On the Forms of Locality over Finite Models

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

Abstract

Most proofs showing limitations of expressive power of first-order logic rely on Ehrenfeucht-Fraisse games. Playing the game often involves a nontrivial combinatorial argument, so it was proposed to find easier tools for proving expressivity bounds. Most of those known for first-order logic are based on its "locality", that is defined in different ways. In this paper we characterize the relationship between those notions of locality. We note that Gaifman's locality theorem gives rise to two notions: one deals with sentences and one with open formulae. We prove that the former implies Hanf's notion of locality, which in turn implies Gaifman's locality for open formulae. Each of these implies the bounded degree property, which is one of the easiest tools for proving expressivity bounds. These results apply beyond the first-order case. We use them to derive expressivity bounds for first-order logic with unary quantifiers and counting. Finally, we apply these results to relational database languages with aggregate functions, and prove that purely relational queries defined in such languages satisfy Gaifman's notion of locality. From this we derive a number of expressivity bounds for languages with aggregates.

1 Introduction

It is well known that first-order logic has limited expressive power. Typically, inexpressibility proofs are based on either a compactness argument, or Ehrenfeucht-Fraisse games. In recent years, expressive power of logics over *finite* models has been studied extensively. This increased interest is mostly due to a number of applications in computer science. For example, most database query languages have well known logical counterparts: traditional relational calculus has precisely the power of first-order logic, the language Datalog, with added negation and evaluated inflationary, corresponds to the least-fixpoint logic, and the query language with while loops is equivalent to the partial-fixpoint logic, cf. [1]. Another area of application is descriptive complexity. It turns out that familiar logics capture complexity classes over the classes of (ordered) finite structures, cf. [21, 12].

Since compactness fails in the case of finite models [12], to prove results about the limits of expressiveness of first-order logic, one has to use Ehrenfeucht-Fraisse games. Moreover, Ehrenfeucht-Fraisse games are often used as the basic step in other, more sophisticated games for different logics, cf. [14]. For example, playing the Ehrenfeucht-Fraisse game is one of the steps in the Ajtai-Fagin game for monadic Σ^1_1 [3]. Since playing the game often involves an intricate combinatorial argument, it was suggested by Fagin, Stockmeyer and Vardi in [15] to build a library of winning strategies for those games. Or, more generally, one would like to have a collection of versatile and easily applicable tools for proving expressivity bounds of first-order logic.

A number of results proving expressivity bounds explain the nature of the limitations of first-order by saying that it can only express *local* properties. Intuitively, one cannot grasp the whole structure; instead, to answer a first-order query, one only looks at small portions of the input.

Several proposals have been made to formalize the notion of locality. Gaifman [16] proved that every first-order formula is equivalent to a local one, in the sense that only a small part of the input is relevant for evaluating a query. Fagin, Stockmeyer and Vardi [15], modifying a result by Hanf [19] for the finite case, proved that if a certain criterion relating the numbers of small neighborhoods in two structures holds, then these structures agree on sentences whose quantifier rank is determined by the size of those neighborhoods. The author and Wong [25] showed that if first-order query operates on graphs, then the number of different in- and out-degrees in the output is below a bound given by the query and the maximal degree in the input graph. That is, if locally the input looks simple, then so does the output of a first-order query. We called this the *bounded degree property*. It was generalized to queries on arbitrary finite structures by Dong, Wong and the author [10].

At a more intuitive level, the weakness of first-order logic is often attributed to its inability to count (e.g, parity of cardinality is not definable), and lacking a mechanism for doing recursion (e.g., transitive closure is not definable). Usually, the proofs of inexpressibility of properties that involve recursive computation are harder than of those based on counting; and the tools we mentioned are typically applied to that class of problems.

Looking at various examples of showing expressivity bounds, one can observe a certain difficulty of proof vs. difficulty of application tradeoff. While the characterization of logics via games was historically the first results of this kind to be proved, it is often the hardest technique to apply. Hanf's technique seems to make life easier: for example, it simplifies the proof that connectivity is not monadic Σ_1^1 [15] quite a lot, compared to [3], but sometimes the combinatorial argument is not completely trivial [7]. Proofs of applicability of Hanf's technique are usually not very hard, see [15, 13, 26, 27]. Further down the road one has Gaifman's locality theorem, whose proof is harder than that of Hanf's technique, but which leads to simpler and cleaner inexpressibility proofs (see [10]). However, no extension of first-order logic is known to satisfy an analog of Gaifman's theorem. Finally, we have the bounded degree property, whose proof is based on Gaifman's theorem, and which leads to particularly simple inexpressibility proofs, cf. [10, 25]. Very recently, with considerable amount of effort, it was shown that the bounded degree property holds for certain queries in a first-order relational language extended with aggregate functions [10] (this language has substantial counting power).

The goal of this paper is to study the relationship between the general notions of locality, and show their applications for proving various expressivity bounds. In fact, our results confirm the intuition of the previous paragraph that certain notions of locality are harder than others, but are easier to apply. Our results are not limited to first-order logic only: they are shown to be applicable to logics with counting and generalized unary quantifiers, as well as relational database query languages with aggregation.

Organization and summary In Section 2, we introduce the notations and describe the basic notions of locality. We start by reviewing Gaifman's theorem, and note that it leads to two notions, called *Gaifman's locality* and *strong Gaifman's locality*. The result of [16] then says that first-order logic has both. We review the modification of Hanf's technique [19] for the finite case [15], and define the notion of Hanf's locality property. We review the bounded degree property of [10, 25] which is implied by Gaifman's locality [10].

In Section 3, we show that Hanf's locality implies Gaifman's locality and the bounded degree property. We use these results to derive expressivity bounds for various logics; we also mention some applications in descriptive complexity. Section 3 begins with a "warm-up" direct proof that Hanf's locality implies the bounded degree property for graph queries.

In Section 4, we show that strong Gaifman's locality implies Hanf's locality. We do not yet know any extension of first-order that has strong Gaifman's locality property, so the main implication of this result is a very simple and intuitive proof that first-order logic has Hanf's locality property.

In Section 5, we deal with relational query languages with aggregate functions. Traditional query languages often correspond to logical languages, and the equivalence of relational algebra and first-order logic is the best known example of such correspondence [1]. However, real query languages often use aggregates (for example, a query may ask for the total number of employees in a department). Several attempts have been made recently to analyze the expressive power of aggregation (see [9, 17, 25] and a survey [18]). In particular, [18] lists an open problem whether such a relational language with aggregate functions has the bounded degree property for purely relational queries. This was proved very recently [10], but the proof is not completely satisfactory, as it relies on syntactic properties of the language rather than its basic logical properties and, more importantly, cannot be extended to show that such queries are local. Here we give a much simplified proof that implies Gaifman's locality, not just the bounded degree property. It is based on simulating relational queries in logic with counting.

Complete proofs are given in the full version [23].

2 Notions of locality

Notations

Unless explicitly stated otherwise, all structures are assumed to be *finite*.

A relational signature σ is a set of relation symbols $\{R_1, ..., R_l\}$, with an associated arity function. In what follows, $p_i(>0)$ denotes the arity of R_i. We write σ_n for σ extended with n new constant symbols. The signature of graphs (that is, one binary predicate R) is denoted by σ_{gr}.

A σ-structure is $\mathcal{A} = \langle A, \overline{R}_1, \ldots, \overline{R}_l \rangle$, where A is a finite set, and $\overline{R}_i \subseteq A^{p_i}$ interprets R_i. The class of finite σ-structures is denoted by STRUCT$[\sigma]$. When there is no confusion, we may write R_i in place of $\overline{R}_i$. Isomorphism of structures is denoted by $\cong$. We shall adopt the convention that the carrier of a structure $\mathcal{A}$ is always denoted by A and the carrier of $\mathcal{B}$ is denoted by B.

To make our results applicable to a number of languages, we assume that an abstract logic comes equipped with the notion of formulae $\psi(x_1, \ldots, x_m)$ with free variables $x_1 \ldots x_m$ and sentences in the language containing relation symbols, functions and constants, and the notion of satisfaction $\models$ between structures and sentences in appropriate vocabulary. We also assume that these are closed under the usual Boolean connectives $\vee, \neg$ and first-order quantification. Note that these notions can be made precise (cf. [11, 22]), but we needn't go into details here, since all logics we consider are extensions of first-order, and the meaning of all the notions above is clear.

With each formula $\psi(x_1, \ldots, x_m)$ in the logical language whose symbols are in σ, we associate a *query* that maps a σ-structure $\mathcal{A}$ into a m-ary relation $\psi[\mathcal{A}] = \langle A, \{(a_1, \ldots, a_m) \in A^m \mid \mathcal{A} \models \psi(a_1, \ldots, a_m)\}\rangle$.

Given a structure $\mathcal{A}$, its *Gaifman graph* [12, 16, 15] $\mathcal{G}(\mathcal{A})$ is defined as $\langle A, E\rangle$ where (a, b) is in E iff there is a tuple $\vec{t} \in \overline{R}_i$ for some i such that both a and b are in $\vec{t}$. 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 $d(a, a) = 0$. Given $a \in A$, its *r-sphere* $S_r^{\mathcal{A}}(a)$ is $\{b \in A \mid d(a, b) \leq r\}$. For a tuple $\vec{t}$, define $S_r^{\mathcal{A}}(\vec{t})$ as $\bigcup_{a \in \vec{t}} S_r^{\mathcal{A}}(a)$.

Given a tuple $\vec{t} = (t_1, \ldots, t_n)$, its *r-neighborhood* $N_r^{\mathcal{A}}(\vec{t})$ is defined as a σ_n structure

$$\langle S_r^{\mathcal{A}}(\vec{t}), \overline{R}_1 \cap S_r^{\mathcal{A}}(\vec{t})^{p_1}, \ldots, \overline{R}_k \cap S_r^{\mathcal{A}}(\vec{t})^{p_k}, t_1, \ldots, t_n\rangle$$

That is, the carrier of $N_r^{\mathcal{A}}(\vec{t})$ is $S_r^{\mathcal{A}}(\vec{t})$, the interpretation of the σ-relations is obtained by restricting them from $\mathcal{A}$ to the carrier, and the n extra constants are the elements of $\vec{t}$. If the structure $\mathcal{A}$ is understood, we shall write $S_r(\vec{t})$ and $N_r(\vec{t})$.

The quantifier rank of a formula, $\mathsf{qr}(\psi)$, is defined as the maximum depth of quantifier nesting in ψ.

Gaifman's locality

Before presenting Gaifman's theorem, note that for any σ-structure $\mathcal{A}$, there is a first order formula $\gamma_\sigma(x, y)$ such that $\mathcal{A} \models \gamma_\sigma(a, b)$ iff $(a, b) \in \mathcal{G}(\mathcal{A})$. Thus, for every fixed k, there are first order formulae $d_{<k}(x, y)$, $d_k(x, y)$ and $d_{>k}(x, y)$ such that $\mathcal{A} \models d_{<k}(a, b)$ iff $d(a, b) < k$, and similarly for d_k and $d_{>k}$. This means that bounded quantification of the form $\forall x \in S_k(\vec{y})$ and $\exists x \in S_k(\vec{y})$ is expressible for every constant k. If every quantifier in a formula is of this form, where $\vec{y}$ are among its free variables, and $k \leq r$, we call the formula *r-local*.

Theorem 2.1 (Gaifman [16]) *Every first-order formula $\psi(x_1, \ldots, x_n)$ is equivalent to a Boolean combination of t-local formulae $\chi(x_{i_1}, \ldots, x_{i_s})$ and sentences of the form*

$$(1) \qquad \exists y_1 \ldots y_m.(\bigwedge_{i=1}^{m} \varphi(y_i) \wedge \bigwedge_{i,j \leq m, i \neq j} d_{>2r}(y_i, y_j))$$

where φ is r-local. Furthermore, $r \leq 7^{\mathsf{qr}(\psi)-1}$, $t \leq (7^{\mathsf{qr}(\psi)-1} - 1)/2$, $m \leq n + \mathsf{qr}(\psi)$, and, if ψ is a sentence, only sentences (1) occur in the Boolean combination. $\square$

This theorem is a result about first-order logic on finite structures. To abstract the notion of being local and extend it to other logics, we introduce the following definitions. For two vectors $\vec{x}$ and $\vec{y}$ of the same length, we write $\vec{x} \approx_d^{\mathcal{A}} \vec{y}$ if $N_d^{\mathcal{A}}(\vec{x}) \cong N_d^{\mathcal{A}}(\vec{y})$. Again, $\mathcal{A}$ is omitted if it is understood.

Definition 2.2
- *A formula $\psi(x_1, \ldots, x_m)$, is* Gaifman-local *if there exists $r > 0$ such that, for every $\mathcal{A} \in$ STRUCT$[\sigma]$ and for every two m-ary vectors $\vec{a}$, $\vec{b}$ of elements of A, $\vec{a} \approx_r \vec{b}$ implies $\mathcal{A} \models \psi(\vec{a})$ iff $\mathcal{A} \models \psi(\vec{b})$. The minimum r for which this holds is called the* locality rank *of ψ, and is denoted by $\mathsf{lr}(\psi)$.*
- *A formula $\psi(x_1, \ldots, x_m)$, is* strongly Gaifman-local *if there exists $r > 0$ such that, for every*

$\mathcal{A}, \mathcal{B} \in \mathrm{STRUCT}[\sigma]$ *and for every two* m*-ary vectors* $\vec{a}$, $\vec{b}$ *of elements of* A *and* B *respectively,* $N_r^{\mathcal{A}}(\vec{a}) \cong N_r^{\mathcal{B}}(\vec{b})$ *implies* $\mathcal{A} \models \psi(\vec{a})$ *iff* $\mathcal{B} \models \psi(\vec{b})$.

- *A sentence* Ψ *is* strongly Gaifman-local *if it is equivalent to a Boolean combination of sentences of the form* $\exists \vec{y}.\psi(\vec{y})$, *where* $\psi(\vec{y})$ *is strongly Gaifman-local formula.*

Now we immediately see:

Proposition 2.3 *Every first-order formula is Gaifman-local, and every first-order sentence is strongly Gaifman-local. Moreover, for every* $\psi(\vec{x})$ *of quantifier rank* n, $\mathsf{lr}(\psi) \leq (7^n - 1)/2$. □

Note that not every first-order formula is strongly Gaifman-local. Consider $\psi(x) \equiv (\forall y.\neg R(y,x)) \wedge \exists z \forall y.\neg R(z,y)$. Assume that it is strongly local, fix r as in the definition and consider two graphs: G_1 is a chain of length $r+1$, and G_2 is obtained from G_1 by adding a loop on the end-node of G_1. Let a_i be the start node of G_i. Then $N_r^{G_1}(a_1) \cong N_r^{G_2}(a_2)$, but $G_1 \models \psi(a_1)$ and $G_2 \models \neg\psi(a_2)$.

Hanf's locality

Let τ be an isomorphism type of a structure in the language σ_1 (σ extended with one constant). A point a in a structure $\mathcal{A}$ d-realizes τ, written as $\tau_d(\mathcal{A}, a) = \tau$, if $N_d(a)$ is of isomorphism type τ.

By $\#_d[\mathcal{A}, \tau]$ we denote the number of elements of A whose d-neighborhoods realize τ, that is, the cardinality of $\{a \in A \mid \tau_d(\mathcal{A}, a) = \tau\}$.

We say that $\mathcal{A}, \mathcal{B} \in \mathrm{STRUCT}[\sigma]$ are *d-equivalent*, if for any isomorphism type τ we have $\#_d[\mathcal{A}, \tau] = \#_d[\mathcal{B}, \tau]$. This is denoted by $\mathcal{A} \leftrightarrows_d \mathcal{B}$. If $d > d'$, then $\mathcal{A} \leftrightarrows_d \mathcal{B}$ implies $\mathcal{A} \leftrightarrows_{d'} \mathcal{B}$ [15].

It was shown by Hanf [19] that two (finite or infinite) models are elementary equivalent if their spheres of finite radius are finite and, for each d and each type τ, either $\#_d[\mathcal{A}, \tau] = \#_d[\mathcal{B}, \tau] < \omega$, or both $\#_d[\mathcal{A}, \tau]$ and $\#_d[\mathcal{B}, \tau]$ are infinite. This was recently modified for the finite case as follows.

Theorem 2.4 (Fagin-Stockmeyer-Vardi [15]) *Let* $n > 0$. *Then there exists an integer* $d > 0$ *such that whenever* $\mathcal{A} \leftrightarrows_d \mathcal{B}$, *then* $\mathcal{A}$ *and* $\mathcal{B}$ *agree on all first-order sentences of quantifier rank up to* n. □

It follows from the proof in [15] that d can be taken to be 3^{n-1}, see also [14]. This leads to the following definition.

Definition 2.5 *A sentence* Ψ *is* Hanf-local *if there exists a number* d *such that any two* d*-equivalent structures agree on* Ψ. *The minimum* d *for which this holds is called the Hanf locality rank of* Ψ, *and is denoted by* $\mathsf{hlr}(\Psi)$.

Thus, Fagin-Stockmeyer-Vardi's theorem says that every first order sentence Ψ is Hanf-local, and $\mathsf{hlr}(\Psi) \leq 3^{\mathsf{qr}(\Psi)-1}$. As we mentioned before, this notion of locality appears to be easier to prove, and several extensions of Theorem 2.4 are known. One such extension deals with *unary quantifiers*. Let σ_k^{unary} be a signature with k unary symbols, and let $\mathcal{K}$ be a class of σ_k^{unary}-structures. Then $\mathcal{FO}(Q_{\mathcal{K}})$ extends first-order logic formulae by adding the following formation rule: if $\psi(x_1, \vec{y}_1), \ldots, \psi(x_k, \vec{y}_k)$ are formulae, then $Q_{\mathcal{K}} x_1 \ldots x_k.(\psi_1, \ldots, \psi_k)$ is a formula with free variables $\vec{y}_1, \ldots, \vec{y}_k$. Its 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 the σ_k^{unary} structure whose ith relation is $\{a \in A \mid \mathcal{A} \models \psi_i(a, \vec{a}_i)\}$ is in $\mathcal{K}$. Examples of unary quantifiers include the usual $\exists$ and $\forall$, as well as Rescher and Härtig quantifiers. We use $\mathcal{FO}(Q_{\mathbf{u}})$ for $\mathcal{FO}$ extended with all unary quantifiers.

Fact 2.6 (see [26, 27]) *Every* $\mathcal{FO}(Q_{\mathbf{u}})$ *sentence is Hanf-local. Moreover,* $\mathsf{hlr}(\Psi) \leq 3^{\mathsf{qr}(\Psi)}$. □

Etessami [13] studied first-order logic with counting $\mathcal{FO} + COUNT$, which is defined as a two sorted logic, with second sort being the sort of natural numbers. On natural numbers one has $1, \max, <$ and the BIT predicate available ($\mathrm{BIT}(i,j)$ iff the ith bit in the binary representation of j is one). It also has counting quantifiers $\exists ix.\varphi(x)$, meaning that φ has i satisfiers; these quantifiers bind x but not i. Etessami noticed that the technique of Nurmonen's proof (which is based on bijective Ehrenfeucht-Fraisse games [20]) applies to it:

Fact 2.7 (see [13]) *Every* $\mathcal{FO} + COUNT$ *sentence is Hanf-local. Moreover,* $\mathsf{hlr}(\Psi) \leq 3^{\mathsf{qr}(\Psi)}$. □

Bounded degree property

We define the notions of *degrees* in the usual way. For a graph G, its *degree set* *deg_set*(G) is the set of all possible in- and out-degrees that are realized in G, and $deg(G)$ is the cardinality of $deg\_set(G)$. These notions generalize to arbitrary σ-structures: Given a relation $\overline{R}_i$ in $\mathcal{A}$, $degree_j(R_i, a)$ is the number of tuples in $\overline{R}_i$ whose jth component is a. Then $deg\_set(\mathcal{A})$ is the set $\{degree_j(R_i, a) \mid \overline{R}_i \in \mathcal{A}, a \in A, j \leq p_i\}$, and $deg(\mathcal{A})$ is its cardinality. The class of σ-structures $\mathcal{A}$ with $deg\_set(\mathcal{A}) \subseteq \{0, 1, \ldots, k\}$ is denoted by $\mathrm{STRUCT}_k[\sigma]$.

Definition 2.8 (see [10]) *A formula* $\psi(x_1, \ldots, x_m)$ *has the* bounded degree property (BDP)*, if there is a function* $f : \mathbb{N} \to \mathbb{N}$ *such that* $deg(\psi[\mathcal{A}]) \leq f(k)$ *for any* $\mathcal{A} \in \mathrm{STRUCT}_k[\sigma]$. □

The BDP was introduced and proved for first-order queries from graphs to graphs (that is, formulae $\psi(x, y)$ in the language σ_{gr}) in [25]. It was also shown there that the BDP proves many inexpressibility results effortlessly. For example, to prove that (deterministic) transitive closure [12, 21] is not first-order, consider the following $C_n \in \mathrm{STRUCT}_1[\sigma_{\mathrm{gr}}]$:

where n is the number of nodes. Since the degree-set of its (deterministic) transitive closure has n elements, it violates the BDP and thus is not first-order definable. Another example in [25] is testing for balanced binary trees (that is, all paths from the root to the leaves are of the same length; note that this involves both recursive computation and counting). Assume this test is definable, and assume G is an input graph. For every two nodes a, b in G, having two successors each, a_1, a_2 and b_1, b_2, we define a new graph $G_{a,b}$ by making b_1, b_2 the successors of a and a_1, a_2 the successors of b. If G were a balanced binary tree, then $G_{a,b}$ is a balanced binary tree iff a and b have the same distance to the root. Thus, we see that there is a first-order query that, when its input is a balanced binary tree $G \in \mathrm{STRUCT}_2[\sigma_{\mathrm{gr}}]$ of length n, returns the set of cliques of elements at the same distance from the root, that is, a graph with $n+1$-element degree-set. This again violates the BDP.

Theorem 2.9 (Dong-Libkin-Wong [10]) *Every Gaifman-local formula has the bounded degree property.* □

Thus, from Gaifman's theorem, we obtain:

Corollary 2.10 *Every first-order formula has the bounded degree property.* □

We saw that simple forms of recursion (deterministic transitive closure) violate the BDP. So does the simplest form of second-order quantification: monadic Σ_1^1 is not local. The BDP was introduced in connection with studying expressive power of database languages with aggregation [18, 25], where it was asked if such languages have it. The positive answer given recently [10] also implies that first-order logic with Rescher and Härtig quantifiers has the BDP, but it was not known (although conjectured) if any of these is Gaifman-local.

3 Hanf's locality implies Gaifman's locality

The main result of this section is:

Theorem 3.1 *Let $\mathcal{L}$ be a logic that is closed under first-order operations. Assume that every sentence in $\mathcal{L}$ is Hanf-local. Then every formula in $\mathcal{L}$ is Gaifman-local.*

Before we sketch the proof of this theorem, we list some corollaries. We immediately obtain

Corollary 3.2 *Let $\mathcal{L}$ be a logic closed under first-order operations. Assume that every sentence in $\mathcal{L}$ is Hanf-local. Then $\mathcal{L}$ has the bounded degree property.* □

Corollary 3.3 $\mathcal{FO}(Q_{\mathbf{u}})$ *and* $\mathcal{FO} + COUNT$ *have the bounded degree property.* □

More precisely, every $\mathcal{FO} + COUNT$ formula without free second-sort variables has the BDP. This generalizes a number of known results. For example, the bounded degree property of first-order logic with Härtig and Rescher quantifiers (proved in [10] by a lengthy and quite involved argument) follows straightforwardly. We also obtain a theorem by Etessami [13] that deterministic transitive closure is not definable in $\mathcal{FO} + COUNT$ in the presence of a successor relation. Note that this is a step towards separating TC^0 from DLOGSPACE, because $\mathcal{FO} + COUNT$ plus order captures uniform TC^0 [4] and $\mathcal{FO}$ with deterministic transitive closure and successor captures DLOGSPACE [12, 21]. Corollary 3.3 allows us to make the next incremental step:

Corollary 3.4 *Let $k \in \mathbb{N}$ and let $\mathcal{S}_k$ be any family of relations whose degrees do not exceed k. Then deterministic transitive closure is not definable in* $\mathcal{FO} + COUNT + \mathcal{S}_k$. □

Furthermore, using locality, we can extend the above results to more complex auxiliary data. Consider a class of structures $\mathcal{C} \subseteq \mathrm{STRUCT}[\sigma']$ for some relational vocabulary σ'. Define a function $s_{\mathcal{C}} : \mathbb{N} \to \mathbb{N}$ by letting $s_{\mathcal{C}}(n)$ be the maximal possible degree in some n-element structure $\mathcal{A} \in \mathcal{C}$. We say that $\mathcal{C}$ is of *moderate degree* (see [15]) if $s_{\mathcal{C}}(n) \leq \log^{o(1)} n$. That is, there is a function $\delta : \mathbb{N} \to \mathbb{N}$ such that $\lim_{n\to\infty} \delta(n) = 0$ and $s_{\mathcal{C}}(n) \leq \log^{\delta(n)} n$.

The following was shown in [10].

Proposition 3.5 (see [10]) *Let ψ be a local graph query, of locality rank r. Then for any structure $\mathcal{A}$, the number of distinct in-degrees in the graph $\psi[\mathcal{A}]$ is at most the number of non-isomorphic $3r+1$-neighborhoods realized in $\mathcal{A}$. The same is true for out-degrees.* □

Now one can use this proposition and calculate that, for structures of moderate degree, one cannot construct a graph that has n distinct in-degrees (where n is the number of nodes) for all n. This, and locality of $\mathcal{FO} + COUNT$, moves us one step closer to separating TC^0 from DLOGSPACE:

Corollary 3.6 *Transitive closure and deterministic transitive closure are* not *definable in $\mathcal{FO} + COUNT$ in the presence of relations of moderate degree.* □

However, the order relation adds all degrees from 0 to the cardinality of the input. Thus, we need a breakthrough like Schwentick's theorem [28] to generalize Corollary 3.4 to the ordered case.

Proof of Theorem 3.1

Before giving the proof of Theorem 3.1, we sketch a direct proof that Hanf's locality implies the graph BDP. The proof below completely avoids Lemma 3.11, which is the main technical tool for proving Theorem 3.1, and the proof that every local formula has the BDP [10]. We start by presenting a simple criterion for d-equivalence.

Proposition 3.7 *$\mathcal{A} \leftrightarrows_d \mathcal{B}$ iff there is a bijection $\pi : A \to B$ such that for any $a \in A$,*

$$N_d^{\mathcal{A}}(a) \;\; \cong \;\; N_d^{\mathcal{B}}(\pi(a)).$$

Proof: Let $\tau_1, \ldots, \tau_m$ be the collection of all isomorphism types of d-neighborhoods realized in $\mathcal{A}$ and $\mathcal{B}$. Let $A_i = \{a \in A \mid \tau_d(\mathcal{A}, a) = \tau_i\}$ and $B_i = \{b \in B \mid \tau_d(\mathcal{B}, b) = \tau_i\}$. Then $\{A_i\}_{i=1,m}$ and $\{B_i\}_{i=1,m}$ form partitions of A and B respectively. Assume $\mathcal{A} \leftrightarrows_d \mathcal{B}$. Then $card(A_i) = card(B_i)$ for every $i = 1, \ldots, m$, and the required π is defined as the union of bijective maps between A_i and B_i for all i. Conversely, if π satisfying $N_d^{\mathcal{A}}(x) \cong N_d^{\mathcal{B}}(\pi(x))$ exists, let τ be an isomorphism type and let $a_1, \ldots, a_k$ be the elements of A such that $\tau_d(\mathcal{A}, a_i) = \tau$. Then $\tau_d(\mathcal{B}, \pi(a_i)) = \tau$, and $\#_d[\mathcal{A}, \tau] \leq \#_d[\mathcal{B}, \tau]$. A symmetric argument shows the reverse inequality. □

Proof of Corollary 3.2 for graph queries (sketch). We start with a simple observation:

Lemma 3.8 *For any signature σ, there exist functions $f_\sigma, F_\sigma : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ such that for any $\mathcal{A} \in \mathrm{STRUCT}_k[\sigma]$,*

$$card(\{\tau \mid \exists a \in A.\tau_d(\mathcal{A}, a) = \tau\}) \leq f_\sigma(k, d) \text{ and}$$

$$\forall a \in A : \qquad card(S_d^{\mathcal{A}}(a)) \leq F_\sigma(k, d).$$

Given a signature σ, by $\sigma^{(n)}$ we denote σ extended with n new unary symbols $U_1, \ldots, U_n$. Given a structure $\mathcal{A}$ and an n-vector $\vec{a}$, by $\mathcal{A}[\vec{a}]$ we denote the $\sigma^{(n)}$ structure that extends $\mathcal{A}$ by interpreting U_is as singletons containing the components of $\vec{a}$.

Lemma 3.9 *Let $a \approx_{2d} b$ in $\mathcal{A}$, and let $d(a,b) > 2d+1$. Assume that $x \notin S_{2d+1}(a,b)$. Then $\mathcal{A}[a,x] \leftrightarrows_d \mathcal{A}[b,x]$.*

Proof: Let h be an isomorphism between $N_{2d}(a)$ and $N_{2d}(b)$. We define $\pi : A \to A$ by $\pi(z) = z$ for $z \notin S_d(a,b)$, $\pi(z) = h(z)$ for $z \in S_d(a)$ and $\pi(z) = h^{-1}(z)$ for $z \in S_d(b)$. Then $N_d^{\mathcal{A}[a,x]}(z) \cong N_d^{\mathcal{A}[b,x]}(\pi(z))$ for every z, and thus $\mathcal{A}[a,x] \leftrightarrows_d \mathcal{A}[b,x]$. □

Now the graph BDP follows: Consider a formula $\psi(x,y)$ and define Ψ as a sentence equivalent to $\forall x \forall y. U_1(x) \wedge U_2(y) \to \psi(x,y)$, where U_1 and U_2 are two new unary symbols. Let $d = \mathsf{hlr}(\Psi)$. Then for any $a \approx_{2d} b$ and $c \notin S_{2d+1}(a,b)$ we have

$$\begin{array}{rl} & \mathcal{A} \models \psi(a,c) \\ \Leftrightarrow & \mathcal{A}[a,c] \models \Psi \\ \Leftrightarrow & \mathcal{A}[b,c] \models \Psi \\ \Leftrightarrow & \mathcal{A} \models \psi(b,c) \end{array}$$

Thus, for any $a \approx_{2d} b$, we have $|out\text{-}deg(a) - out\text{-}deg(b)| \leq F_\sigma(k, 2d+1)$ where $\mathcal{A} \in \mathrm{STRUCT}_k[\sigma]$. Hence, the number of outdegrees in $\psi[\mathcal{A}]$ is at most

$f_\sigma(k, 2d) \cdot (2F\sigma(k, 2d+1)+1)$, that is, it depends on k, d and σ only. Since d only depends on ψ, we have the BDP. $\square$

This simple sketch contains all the main components of the proof of Theorem 3.1. These are: characterization of d-equivalence in terms of maps preserving d-neighborhoods, going from open formulae to sentences by adding extra unary predicates, and the fact that for each r, $\approx_r$-equivalence of $n+1$-tuples can be guaranteed by $\approx_d$-equivalence of n-tuples for sufficiently large d that depends on r only.

Now we sketch the proof of Theorem 3.1. Recall that $\sigma^{(n)}$ is σ extended with n new unary symbols $U_1, \ldots, U_n$. Let $\psi(x_1, \ldots, x_n)$ be a formula with n free variables. By $\Psi^{(n)}$ we denote a sentence in $\mathcal{L}$ that is equivalent to $\forall x_1 \ldots \forall x_n.(U_1(x_1) \wedge \ldots \wedge U_n(x_n)) \rightarrow \psi(x_1, \ldots, x_n)$; it exists since $\mathcal{L}$ is closed under first-order operations.

Lemma 3.10 *For any $\mathcal{A}$ and any n-vector $\vec{a}$, $\mathcal{A} \models \psi(\vec{a})$ iff $\mathcal{A}[\vec{a}] \models \Psi^{(n)}$.* $\square$

Let $\vec{a}$ be an n-vector. By $\vec{a}x$ we denote the $n+1$-vector whose first n components are those of $\vec{a}$ and the last one is x.

Lemma 3.11 *Let $\mathcal{A}$ be a σ-structure. Suppose $r > 0$ and $\vec{a} \approx_{3r+1} \vec{b}$. Then there exists a permutation π on A such that, for all $x \in A$, it is the case that $\vec{a}x \approx_r \vec{b}\pi(x)$.*

Proof sketch: We use the notation h^{-m} for $\underbrace{h^{-1} \circ \ldots \circ h^{-1}}_{m \text{ times}}$. A usual, h^m is h iterated m times.

Claim 3.12 *Let $\mathcal{A} \in$ STRUCT$[\sigma]$ and let $\vec{a} \approx_d \vec{b}$. Assume that $h : N_d(\vec{a}) \rightarrow N_d(\vec{b})$ is an isomorphism. Then for every $x \in S_d(\vec{b}) - S_d(\vec{a})$, there exists a number $m(x) > 0$ such that*

a) $h^{-m(x)}(x) \in S_d(\vec{a}) - S_d(\vec{b})$;

b) If $m(x) > 1$, then for every $0 < k < m(x)$, $h^{-k}(x) \in S_d(\vec{a}) \cap S_d(\vec{b})$.

Proof of Claim 3.12: Let $x \in S_d(\vec{b})$. Let $x_1 = h^{-1}(x_1) \in S_d(\vec{a})$. If $x_1 \notin S_d(\vec{b})$, then $m(x) = 1$ and we are done. Otherwise, $x_1 \in S_d(\vec{a}) \cap S_d(\vec{b})$. Consider $x_2 = h^{-1}(x_1)$. Again, if $x_2 \notin S_d(\vec{b})$, then $m(x) = 2$ and we are done; otherwise, $x_2 \in S_d(\vec{a}) \cap S_d(\vec{b})$.

Continuing this process we build a sequence $x_1, x_2, \ldots$ with $x_{i+1} = h^{-1}(x_i)$. There are two possibilities. First, for some $x_i \in S_d(\vec{a}) \cap S_d(\vec{b})$, $x_{i+1} \notin S_d(\vec{b})$. Then $m(x) = i+1$ and the claim is proved. Otherwise, we have that $h^{-i}(x) \in S_d(\vec{a}) \cap S_d(\vec{b})$ for all $i \in \mathbb{N}_+$. Since $\mathcal{A}$ is finite, find lexicographically least pair (i, j) with $i < j$ such that $h^{-i}(x) = h^{-j}(x)$ (where we assume $h^0(x)$ to be x). Since all $h^{-k}(x) \in S_d(\vec{a})$, $i > 0$ implies that we can apply h and get $h^{1-i}(x) = h^{1-j}(x)$, which contradicts minimality of (i, j). Thus, $i = 0$ and $x = h^{-j}(x)$ and j is the minimum such. But this is impossible since $j > 0$, $h^{-j}(x) \in S_d(\vec{a})$, but $x \in S_d(\vec{b}) - S_d(\vec{a})$. This proves the claim.

Reversing Claim 3.12, we see that for every $x \in S_d(\vec{a}) - S_d(\vec{b})$ there exists a number $k(x)$ such that $h^{k(x)}(x) \in S_d(\vec{b}) - S_d(\vec{a})$ and, for every $1 \leq j < k(x)$, $h^j(x) \in S_d(\vec{b}) \cap S_d(\vec{a})$. Now, using Claim 3.12 and its converse, we prove

Claim 3.13 *Let $\mathcal{A} \in$ STRUCT$[\sigma]$ and let $\vec{a} \approx_d \vec{b}$. Assume that $h : N_d(\vec{a}) \rightarrow N_d(\vec{b})$ is an isomorphism, and define $m(x)$ as in Claim 3.12. Now define $\pi : S_d(\vec{a}) \cup S_d(\vec{b}) \rightarrow S_d(\vec{a}) \cup S_d(\vec{b})$ as follows:*

$$\mu(x) \quad = \quad \begin{cases} h(x) & \text{if } x \in S_d(\vec{a}) \\ h^{-m(x)}(x) & \text{if } x \in S_d(\vec{b}) - S_d(\vec{a}) \end{cases}$$

Then μ is a permutation on $S_d(\vec{a}) \cup S_d(\vec{b})$. $\square$

We are now ready to finish the proof of the lemma. Let $\vec{a} \approx_{3r+1} \vec{b}$ and let $h : N_{3r+1}(\vec{a}) \rightarrow N_{3r+1}(\vec{b})$ be an isomorphism. Let $d = 2r+1$. Then h is also an isomorphism between $N_d(\vec{a})$ and $N_d(\vec{b})$. Define $\mu : S_d(\vec{a}) \cup S_d(\vec{b}) \rightarrow S_d(\vec{a}) \cup S_d(\vec{b})$ as in Claim 3.13, and then define a permutation π on A by

$$\pi(x) \quad = \quad \begin{cases} \mu(x) & \text{if } x \in S_d(\vec{a}) \cup S_d(\vec{b}) \\ x & \text{otherwise} \end{cases}$$

It can now be shown that $\vec{a}x \approx_r \vec{b}\pi(x)$ for all x. This finishes the proof of Lemma 3.11. $\square$

Lemma 3.14 *Suppose that in $\mathcal{A}$ we have $\vec{a}x \approx_r \vec{b}y$. Then*

$$N_r^{\mathcal{A}[\vec{a}]}(x) \quad \cong \quad N_r^{\mathcal{A}[\vec{b}]}(y)$$

Proof: Since carrier of both $\mathcal{A}[\vec{a}]$ and $\mathcal{A}[\vec{b}]$ is A, and U_is are unary, we have $S_r^{\mathcal{A}[\vec{a}]}(x) = S_r^{\mathcal{A}[\vec{b}]}(x) = S_r^{\mathcal{A}}(x)$ and similarly for y. Let $h : N_r^{\mathcal{A}}(\vec{a}x) \rightarrow N_r^{\mathcal{A}}(\vec{b}y)$ be an isomorphism; in particular, it maps $S_r^{\mathcal{A}}(x)$ onto $S_r^{\mathcal{A}}(y)$. Then one can show that h is an isomorphism between $N_r^{\mathcal{A}[\vec{a}]}(x)$ and $N_r^{\mathcal{A}[\vec{b}]}(y)$. $\square$

Now Theorem 3.1 follows: Let $\psi(x_1, \ldots, x_n)$ be given, and let $d = \mathsf{hlr}(\Psi^{(n)})$. Let $r = 3d+1$. We claim that $\mathsf{lr}(\psi) \le r$. Suppose $\mathcal{A}$ is a σ-structure, and $\vec{a} \approx_r \vec{b}$ in $\mathcal{A}$. By Lemma 3.11, we have a permutation $\pi : A \to A$ such that $\vec{a}x \approx_d \vec{b}\pi(x)$ for every $x \in A$. Thus, by Lemma 3.14, we have $N_d^{\mathcal{A}[\vec{a}]}(x) \cong N_d^{\mathcal{A}[\vec{b}]}(\pi(x))$, and now from proposition 3.7 we get that $\mathcal{A}[\vec{a}] \leftrightarrows_d \mathcal{A}[\vec{b}]$.

Since $d = \mathsf{hlr}(\Psi^{(n)})$, we have by Lemma 3.10:

$$\begin{aligned} & \mathcal{A} \models \psi(\vec{a}) \\ \Leftrightarrow\ & \mathcal{A}[\vec{a}] \models \Psi^{(n)} \\ \Leftrightarrow\ & \mathcal{A}[\vec{b}] \models \Psi^{(n)} \\ \Leftrightarrow\ & \mathcal{A} \models \psi(\vec{b}) \end{aligned}$$

which finishes the proof of locality of $\psi(\cdot)$. □

The proof above also shows that $\mathsf{lr}(\psi) \le 3 \cdot \mathsf{hlr}(\Psi^{(n)}) + 1$. In the case of first-order formulae, $\Psi^{(n)}$ increases the quantifier rank by n, and we obtain a new bound that improves Gaifman's $(7^{\mathsf{qr}(\psi)} - 1)/2$.

Corollary 3.15 *Let $\psi(x_1, \ldots, x_n)$ be a first-order formula. Then $\mathsf{lr}(\psi) \le 3^{\mathsf{qr}(\psi)+n} + 1$.* □

Note that this improves the locality rank implied by Gaifman's theorem, not the bound on the size of neighborhood in an *explicitly* constructed formula used in Gaifman's proof.

4 Strong Gaifman's locality implies Hanf's locality

The main result of the section is:

Theorem 4.1 *Let $\mathcal{L}$ be a logic that is closed under first-order operations. Assume that every sentence in $\mathcal{L}$ is strongly Gaifman-local. Then every sentence in $\mathcal{L}$ is Hanf-local.*

From this and Gaifman's theorem, the theorem by Fagin, Stockmeyer and Vardi follows immediately (though not the bound produced by the proof in [15]). We also believe that the proof, sketched below, is simpler than that in [15] and shows clearly why this result is indeed a form of locality, as claimed in [15].

Recall that $\mathcal{A} \leftrightarrows_d \mathcal{B}$ is equivalent to the existence of a bijection $\pi : A \to B$ such that $N_d^{\mathcal{A}}(x) \cong N_d^{\mathcal{B}}(\pi(x))$ for all x. Now we say that $\mathcal{A}$ and $\mathcal{B}$ are (n,d)-*equivalent*, denoted $\mathcal{A} \leftrightarrows_{n,d} \mathcal{B}$, if there is a bijection $\pi : A^n \to B^n$ such that for any n-vector $\vec{a}$ from A, $N_d^{\mathcal{A}}(\vec{a}) \cong N_d^{\mathcal{B}}(\pi(\vec{a}))$.

Our main technical tool is the proposition below. A similar idea was used in [26] to show that d-equivalence, for large enough d, guarantees a win for duplicator in the r-round *bijective* Ehrenfeucht-Fraisse game. Proposition 4.2 can also simplify the proof in [26].

Proposition 4.2 *Let $n > 0$ and $d > 0$. Then $\mathcal{A} \leftrightarrows_{n,3d+1} \mathcal{B}$ implies $\mathcal{A} \leftrightarrows_{n+1,d} \mathcal{B}$.*

Proof sketch: Suppose $\mathcal{A} \leftrightarrows_{n,3d+1} \mathcal{B}$. Then there exists a bijection $\mu : A^n \to B^n$ such that $N_{3d+1}^{\mathcal{A}}(\vec{a}) \cong N_{3d+1}^{\mathcal{B}}(\mu(\vec{a}))$. In particular, $card(A)^n = card(B)^n$ and thus $card(A) = card(B)$.

We now construct a bijection $\pi : A^{n+1} \to B^{n+1}$ as follows. With each vector $\vec{a} \in A^n$, associate a bijection $f_{\vec{a}} : A \to B$, and define $\pi(\vec{a}x)$ as $\mu(\vec{a})f_{\vec{a}}(x)$. (Recall that $\vec{a}x$ is an $n+1$-vector whose first n components are those of $\vec{a}$ and the last one is x.) To define $f_{\vec{a}}$, let $\vec{b} = \mu(\vec{a})$, and fix an isomorphism $h : N_{3d+1}^{\mathcal{A}}(\vec{a}) \to N_{3d+1}^{\mathcal{A}}(\vec{b})$. Since h maps $S_{2d+1}^{\mathcal{A}}(\vec{a})$ onto $S_{2d+1}^{\mathcal{A}}(\vec{b})$, we have $card(A - S_{2d+1}^{\mathcal{A}}(\vec{a})) = card(B - S_{2d+1}^{\mathcal{B}}(\vec{b}))$.

Consider an arbitrary type τ of a d-neighborhood of a single point. Assume that $a \in S_{2d+1}^{\mathcal{A}}(\vec{a})$ realizes τ in $\mathcal{A}$. Since h is an isomorphism of $3d+1$-neighborhoods, we see that $S_d^{\mathcal{A}}(a) \subseteq S_{3d+1}^{\mathcal{A}}(\vec{a})$ and thus $h(a) \in S_{2d+1}^{\mathcal{B}}(\vec{b})$ realizes τ. Thus, the number of elements in $S_{2d+1}^{\mathcal{A}}(\vec{a})$ and $S_{2d+1}^{\mathcal{B}}(\vec{b})$ that realize τ is the same. Note that $\mathcal{A} \leftrightarrows_{n,3d+1} \mathcal{B}$ implies $\mathcal{A} \leftrightarrows_d \mathcal{B}$. Therefore,

$$\begin{aligned} & card(\{a \in A - S_{2d+1}^{\mathcal{A}}(\vec{a}) \mid \tau_d(\mathcal{A}, a) = \tau\}) \\ = \ & card(\{b \in B - S_{2d+1}^{\mathcal{B}}(\vec{b}) \mid \tau_d(\mathcal{B}, b) = \tau\}) \end{aligned}$$

for any τ. Thus, we can find a bijection $g : A - S_{2d+1}^{\mathcal{A}}(\vec{a}) \to B - S_{2d+1}^{\mathcal{B}}(\vec{b})$ such that $N_d^{\mathcal{A}}(a) \cong N_d^{\mathcal{B}}(g(a))$ for any $a \in A - S_{2d+1}^{\mathcal{A}}(\vec{a})$. We now define $f_{\vec{a}}$ by

$$f_{\vec{a}}(x) \quad = \quad \begin{cases} h(x) & \text{if } x \in S_{2d+1}^{\mathcal{A}}(\vec{a}) \\ g(x) & \text{if } x \notin S_{2d+1}^{\mathcal{A}}(\vec{a}) \end{cases}$$

It is now easy to see that π defined by $\pi(\vec{a}x) = \mu(\vec{a})f_{\vec{a}}(x)$ is a bijection that satisfies $N_d^{\mathcal{A}}(\vec{a}_0) \cong N_d^{\mathcal{B}}(\pi(\vec{a}_0))$. Hence, $\mathcal{A} \leftrightarrows_{n+1,d} \mathcal{B}$. □

Immediately from here we obtain:

Corollary 4.3 *For any $r > 0$ and any $n \ge 1$ there exists a number d such that $\mathcal{A} \leftrightarrows_d \mathcal{B}$ implies $\mathcal{A} \leftrightarrows_{n,r} \mathcal{B}$.*

In fact, d can be taken to be $3^{n-1}r + (3^{n-2}-1)/2$ for $n > 1$ and $d = r$ for $n = 1$. □

Now the proof of Theorem 4.1 follows. It is enough to consider a sentence Ψ which is equivalent to $\exists x_1 \ldots \exists x_n . \psi(x_1, \ldots, x_n)$, where $\psi(\vec{x})$ is strongly Gaifman-local. Assume that r witnesses strong locality of ψ: that is, $N_r^{\mathcal{A}}(\vec{a}) \cong N_r^{\mathcal{B}}(\vec{b})$ implies $\mathcal{A} \models \psi(\vec{a})$ iff $\mathcal{B} \models \psi(\vec{b})$. Let d be given by Corollary 4.3. We claim that $\mathsf{hlr}(\Psi) \le d$. Indeed, assume $\mathcal{A} \leftrightarrows_d \mathcal{B}$. Let $\mathcal{A} \models \Psi$. Then $\mathcal{A} \models \psi(\vec{a})$ for some $\vec{a} \in A^n$. By Corollary 4.3, $\mathcal{A} \leftrightarrows_{n,r} \mathcal{B}$, and thus we find $\vec{b} \in B^n$ such that $N_r^{\mathcal{B}}(\vec{b}) \cong N_r^{\mathcal{A}}(\vec{a})$. From strong Gaifman-locality of ψ we see $\mathcal{B} \models \psi(\vec{b})$ and thus $\mathcal{B} \models \Psi$. The converse (that is, $\mathcal{B} \models \Psi$ implies $\mathcal{A} \models \Psi$) is similar. Hence, $\mathsf{hlr}(\Psi) \le d$, which completes the proof. □

Combining the proof above with Gaifman's theorem, we see that for an arbitrary *first-order* sentence Ψ, we have the bound $\mathsf{hlr}(\Psi) \le 2 \cdot 3^{\mathsf{qr}(\Psi)} \cdot 7^{\mathsf{qr}(\Psi)-1}$, which is much worse than $3^{\mathsf{qr}(\Psi)}$ that is given by [15]. However, it is not the bound itself, but its existence that is used in most applications. Also, the above proof reveals the close connection between Gaifman's and Hanf's theorems.

Another corollary of Theorem 4.1 is that the two parts of Gaifman's theorem are not independent:

Corollary 4.4 *Let $\mathcal{L}$ be a logic that is closed under first-order operations. Assume that every sentence in $\mathcal{L}$ is strongly Gaifman-local. Then every formula in $\mathcal{L}$ is Gaifman-local.* □

5 An application: expressiveness of a query language with aggregate functions

Most of traditional database theory deals with query languages that have well studied logical counterparts. For example, relational algebra has the power of first-order logic and Datalog with negation has the power of least-fixpoint logic (under inflationary semantics), see [1]. However, real query languages use some features that are not adequately captured by these logical formalisms. One of them is dealing with interpreted functions, and it was addressed recently [6]. The other is aggregation, which received a lot of attention in connection with studying bag semantics of query languages, cf. [9, 17, 18, 25]. Aggregation allows queries about a column in a relation as a whole, for example, the sum of all elements in a column.

First results on expressive power of aggregation appeared in [9], but they were based on an assumption of strict containment of some complexity classes. The first definitive proof appeared in [25], where an attempt was made to show that the relational language extended with aggregate functions has the bounded degree property for purely relational queries. Although this question was unanswered in [25], that paper did prove that connectivity and parity tests are not definable with the help of aggregation. The BDP was proved very recently [10], though the proof is far from satisfactory. It relies on a particular syntactic presentation of the language, and starts by proving a complicated normal form result that gets aggregation "out of the way". (Note that the idea itself seems to be essential for proving expressivity bounds for logics that count, see [5].) However, the intuition behind the proof of the normal form is far from obvious, and the proof does not extend to show locality of relational queries with aggregates.

In this section we prove, via an encoding in $\mathcal{FO} + COUNT$, that every relational query in a language with aggregates is Gaifman-local. Our technique has simple intuition behind it: all the counting happens "on the side" and does not affect locality. The proof depends less on a particular presentation of the language, because it only changes the encoding part.

Let us present the language, called $\mathcal{RA}^{\text{aggr}}(\mathbb{N})$, following [8, 18]. Assume the existence of two base types: type $\mathbb{N}$ of natural numbers, and an unspecified base type $\mathbf{b}$ whose domain is a countably infinite set D. A record type is of the form $t_1 \times \ldots \times t_n$, $n \ge 1$, where each t_i is $\mathbb{N}$ or $\mathbf{b}$; the semantics is n-tuples such that the ith component is of type t_i. We also consider the set type $\{t\}$, where t is restricted to be a record type; its objects are finite sets of objects of type t. Expressions of $\mathcal{RA}^{\text{aggr}}(\mathbb{N})$ are defined in Figure 1. Here s and t are record types, and T ranges over both record and set types.

The semantics is as follows (see [8, 18, 25] for detailed exposition): $+, \dot{-}, *$, *div*, *mod* are the standard operations on natural numbers; $K0$ and $K1$ return 0 and 1 respectively; $=$ is the equality test on base types $\mathbf{b}$ and $\mathbb{N}$ (true is represented by $\{0\}$ and false by $\{\}$); $<_{\mathbb{N}}$ is the usual order on the naturals; the semantics for identity, composition, tupling and projection is standard; $K\{\}$ always returns the empty set; *empty* tests if a set is empty; η forms singleton sets; $\cup$ is set union,

$$\frac{}{+,*,\dot{-},\mathit{div},\mathit{mod}:\mathbb{N}\times\mathbb{N}\to\mathbb{N}} \qquad \frac{}{K0,K1:T\to\mathbb{N}} \qquad \frac{}{=_{\mathbf{b}}:\mathbf{b}\times\mathbf{b}\to\{\mathbb{N}\}} \qquad \frac{}{=_{\mathbb{N}},<_{\mathbb{N}}:\mathbb{N}\times\mathbb{N}\to\{\mathbb{N}\}}$$

$$\frac{}{id:T\to T} \qquad \frac{f:u\to t \quad g:s\to u}{f\circ g:s\to t} \qquad \frac{f_i:t\to t_i, i=1,\ldots,n}{(f_1,\ldots,f_n):t\to t_1\times\ldots\times t_n} \qquad \frac{i\le n}{\pi_{i,n}:t_1\times\ldots\times t_n\to t_i}$$

$$\frac{}{K\{\}:T\to\{s\}} \qquad \frac{}{\mathit{empty}:\{t\}\to\{\mathbb{N}\}} \qquad \frac{}{\eta:t\to\{t\}} \qquad \frac{}{\cup:\{t\}\times\{t\}\to\{t\}}$$

$$\frac{f:T\times s\to\{t\}}{\mathit{ext}[f]:T\times\{s\}\to\{t\}} \qquad \frac{}{\mathit{cartprod}_n:\{t_1\}\times\ldots\times\{t_n\}\to\{t_1\times\ldots\times t_n\}} \qquad \frac{f:s\to\mathbb{N}}{\sum[f]:\{s\}\to\mathbb{N}}$$

Figure 1. Expressions of $\mathcal{RA}^{\mathrm{aggr}}(\mathbb{N})$

$\mathit{cartprod}_n$ is the cartesian product of n sets. Given functions $f : T \times s \to \{t\}$ and $g : s \to \mathbb{N}$, a set X of type $\{s\}$, and an object y of type T, $\mathit{ext}[f](y, X)$ evaluates to $\bigcup_{x\in X} f(y, x)$, and $\sum[g](X)$ is $\sum_{x\in X} g(x)$. For example, $\sum[K1](X)$ is $\mathit{card}(X)$.

Without the type of natural numbers, this language is equivalent to the relational algebra [8], and thus expresses precisely the first-order queries. Summation and arithmetic give it the power of aggregate functions; for example, the aggregate TOTAL (cf. [1]) is given by $\sum[id]$. Most commercial systems use a richer collection of aggregates by allowing rational arithmetic so that aggregates such as "average" can be defined. Here we prove the result for the language with natural arithmetic; we briefly address the problem of extending the results to rational arithmetic in the next section.

Commercial query languages also use grouping with aggregation; for example, one may ask ask for the average salary in each department. The use of grouping is modeled by *nesting* of sets. At first glance, it seems that sets of sets must be produced to answer this query. That is, the restriction that ts and ss in Figure 1 be record types is eliminated. However, such a nested language has the conservative extension property [24], which says that every query can be written using the height of set nesting not exceeding that of its input and output types. In particular, it means that every query whose input and output types are flat relations, can be written in $\mathcal{RA}^{\mathrm{aggr}}(\mathbb{N})$ (that is, without using nesting), even if it uses grouping. Thus, $\mathcal{RA}^{\mathrm{aggr}}(\mathbb{N})$ is a good candidate for modeling relational languages with aggregates.

Abbreviate $\mathbf{b} \times \ldots \times \mathbf{b}$, m times, as $\mathbf{b}^m$. Then a σ-structure is represented as an object of type $\{\mathbf{b}^{p_1}\} \times \ldots \times \{\mathbf{b}^{p_l}\}$, where σ has l relations of arities $p_1, \ldots, p_l$. We denote this type by $\sigma_{\mathbf{b}}$. Types of this form are called *relational*. A query in $\mathcal{RA}^{\mathrm{aggr}}(\mathbb{N})$ is *relational* if both its input and output types are.

We assume without loss of generality that the output of a relational query is one set of m-tuples. Then such a query is a mapping from σ-structures over D into finite subsets of D^m. It can be easily seen that for any such query Q, an element $d \in D$ occurs in a tuple in $Q(\mathcal{A})$ for some structure $\mathcal{A}$ with carrier A only if $d \in A$. Thus, we define $\psi_Q(x_1, \ldots, x_m)$ by letting $\mathcal{A} \models \psi_Q(\vec{a})$ iff $\vec{a} \in Q(\mathcal{A})$. We then say that Q is *Gaifman-local* if so is the associated formula ψ_Q.

We now prove the main result of the section:

Theorem 5.1 *Every relational query in $\mathcal{RA}^{\mathrm{aggr}}(\mathbb{N})$ is Gaifman-local.*

Proof sketch: Consider a relational query $Q : \sigma_{\mathbf{b}} \to \{\mathbf{b}^m\}$. Extend σ with one binary relational symbol S, and let p be a function on natural numbers. Define a query $Q_p : \sigma_{\mathbf{b}} \times \{\mathbf{b} \times \mathbf{b}\} \to \{\mathbf{b}^m\}$ as follows. Its input is a pair: a σ-structure $\mathcal{A}$ and a binary relation S. Let C be the set of elements in S, that is, the union of its first and second projections. Then Q_p is defined by

$$Q_p(\mathcal{A}, S) \;=\; \begin{cases} Q(\mathcal{A}) & \text{if } (*_{Q,p}) \text{ holds} \\ \emptyset & \text{otherwise,} \end{cases}$$

where $(*_{Q,p})$ is the following condition:

$$(*_{Q,p}) \qquad \begin{pmatrix} C \cap A = \emptyset, \text{ and} \\ \mathit{card}(C) \ge p(\mathit{card}(A)), \text{ and} \\ S \text{ is a linear order} \end{pmatrix}$$

Lemma 5.2 *If Q_p is Gaifman-local, then so is Q.*

Proof sketch: Let Q_p be local, and let $r = \mathsf{lr}(Q_p)$. Consider Q, its input (which is a structure $\mathcal{A}$), and let $\vec{a} \approx_r^{\mathcal{A}} \vec{b}$, where $\vec{a}$ and $\vec{b}$ are m-vectors of elements of A. Let $n = \mathit{card}(A)$ and let $n' > p(n)$. Let C be an n'-element subset of D such that $C \cap A = \emptyset$. Let S be an arbitrary linear ordering on C. We define $\mathcal{A}_S$ as

$\mathcal{A}$ extended with the binary relation S. Since $S_d^{\mathcal{A}_S}(\vec{a})$ does not contain any element of C, and neither does $S_d^{\mathcal{A}_S}(\vec{b})$ for any d, we obtain $\vec{a} \approx_r^{\mathcal{A}_S} \vec{b}$ and thus, by the locality of Q_p, $\vec{a} \in Q_p(\mathcal{A}_S)$ iff $\vec{b} \in Q_p(\mathcal{A}_S)$.

Note that all the conditions in $(*_{Q,p})$ hold, and thus $Q_p(\mathcal{A}_S) = Q(\mathcal{A})$. Hence, $\vec{a} \in Q(\mathcal{A})$ iff $\vec{b} \in Q(\mathcal{A})$, which proves that Q is local, and $\mathsf{lr}(Q) \leq r$. □

Lemma 5.3 *For every relational query Q in $\mathcal{RA}^{\text{aggr}}(\mathbb{N})$, there exists a polynomial p such that ψ_{Q_p} is definable in $\mathcal{FO} + COUNT$.*

Proof sketch: It can be shown that, for any $\mathcal{RA}^{\text{aggr}}(\mathbb{N})$ query there exist two polynomials p_1 and p_2 such that, for any input object x of size n, (a) $Q(x)$ contains at most $p_1(n)$ elements, and (b) every number produced in the process of evaluating Q is bounded by $p_2(n)$. From this we also derive that for each fixed number k, there exists a coding scheme of k-tuples of elements under $p_2(n)$ such that both coding and decoding are definable in $\mathcal{FO} + COUNT$ and there is a polynomial p such that all codes of tuples are bounded by $p(n)$.

Now we encode $\mathcal{RA}^{\text{aggr}}(\mathbb{N})$ objects in $\mathcal{A}_S$ by encoding elements of base type by themselves, and each number n by the nth element in the ordering S. Using this, we give a (rather long and tedious) encoding of $\mathcal{RA}^{\text{aggr}}(\mathbb{N})$ operators in $\mathcal{FO} + COUNT$. Note that both types $\mathbf{b}$ and $\mathbb{N}$ are coded as elements of the first sort, and the counting power is used to simulate operations of Q. By choosing p large enough so that all numbers produced during evaluating Q and all encodings of tuples are under $p(n)$, we guarantee that the simulation can be done in $\mathcal{FO} + COUNT$. It turns out that only coding of tuples of fixed length is needed. For more details, see the full version [23].

Now by Fact 2.7 and Theorem 3.1, Q_p is Gaifman-local, and thus Q is. This proves the theorem. □

Corollary 5.4 (see [10]) *Every relational query in $\mathcal{RA}^{\text{aggr}}(\mathbb{N})$ has the bounded degree property. Consequently, (deterministic) transitive closure is not definable in $\mathcal{RA}^{\text{aggr}}(\mathbb{N})$: $\mathcal{RA}^{\text{aggr}}(\mathbb{N}) \subset$ DLOGSPACE.* □

6 Conclusion

We examined the main notions of locality of first-order formulae, and proved that these notions are closely related. We showed that Hanf's locality implies Gaifman's locality for open formulae, and Gaifman's locality for sentences implies Hanf's locality. Each of these implies the bounded degree property, which is one of the easiest tools for proving inexpressibility results. Our presentation goes beyond the first-order case, and thus allows us to infer new results for logics with unary quantifiers and counting. We gave a much simplified proof that relational query language with aggregate functions cannot express the transitive closure; and we showed for the first time that its relational fragment is Gaifman-local.

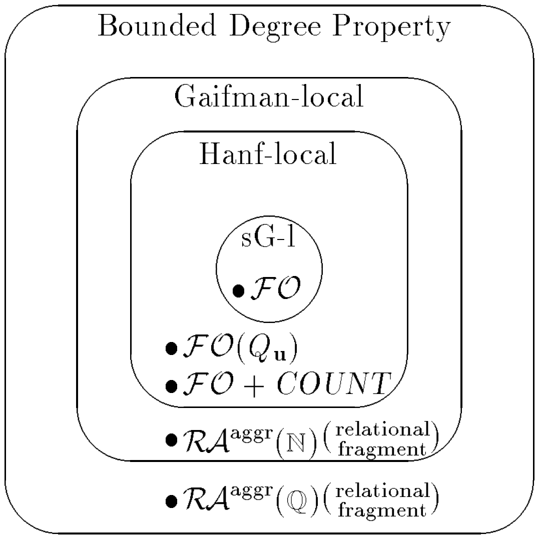

Figure 2. Summary of the notions of locality and relationship between them.

The results are summarized in Figure 2. We abbreviate strongly Gaifman-local by "sG-l". By $\mathcal{RA}^{\text{aggr}}(\mathbb{Q})$ we mean an extension of $\mathcal{RA}^{\text{aggr}}(\mathbb{N})$ to include rational arithmetic (that is, type $\mathbb{Q}$ and additional operations $-$ and $\div$). It follows from [10] that relational queries in $\mathcal{RA}^{\text{aggr}}(\mathbb{Q})$ have the BDP.

We mention two open problems. First, we would like to find more examples of Gaifman's locality, perhaps by extending techniques in [13, 15, 20, 26] to prove Hanf's locality for new languages. In terms of applications to database languages, we believe that Theorem 5.1 holds for $\mathcal{RA}^{\text{aggr}}(\mathbb{Q})$.

A much more challenging question is to extend these techniques to the ordered setting. The order relation gives easy counterexamples to all the forms of locality, so one should try to prove locality of *order-independent* queries, such as the transitive closure. Resolving these questions is probably as hard as separating complexity classes. Indeed, proving that order-independent queries in $\mathcal{FO} + COUNT + <$ are local would imply that TC^0 is strictly contained in DLOGSPACE. Also, one could try to show that $\mathcal{RA}^{\mathrm{aggr}}(\mathbb{N})$ can be evaluated with TC^0 data complexity. This would imply that $\mathcal{RA}^{\mathrm{aggr}}(\mathbb{N})$ with order cannot express transitive closure if and only if $\mathrm{TC}^0 \neq$ NLOGSPACE, making another interesting connection between separation of complexity classes and separation of query languages, in the spirit of [2].

Remark: After the first draft of this paper was written and submitted, I was informed by Lauri Hella and Juha Nurmonen that they had considered similar problems and obtained a number of closely related results, although they had not written them up. We are currently working on a joint journal paper that, in addition to clarifying the relationship between various notions of locality, will contain a number of new results (for example, characterizations of those notions of locality on structures of small degree.)

Acknowledgements: I thank Lauri Hella, Rick Hull, Juha Nurmonen, Limsoon Wong and anonymous referees for helpful comments.

References

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

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

[3] M. Ajtai and R. Fagin. Reachability is harder for directed than for undirected graphs. *Journal of Symbolic Logic*, 55(1):113–150, March 1990.

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

[5] M. Benedikt, H.J. Keisler. On expressive power of unary counters. *Proc. Int. Conf. on Database Theory*, Springer LNCS 1186, 1997, pages 291–305.

[6] M. Benedikt, L. Libkin. On the structure of queries in constraint query languages. In *LICS'96*, pages 25–34.

[7] M. Benedikt, T. Griffin and L. Libkin. Verifiable properties of database transactions. In *PODS'96*, pages 117–127.

[8] P. Buneman, S. Naqvi, V. Tannen, L. Wong. Principles of programming with complex objects and collection types. *TCS* 149 (1996), 3–48.

[9] M. Consens and A. Mendelzon. Low complexity aggregation in GraphLog and Datalog, *Theoretical Computer Science* 116 (1993), 95–116.

[10] G. Dong, L. Libkin, L. Wong. Local properties of query languages. *Proc. Int. Conf. on Database Theory*, Springer LNCS 1186, 1997, pages 140–154.

[11] H.-D. Ebbinghaus. Extended logics: the general framework. In J. Barwise and S. Feferman, editors, *Model-Theoretic Logics*, Springer-Verlag, 1985, pages 25–76.

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

[13] K. Etessami. Counting quantifiers, successor relations, and logarithmic space, *JCSS*, to appear. Extended abstract in *Structure'95.*

[14] R. Fagin. Easier ways to win logical games. In *Proc. DIMACS Workshop on Descriptive Complexity and Finite Models*, AMS 1997.

[15] R. Fagin, L. Stockmeyer, M. Vardi, On monadic NP vs monadic co-NP, *Information and Computation*, 120 (1994), 78–92.

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

[17] S. Grumbach and T. Milo. Towards tractable algebras for bags. *JCSS*, 52 (1996), 570–588.

[18] S. Grumbach, L. Libkin, T. Milo and L. Wong. Query languages for bags: expressive power and complexity. *SIGACT News*, 27 (1996), 30–37.

[19] 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.

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

[21] N. Immerman. Descriptive complexity: A logician's approach to computation. *Notices of the AMS* 42 (1995), 1127–1133.

[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] L. Libkin. On the forms of locality over finite models. Bell Labs, Technical Memo, 1996.

[24] L. Libkin and L. Wong. Aggregate functions, conservative extension, and linear orders. *In* Proc. 4th Internat. Workshop on Database Programming Languages, New York, 1993, Springer Verlag, pages 282–294.

[25] L. Libkin, L. Wong. New techniques for studying set languages, bag languages, and aggregate functions. In *PODS'94*, pages 155–166. Full version to appear in *JCSS.*

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

[27] J. Nurmonen. Unary quantifiers and finite structures. PhD Thesis, University of Helsinki, 1996.

[28] T. Schwentick. Graph connectivity and monadic NP. In *Proceedings of 35th Symposium on Foundations of Computer Science*, pages 614–622, 1994.