_{2r}^{\mathcal{A},\mathcal{B}}\vec{b}$; that is, $\vec{a}_j^{\mathcal{I}} \approx_{2r-1}^{\mathcal{A},\mathcal{B}} \vec{b}_j^{\mathcal{I}}$ and $d(\vec{a}_j^{\mathcal{I}}, \vec{a}_k^{\mathcal{I}}), d(\vec{b}_j^{\mathcal{I}}, \vec{b}_k^{\mathcal{I}}) > 2r$ for every $j \neq k$. Let h_j^0 be the restriction of an isomorphism $h_j : N^{\mathcal{A}}_{2r-1}(\vec{a}^{\mathcal{I}}_j) \to N^{\mathcal{B}}_{2r-1}(\vec{b}^{\mathcal{I}}_j)$ to $N^{\mathcal{A}}_r(\vec{a}^{\mathcal{I}}_j)$. Note that $S^{\mathcal{A}}_r(\vec{a}^{\mathcal{I}}_j)$ and $S^{\mathcal{B}}_r(\vec{b}^{\mathcal{I}}_j)$ have the same size, and, since $S^{\mathcal{A}}_r(\vec{a}^{\mathcal{I}}_j) \cap S^{\mathcal{A}}_r(\vec{a}^{\mathcal{I}}_k) = \emptyset$ for $j \neq k$ (and likewise for $\vec{b}$), $|A - S^{\mathcal{A}}_r(\vec{a})| = |B - S^{\mathcal{B}}_r(\vec{b})|$.

Let τ be an isomorphism type of a $r-1$-neighborhood of a single point. Since the number of elements realizing τ is the same in $S^{\mathcal{A}}_r(\vec{a}^{\mathcal{I}}_j)$ and $S^{\mathcal{B}}_r(\vec{b}^{\mathcal{I}}_j)$ for each j, we obtain that τ has the same number of realizers in $A - S^{\mathcal{A}}_r(\vec{a})$ and in $B - S^{\mathcal{B}}_r(\vec{b})$. Since $\mathcal{A} \leftrightarrows_{r-1} \mathcal{B}$, there exists a bijection $g : A - S^{\mathcal{A}}_r(\vec{a}) \to B - S^{\mathcal{B}}_r(\vec{b})$ such that $x \approx^{\mathcal{A},\mathcal{B}}_{r-1} g(x)$ for every $x \in A - S^{\mathcal{A}}_r(\vec{a})$. We now define $f(x)$ to be $h^0_j(x)$ on $S^{\mathcal{A}}_r(\vec{a}^{\mathcal{I}}_j)$ and $g(x)$ on $A - S^{\mathcal{A}}_r(\vec{a})$. The same argument as in the proof of Lemma 4.3 shows that $\vec{a}x\theta^{\mathcal{A},\mathcal{B}}_r\vec{b}f(x)$. $\square$

Using Lemma 5.2, we can prove, similarly to Theorem 4.4, that for any $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$ formula $\varphi(\vec{x};\vec{\jmath})$ with $\mathsf{rk}(\varphi) = k$, and any structure $\mathcal{A}$, $\vec{a}\theta^{\mathcal{A},\mathcal{A}}_{2^k}\vec{b}$ implies that $\mathcal{A} \models \varphi(\vec{a};\vec{\jmath}) \leftrightarrow \varphi(\vec{b};\vec{\jmath})$ for every $\vec{\jmath}$. It suffices to observe that $\vec{a}\theta^{\mathcal{A},\mathcal{A}}_1\vec{b}$ implies $\vec{a} \approx^{\mathcal{A}}_0 \vec{b}$, and the inductive prove is same as before. Next, we note that $\vec{x} \approx^{\mathcal{A}}_{d-1} \vec{y}$ implies $\vec{x}\theta^{\mathcal{A},\mathcal{A}}_d\vec{y}$, since the partition can be chosen to consist of one set with all the indices. Thus, for a formula $\varphi(\vec{x};\vec{\jmath})$ of rank k, $\vec{a} \approx^{\mathcal{A}}_{2^k-1} \vec{b}$ implies $\mathcal{A} \models \varphi(\vec{a};\vec{\jmath}) \leftrightarrow \varphi(\vec{b};\vec{\jmath})$ for every $\vec{\jmath}$. Therefore, $\mathsf{lr}(\varphi) \leq 2^k - 1$, and $\text{Loc\_rank}_{\mathcal{L}}(k) = 2^k - 1$. $\square$

Remark: The reason why the separation property itself could not be used to prove this theorem, is the following. It is possible to find, for any n, a formula $\alpha_n(x)$ such that $\alpha_n \in \mathsf{SP}(2^n)$ but $\alpha_n \notin \mathsf{SP}(r)$ for any $r < 2^n$. In fact, the formulae $\alpha_n(x)$ we introduced in the proof to show the lower bound, are such.

While for the formulae d_k witnessing exponential locality rank we have $\mathsf{qr}(d_k) = k$, the number of quantifiers in d_k is $2^k - 1$. Thus, one may conjecture that $\mathsf{lr}(\varphi)$ is polynomial (or even linear) in the total number of quantifiers. This conjecture is refuted by the result below, which shows that alternation gives us another way of getting exponential locality rank.

THEOREM 5.3. *For each $n > 3$, there is a prenex first-order formula $\psi_n(x,y)$ whose prefix consists of n quantifiers, such that $\mathsf{lr}(\psi_n) \geq 2^{\lfloor \frac{n}{2} \rfloor - 1}$.*

Proof. We fix the language to be $\{E, U_1, U_2\}$, where E is binary and U_1, U_2 are unary. Assume first that n is even and $k = n/2$. We construct the structure $\mathcal{A}_n$ as follows: U_1 and U_2 are interpreted as two disjoint sets whose union is A, each of cardinality 2^k. The binary predicate E is interpreted as a successor relation on U_1 and U_2; that is, $\mathcal{A}$ can be viewed as a pair of successor relations of the same length. We use $\mathcal{A}^1_n$ for $\mathcal{A}_n$ restricted to U_1, and likewise for $\mathcal{A}^2_n$.

We now define $\psi_n(x,y)$ as a formula such that $\mathcal{A}_n \models \psi_n(a,b)$ iff $a \in U_1, b \in U_2$ and $(\mathcal{A}^1_n, a) \equiv_k (\mathcal{A}^2_n, b)$; that is, for every first-order formula $\gamma(x)$ of quantifier rank k in the language of graphs, $\mathcal{A}^1_n \models \gamma(a)$ iff $\mathcal{A}^2_n \models \gamma(b)$. The formula ψ_n is defined as

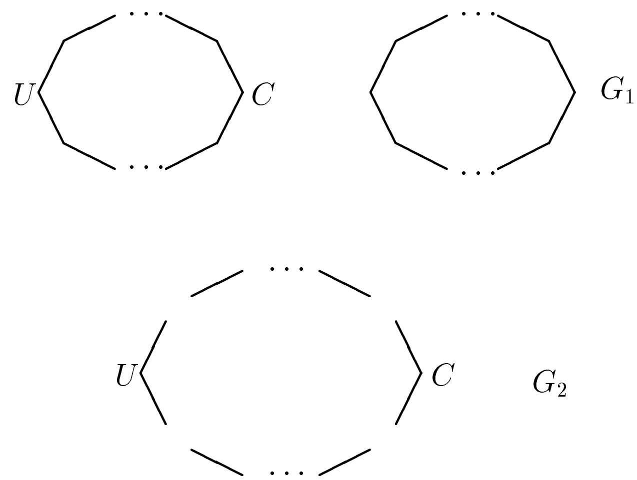

Fig. 3. Example for Theorem 5.4

follows:

$$\forall x_1 \exists y_1 \forall x_2 \exists y_2 \ldots \forall x_k \exists y_k . (U_1(x) \wedge U_2(y) \wedge \bigwedge_{i=1}^{k} \mathit{Diff}(x_i, y_i) \wedge F_n(\vec{x}, \vec{y}, x, y))$$

where $\mathit{Diff}(u,v)$ is $(U_1(u) \leftrightarrow U_2(v)) \wedge (U_1(v) \leftrightarrow U_2(u))$ and $F_n(\vec{x}, \vec{y}, x, y)$ states the configurations given by $\vec{x}, \vec{y}, x, y$ in $\mathcal{A}_n^1$ and $\mathcal{A}_n^2$ are isomorphic. That is,

$$F_n(\vec{x}, \vec{y}, x, y) \ = \bigvee_{I \subseteq \{1,\ldots,k\}} ((\bigwedge_{i \in I} U_1(x_i) \wedge \bigwedge_{j \notin I} U_1(y_j)) \wedge ((\vec{x}_I, \vec{y}_{\bar{I}}, x) \approx_0 (\vec{x}_{\bar{I}}, \vec{y}_I, y)))$$

Here $(\vec{x}_I, \vec{y}_{\bar{I}}, x) \approx_0 (\vec{x}_{\bar{I}}, \vec{y}_I, y)$ states that exactly the same atomic formulae holds for the elements of these tuples; it can be expressed as a quantifier-free formula.

Next, consider $a, a' \in U_1$ as the distance 2^{k-1} and $2^{k-1}+1$ from the start node, respectively. Let $b \in U_2$ be at the distance $2^{k-1}+1$ from the start node. Note that for $r = 2^{k-1}-1$, $N_r^{\mathcal{A}_n}(a,b) \cong N_r^{\mathcal{A}_n}(a',b)$.

Since $(\mathcal{A}_n^1, a') \cong (\mathcal{A}_n^2, b)$, we conclude that they agree on all formulae, in particular, those of quantifier rank up to k. Hence, $\mathcal{A}_n \models \psi_n(a,b)$. Now, define $\gamma(z) = \exists v (d_{k-1}(v,z) \wedge \forall u \ \neg E(u,v))$. This formula is of quantifier rank k; it says that z is at the distance at most 2^{k-1} from the start node. Hence, $\mathcal{A}_n^1 \models \gamma(a)$ and $\mathcal{A}_n^2 \models \neg\gamma(b)$. Since $\mathsf{qr}(\gamma) = k$, we obtain $\mathcal{A}_n \models \neg\psi_n(a,b)$. This implies $\mathsf{lr}(\psi_n) > r$, that is, $\mathsf{lr}(\psi_n) \geq 2^{\frac{n}{2}-1}$.

If n is odd, we apply the above to $n+1$ to obtain $\mathsf{lr}(\psi_n) \geq 2^{\lfloor \frac{n}{2} \rfloor - 1}$. This concludes the proof. $\square$

Similarly to Loc_rank(n), define Hanf_rank$_{\mathcal{L}}(n) = \max\{\mathsf{hlr}(\Phi) \mid \Psi \in \mathcal{L}, \mathsf{rk}(\Psi) = n\}$.

THEOREM 5.4. *Let $\mathcal{L}$ be* FO, *or* FO($\mathbf{C}$), *or* FO($Q_{\mathbf{u}}$). *Then, for any $n > 1$,*

$$\text{Hanf\_rank}_{\mathcal{L}}(n) = 2^{n-1} - 1.$$

Proof. We first prove the lower bound. Consider structures of the signature $\{E, U, C\}$, where E is binary, U, C are unary; that is, graphs with two sets of vertices. Let $\Phi_k = \exists x.\alpha_k(x) \wedge C(x)$, where α_k is defined as in the proof of Theorem 5.1. That is, Φ_k says that there are two nodes in a graph that are at a distance at most $2^k - 1$ and they belong to C and U respectively. We now construct two graphs, G_1 being a union of two cycles of length $2^{k+1} - 2$, with C and U containing one element each, at the distance $2^k - 1$ in one of the cycles. Graph G_2 is one cycle of length $2^{k+2} - 4$, with C and U interpreted as single element each, at the distance $2^{k+1} - 2$. See Figure 3. Then $G_1 \leftrightarrows_{2^k-2} G_2$ is verified easily, and further $G_1 \models \Phi_k$ while $G_2 \models \neg\Phi_k$. Thus, $\mathsf{hlr}(\Phi_k) > 2^k - 2$. Since $\mathsf{rk}(\Phi_k) = k + 1$, we conclude Hanf_rank$_{\mathcal{L}}(n) \geq 2^{n-1} - 1$.

To see the upper bound, we show that, if $\mathcal{A} \leftrightarrows_{2^{n-1}-1} \mathcal{B}$, the duplicator can win the n-round bijective Ehrenfeucht-Fraïssé game. In fact, we show that the duplicator can maintain the condition that if $(a_1, \ldots, a_i)$ and $(b_1, \ldots, b_i)$ are elements played on $\mathcal{A}$ and $\mathcal{B}$ respectively, then $\vec{a}\theta^{\mathcal{A},\mathcal{B}}_{2^{n-i}}\vec{b}$. This suffices, since $\vec{a}\theta^{\mathcal{A},\mathcal{B}}_1\vec{b}$ implies that $\{(a_l, b_l) \mid l = 1, \ldots, n\}$ is a partial isomorphism.

For the first move, the duplicator can always play so that $a_1 \approx^{\mathcal{A},\mathcal{B}}_{2^{n-1}-1} b_1$, and hence $a_1\theta^{\mathcal{A},\mathcal{B}}_{2^{n-1}}b_1$. Assuming $\vec{a}\theta^{\mathcal{A},\mathcal{B}}_{2^{n-k}}\vec{b}$, we apply Lemma 5.2 from the proof of Theorem 5.1 to find a bijection $f : A \to B$ such that $\vec{a}x\theta^{\mathcal{A},\mathcal{B}}_{2^{n-k-1}}\vec{b}f(x)$. This is the bijection used by the duplicator, as it guarantees that the condition $\vec{a}\theta^{\mathcal{A},\mathcal{B}}_{2^{n-i}}\vec{b}$ is maintained. Hence, $\mathcal{A} \leftrightarrows_{2^{n-1}-1} \mathcal{B}$ implies that $\mathcal{A}$ and $\mathcal{B}$ agree on all FO($Q_{\mathbf{u}}$) (FO, FO($\mathbf{C}$)) sentences of rank up to n, thus showing Hanf_rank$(n) \leq 2^{n-1} - 1$. $\Box$

6. OUTPUTS OF LOCAL QUERIES

Recall that $\#\mathsf{ntp}_d(\mathcal{A})$ is the cardinality of $\{ntp^{\mathcal{A}}_d(a) \mid a \in A\}$; that is, the number of different isomorphism types of d-neighborhoods of points realized in $\mathcal{A}$, and $deg(\mathcal{A})$ is the number of different degrees realized in $\mathcal{A}$. Fact 2.7 relates $deg(\varphi[\mathcal{A}])$ and $\#\mathsf{ntp}_d(\mathcal{A})$ when $\varphi(x, y)$ is a Gaifman-local graph query. Its proof in [7] relied on an analog of Lemma 3.9, which establishes a permutation of elements of the carrier. To extend the result to *arbitrary* local queries, one needs an analog of Lemma 3.9 that establishes a permutation of *vectors* over the carrier. It turns out that Lemma 4.3 provides such a tool, and allows us to solve an open problem from [7].

We do not specify which logic the formulae $\varphi(\vec{x})$ belong to; it could be any that is Gaifman-local (e.g., FO($\mathbf{C}$), $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$). In case of two-sorted logic, like $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$, we require that all free variables be of the first sort.

THEOREM 6.1. *Let $\varphi(x_1, \ldots, x_m) \in \mathsf{SP}(r)$, $m > 1$. Then, for any structure $\mathcal{A}$, $deg(\varphi[\mathcal{A}]) \leq m \cdot \#\mathsf{ntp}_{2^{m-1}r}(\mathcal{A})$.*

Proof. First, we show that for any structure $\mathcal{A}$, whenever $\vec{a} \sim^{\mathcal{A}}_{2^k r} \vec{b}$, there exists a bijection $f : A^k \to A^k$ such that, for any $\vec{x} \in A^k$, $\vec{a}\vec{x} \sim_r \vec{b}f(\vec{x})$. The proof is by induction on k. For $k = 1$, this follows from Lemma 4.3. Assume $\vec{a} \sim_{2^{k+1}r} \vec{b}$. Then, by the hypothesis, there exists a bijection $g : A^k \to A^k$ such that $\vec{a}\vec{x} \sim_{2r} \vec{b}g(\vec{x})$ for any $\vec{x} \in A^k$. By Lemma 4.3, there exists a bijection $\pi_{\vec{x}} : A \to A$ such that for every $x_0 \in A$, $\vec{a}\vec{x}x_0 \sim_r \vec{b}g(\vec{x})\pi_{\vec{x}}(x_0)$. We now define a new bijection $f : A^{k+1} \to A^{k+1}$ as follows: for $\vec{z} \in A^{k+1}$, let $\vec{x}$ be its first k components, and x_0 the last component. Then $f(\vec{z}) = g(\vec{x})\pi_{\vec{x}}(x_0)$. It follows that $\vec{a}\vec{z} \sim_r \vec{b}f(\vec{z})$, and it is routine to verify that f is a bijection.

Let $a \approx^{\mathcal{A}}_{2^{m-1}r} b$. Then $a \sim^{\mathcal{A}}_{2^{m-1}r} b$, and the claim above shows that there exists a bijection $f : A^{m-1} \to A^{m-1}$ such that $a\vec{x} \sim^{\mathcal{A}}_r bf(\vec{x})$ for each $\vec{x}$. Since $\varphi \in \mathsf{SP}(r)$, we obtain $\mathcal{A} \models \varphi(a\vec{x})$ iff $\mathcal{A} \models \varphi(bf(\vec{x}))$. Thus, $degree_1(a) = degree_1(b)$ in $\varphi[\mathcal{A}]$, and hence the number of different values of $degree_1(x), x \in A$, is at most $\#\mathsf{ntp}_{2^{m-1}r}(\mathcal{A})$. Thus, $deg(\varphi[\mathcal{A}]) \le m \cdot \#\mathsf{ntp}_{2^{m-1}r}(\mathcal{A})$. □

COROLLARY 6.2. *Let $\varphi(x_1, \ldots, x_m)$, $m > 1$, be a $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$ formula, with all free variables of the first sort. Then $deg(\varphi[\mathcal{A}]) \le m \cdot \#\mathsf{ntp}_{2^{\mathsf{rk}(\varphi)+m-1}}(\mathcal{A})$ for any structure $\mathcal{A}$.* □

Combining Propositions 6.1 and 4.2, we answer an open question from [7].

COROLLARY 6.3. *If $\varphi(x_1, \ldots, x_m)$ is a Gaifman-local query, $m > 1$ and $\mathsf{lr}(\varphi) = r > 0$, then for any structure $\mathcal{A}$, $deg(\varphi[\mathcal{A}]) \le m \cdot \#\mathsf{ntp}_{2^m r + 2^{m-1}}(\mathcal{A})$.*

Proof. In the proof of Theorem 6.1 we showed that $a \approx^{\mathcal{A}}_{2^k d} b$ implies that there exists a bijection $f : A^k \to A^k$ such that $a\vec{x} \sim^{\mathcal{A}}_d bf(\vec{x})$. Let $d = 2r + 1$. Assume $a \approx^{\mathcal{A}}_{2^m r + 2^{m-1}} b$. Then we have a bijection $g : A^{m-1} \to A^{m-1}$ such that $a\vec{x} \sim^{\mathcal{A}}_{2r+1} bg(\vec{x})$, and by Proposition 4.2, $a\vec{x} \approx^{\mathcal{A}}_r bg(\vec{x})$ for any $\vec{x}$. By locality of φ, $\mathcal{A} \models \varphi(a\vec{x})$ iff $\mathcal{A} \models \varphi(bg(\vec{x}))$, and thus $degree_1(a) = degree_1(b)$ in $\varphi[\mathcal{A}]$; hence, as in the proof of Theorem 6.1, we conclude that $deg(\varphi[\mathcal{A}]) \le m \cdot \#\mathsf{ntp}_{2^m r + 2^{m-1}}(\mathcal{A})$. □

If $\mathcal{A} \in \mathrm{STRUCT}_k[\sigma]$, there is a bound on $\#\mathsf{ntp}_r(\mathcal{A})$ that depends only on k, r and σ. Indeed, $\mid S^{\mathcal{A}}_r(a) \mid$ is at most of the order of k^r, and the number of different structures (of a fixed signature) of size n is at most exponential in $p(n)$, where p is a polynomial. Thus, we have

COROLLARY 6.4. *For every relational signature σ, every positive integer r, there exist integers c and d such that, if $\varphi(\vec{x})$ is a σ-formula and $\mathsf{lr}(\varphi) \le r$, then for any structure $\mathcal{A} \in \mathrm{STRUCT}_k[\sigma]$, we have*

$$deg(\varphi[\mathcal{A}]) \le c^{k^d}. \qquad \square$$

This proves that local formulae have the BDP. From counting neighborhoods, of a fixed radius r when degrees are bounded by k, one can obtain that for every relational signature σ, every $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$ σ-formula $\varphi(\vec{x})$ with all free variables of the first sort, and $k > 0$, there exist constants a, b, c such that for $\mathcal{A} \in \mathrm{STRUCT}_k[\sigma]$,

$deg(\varphi[\mathcal{A}]) \leq a^{b^{c^{\mathrm{rk}(\varphi)}}} \stackrel{\mathrm{def}}{=} g(\varphi, k)$. This can be used to derive lower bounds on the rank of $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$ (and thus FO, FO($\mathbf{C}$) and FO($Q_{\mathbf{u}}$)) formulae that compute certain queries, as well as upper bounds on the size of structures on which a given formula has certain behavior. For example, consider the transitive closure query. If its input is an n-element successor relation (whose degrees are either 0 or 1), the output realizes all degrees from 0 to n. Thus, if $\varphi(x, y)$ is a $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$ formula in the language of graphs, and m is such that $\varphi[G]$ is the transitive closure of G whenever G has at most m vertices, then m cannot exceed $g(\varphi, 1)$.

7. CONCLUSION

In this paper, we defined a logic $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$ that subsumes a number of counting extensions of FO (such as FO($\mathbf{C}$) and FO($Q_{\mathbf{u}}$)) and has enormous counting power; at the same time, its formulae are local. This can be interpreted as showing independence of counting from recursive computation over unordered structures. Note that adding both fixpoints and counting to FO was studied rather extensively (see, e.g., [5; 13; 14; 28]) in connection with capturing PTIME over unordered structures. We then proved tight bounds on locality rank for a variety of counting logics, and described outputs of local queries.

We now briefly discuss applications and new directions. First, one can add counting over the numerical domain to $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$. This has to be done with care, as the value of a counting term may then be infinite. Such an extension was given recently in [18]. Moreover, [18] showed that $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$ can be extended by means of aggregate functions. An example would be a formula $\psi(x, j) \equiv \sum_{(y,i)} \varphi(x, y, i)$ such that $\psi(x, j)$ holds if $j = \sum_{(y,i):\varphi(x,y,i) \text{ holds}} i$. This is a typical operation in database query languages [1]. It was shown in [18] that $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$ has sufficient power to code such aggregate functions. Moreover, [18] proved that $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$ formulae are Hanf-local, thereby answering a number of open questions on the expressive power of database query languages with aggregates.

Over *ordered* structures, FO($\mathbf{C}$) captures complexity class TC^0 (the class of problems accepted by constant-depth polynomial-size unbounded fan-in threshold circuits, under DLOGTIME-uniformity [3]). As shown in [17], this class contains non-local order-invariant problems. Nevertheless, we plan to work on defining the appropriate notions of locality in the ordered setting. This may shed light not only on complexity-theoretic problems, but also on expressiveness of database languages: in the presence of an order, proving bounds for a language with aggregates hinges on separation of TC^0 from other classes [18]. Interestingly, there are other cases where expressiveness problems for database query languages cannot be resolved without separating complexity classes [2].

The idea of putting the unrestricted numerical domain on the side was influenced by the development of metafinite model theory [12]. We hope that this work may help formulate analogs of locality theorems in the metafinite context.

Acknowledgements I thank Lauri Hella and Moshe Vardi for asking questions that led to some of the results in this paper. The definition of $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$ owes a lot to my discussions with Lauri, Juha Nurmonen and Limsoon Wong. I also thank Michael Benedikt, Kousha Etessami, Neil Immerman, and anonymous referees for their comments.

REFERENCES

[1] S. Abiteboul, R. Hull and V. Vianu. *Foundations of Databases*, Addison Wesley, 1995.

[2] S. Abiteboul and V. Vianu. Computing with first-order logic. *Journal of Computer and System Sciences* 50 (1995), 309–335.

[3] D.M. Barrington, N. Immerman, H. Straubing. On uniformity within NC^1. *Journal of Computer and System Sciences*, 41:274–306, 1990.

[4] M. Benedikt, H.J. Keisler. Expressive power of unary counters. *Proc. Int. Conf. on Database Theory (ICDT'97)*, Springer LNCS 1186, January 1997.

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

[6] A. Dawar, S. Lindell, S. Weinstein. First order logic, fixed point logic, and linear order. In *Proc. Comp. Sci. Logic (CSL'95)*, pages 161–177.

[7] G. Dong, L. Libkin and L. Wong. Local properties of query languages. *Proc. Int. Conf. on Database Theory (ICDT'97)*, Springer LNCS 1186, 1997, pages 140–154. Full version to appear in *Theoretical Computer Science.*

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

[9] K. Etessami. Counting quantifiers, successor relations, and logarithmic space, *Journal of Computer and System Sciences*, 54 (1997), 400–411.

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

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

[12] E. Grädel and Y. Gurevich. Metafinite model theory. *Information and Computation* 140 (1998), 26–81.

[13] E. Grädel and M. Otto. Inductive definability with counting on finite structures. In *Proc. Computer Science Logic'92*, Springer LNCS vol. 702, 1993, pages 231–247.

[14] S. Grumbach and C. Tollu. On the expressive power of counting. *Theoretical Computer Science* 149 (1995), 67–99.

[15] W. Hanf. Model-theoretic methods in the study of elementary logic. In J.W. Addison et al, eds, *The Theory of Models*, North Holland, 1965, pages 132–145.

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

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

[18] L. Hella, L. Libkin, J. Nurmonen and L. Wong. Logics with aggregate operators. In *Proc. 14th IEEE Symp. on Logic in Computer Science (LICS'99)*, Trento, Italy, July 1999, pages 35–44.

[19] L. Hella and G. Sandu. Partially ordered connectives and finite graphs. In M. Krynicki, M. Mostowski and L. Szczerba, eds, *Quantifiers: Logics, Models and Computation II*, Kluwer Academic Publishers, 1995, pages 79–88.

[20] N. Immerman. *Descriptive Complexity.* Springer Verlag, 1998.

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

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

[23] Ph. Kolaitis, M. Vardi. Infinitary logic and 0-1 laws. *Information and Computation*, 98 (1992), 258–294.

[24] L. Libkin. On the forms of locality over finite models. In *Proc. 12th IEEE Symp. on Logic in Computer Science (LICS'97)*, Warsaw, Poland, June–July 1996, pages 204–215.

[25] L. Libkin, L. Wong. Query languages for bags and aggregate functions. *Journal of Computer and System Sciences* 55 (1997), 241–272.

[26] S. Lifsches and S. Shelah. Distorted sums of models. Unpublished manuscript.

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

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