

Hanf Locality and Invariant Elementary Definability *

Steven Lindell
Haverford College

Henry Towsner
University of Pennsylvania

Scott Weinstein
University of Pennsylvania

July 17, 2025

Abstract

We introduce some notions of invariant elementary definability which extend the notions of first-order order-invariant definability, and, more generally, definability invariant with respect to arbitrary numerical relations. In particular, we study invariance with respect to expansions which depend not only on (an ordering of) the universe of a structure, but also on the particular relations which determine the structure; we call such expansions *presentations* of a structure. We establish two locality results in this context. The first is an extension of the original Hanf Locality Theorem to boolean queries which are invariantly definable over classes of locally finite structures with respect to *elementary, neighborhood-bounded* presentations. The second is a non-uniform version of the Fagin-Stockmeyer-Vardi Hanf Threshold Locality Theorem to boolean queries which are invariantly definable over classes of bounded degree structures with respect to elementary, neighborhood-bounded, *local* presentations.

1 Introduction

Locality theorems characterize various senses in which properties of a relational structure that are definable in a particular way depend only on “local information” about the structure, typically, isomorphism types of finite radius neighborhoods in the Gaifman-graph of the structure. Such theorems provide simple and perspicuous proofs of descriptive lower bound results, that is, proofs that various combinatorial properties cannot

*Henry Towsner is partially supported by NSF grant DMS-1340666.

be defined with specified logical resources (see, Gaifman (1982), Fagin et al. (1995), Nurmonen (1996), Hella et al. (1999)). They are also central to the proofs of various preservation theorems over classes of finite structures (see Atserias et al. (2006), Atserias et al. (2008), Dawar (2010)) and of numerous algorithmic metatheorems (see Grohe and Kreutzer (2011)). In this paper, we explore Hanf locality of boolean queries which are invariantly definable on classes of (locally) finite structures.

We introduce some notions of invariant elementary definability which extend the notions of first-order order-invariant definability, and, more generally, definability invariant with respect to arbitrary numerical relations; we study invariance with respect to expansions which depend not only on (an ordering of) the universe of a structure, but also on the particular relations which determine the structure. We call such expansions *presentations* of a structure. Examples of presentations are breadth-first and depth-first traversals of finite graphs, and local orders of locally finite graphs. We establish two locality results in this context. The first, Theorem 4.6, is an extension of the original Hanf Locality Theorem to boolean queries which are invariantly definable over classes of locally finite structures with respect to *elementary, neighborhood-bounded* presentations. The second, Theorem 4.8, is a non-uniform version of the Fagin-Stockmeyer-Vardi Hanf Threshold Locality Theorem to boolean queries which are invariantly definable over classes of bounded degree structures with respect to *elementary, neighborhood-bounded, local* presentations.

Related Work The study of locality properties of first-order logic began with Hanf (1965), who proved that relational structures which are fully locally equivalent up to a countable multiplicity threshold are elementarily equivalent. Gaifman (1982) analyzed locality of relations definable within a single structure and established local normal forms for first-order formulas that have application to descriptive lower bounds over classes of finite structures as well as infinite structures. Fagin et al. (1995) established a refinement of the locality theorem of Hanf (1965) which shows that for each first-order sentence θ , there is a locality radius r_θ , depending only on the quantifier rank of θ , such that structures locally equivalent up to radius r_θ and up to a countable multiplicity threshold, are θ -equivalent. Moreover, they showed that over structures of bounded degree d , the multiplicity threshold depends exponentially on d . Grohe and Schwentick (2000) address the question of locality of queries which are not first-order definable over finite structures. In particular, they establish Gaifman locality of queries which are order-invariant first-order definable over finite structures. Hella et al. (1999) provides an extensive and detailed study of the relations among several notions of locality and establishes locality results for several extensions of first-order logic, among them FO+COUNT. The paper also provides a counterexample to Gaifman locality

of queries definable order-invariantly over finite structures by FO+COUNT-formulas, and suggests the possibility that such a locality result might hold non-uniformly in the size of a structure, that is, when the radius of locality is allowed to depend not only on the quantifier rank of a formula defining the query but also on the cardinality of the structure over which it is evaluated. Anderson et al. (2012) establishes just such a non-uniform form of Gaifman locality for queries definable in FO by formulas invariant with respect to arbitrary numerical predicates over finite structures, that is, predicates whose interpretation depends only on an arbitrary linear order of the universe; in particular, they demonstrate Gaifman locality for any query first-order definable invariantly with respect to arbitrary numerical predicates, where the radius of locality depends polylogarithmically on the size of the structure over which it is evaluated. Anderson et al. (2012) also establishes a similar non-uniform form of Hanf locality for queries definable by first-order formulas invariant with respect to arbitrary numerical predicates over string structures, again with radius of locality depending polylogarithmically on the size of the structure. The question of such a Hanf locality result for arbitrary finite structures remains open. In this paper we explore Hanf locality of queries which are definable over some “tame” classes of structures (locally finite structures and bounded degree structures) invariantly with respect to certain “tame” presentations (neighborhood bounded and local presentations). We consider presentations of a structure other than built-in numerical relations, in that they depend (elementarily) on relations of the structure other than an arbitrary ordering of its universe. In this context, we establish a Hanf type locality result for invariant definability over locally finite structures, and a Hanf threshold type result for invariant definability over bounded degree structures. The latter result is “non-uniform”, in the sense that the radius of locality depends not only on the defining formulas, but also on the degree of structures over which it is evaluated. Our threshold theorem applies to boolean queries which are invariantly definable via local-orders over structures with Gaifman graphs of bounded degree; invariant definability via orders and local-orders over “tame” classes of finite graphs, such as graphs of bounded degree and bounded tree-width, has been studied in Benedikt and Segoufin (2009) where certain “collapse” results were established.

2 Preliminaries

The next definition formulates some topological notions requisite for describing locality precisely. We write \mathfrak{K}_τ (\mathfrak{F}_τ) for the collection of (finite) structures of finite relational similarity type τ , that is, τ consists of finitely many relation and constant symbols.

Definition 2.1. *Let G be a simple graph, $a, b \in G$ and $r \geq 0$.*

- (a) We write $\delta_G(a, b)$ for the distance between a and b in G .
- (b) $B_r(G, a) = \{c \in G \mid \delta_G(a, c) \leq r\}$, for $r \in \omega$. $B_\infty(G, a) = \bigcup_{r \in \omega} B_r(G, a)$.
- (c) Let $A \in \mathfrak{K}_\tau$. The Gaifman-graph of A (denoted G_A) is the simple graph whose node set is A and whose edge relation holds between a pair of nodes a, b with $a \neq b$ if and only if for some $R \in \tau$ and tuple $\bar{c} \in R^A$ there are i, j with $c_i = a$ and $c_j = b$.
- (d) We say a structure A has degree bounded by d (finite degree) if and only if for every $a \in A$ the degree of a in G_A is at most d (degree of a in G_A is finite). We write \mathfrak{K}_τ^d (\mathfrak{F}_τ^d) for the collection of (finite) τ -structures with degree bounded by $d \in \omega$, and $\mathfrak{K}_\tau^{<\omega}$ for the collection of τ -structures of finite degree. We will suppress the subscript τ when it is clear from the context.
- (e) Let $r \in \omega$ or $r = \infty$, and let $a \in A$.

$$A^r(a) = \langle (A \upharpoonright B_r(G_A, a)), a \rangle.$$

The structure $A^r(a)$ is called the pointed r -neighbor-hood of a in A . Note that $A^\infty(a)$ is the substructure of A induced by the connected component of a in G_A .

$\mathfrak{K}_\tau^{<\omega}$ is also known as the class of *locally finite* τ -structures, because $A \in \mathfrak{K}_\tau^{<\omega}$ if and only if for every $a \in A$ and every $r \in \omega$, $B_r(G_A, a)$ is finite. Note that if A is locally finite, then for every $a \in A$, $B_\infty(G_A, a)$ is either finite or countably infinite.

Let $t \in \omega \cup \{\omega\}$. We say sets X and Y are equipollent up to threshold t (written $X \sim_t Y$) if and only if either $|X| = |Y|$ or $|X|, |Y| \geq t$.

Let $\mathfrak{J} \subseteq \mathfrak{K}_\tau$ and $A, B \in \mathfrak{K}_\tau$. We write $A \equiv_{\mathfrak{J}} B$ if A and B agree on \mathfrak{J} , that is, $(A \in \mathfrak{J} \iff B \in \mathfrak{J})$, and similarly, $A \equiv_\theta B$ for $(A \models \theta \iff B \models \theta)$, when θ is a first-order sentence in signature τ . $A \equiv B$ (A is elementarily equivalent to B) if and only if $A \equiv_\theta B$, for all first-order sentences θ .

Definition 2.2. Suppose $A, B \in \mathfrak{K}_\tau$ and $\mathfrak{K} \subseteq \mathfrak{K}_\tau$.

- (a) Let $r \in \omega$ and $t \in \omega \cup \{\omega\}$. A is Hanf threshold r, t -equivalent to B (written $A \approx_{r,t} B$) if and only if for every pointed τ -structure C ,

$$\{a \in A \mid A^r(a) \cong C\} \sim_t \{b \in B \mid B^r(b) \cong C\},$$

that is, each isomorphism type of pointed r -neighbor-hood is realized the same number of times in A and B up to threshold t .

- (b) A class of τ -structures \mathfrak{J} is Hanf r, t -local on \mathfrak{K} if and only if for all $A, B \in \mathfrak{K}$, if $A \approx_{r,t} B$, then $A \equiv_{\mathfrak{J}} B$, that is, \mathfrak{J} is closed under r, t -equivalence with respect to structures in \mathfrak{K} .
- (c) Let $r \in \omega$. A is Hanf r -equivalent to B (written $A \approx_r B$) if and only if $A \approx_{r,\omega} B$. (Thus, if A and B are countable and Hanf r -equivalent, there is an isomorphism preserving bijection between their pointed r -neighborhoods.)
- (d) A class of τ -structures \mathfrak{J} is Hanf r -local on \mathfrak{K} if and only if for all $A, B \in \mathfrak{K}$, if $A \approx_r B$, then $A \equiv_{\mathfrak{J}} B$, that is, \mathfrak{J} is closed under r -equivalence with respect to structures in \mathfrak{K} .
- (e) A is Hanf equivalent to B (written $A \approx B$) if and only if for all $r \in \omega$, $A \approx_r B$, that is, A and B are fully locally equivalent up to threshold ω .
- (f) A class of τ -structures \mathfrak{J} is Hanf local on \mathfrak{K} if and only if for all $A, B \in \mathfrak{K}$, if $A \approx B$, then $A \equiv_{\mathfrak{J}} B$, that is, \mathfrak{J} is closed under Hanf equivalence with respect to structures in \mathfrak{K} .

3 Hanf Locality

In general, elementary boolean queries are not Hanf local on arbitrary structures. For example, the simple graphs G and H consisting of a single countable clique and the disjoint union of two countable cliques are distinguished by a first-order sentence, yet $G \approx H$. On the other hand, Hanf (1965) showed that every elementary class, that is, a class defined by a set of first-order sentences, is Hanf local on the class of locally finite structures.

Theorem 3.1 (Hanf). *For every τ and $A, B \in \mathfrak{K}_{\tau}^{<\omega}$, if $A \approx B$, then $A \equiv B$.*

Note that for finite structures A and B , $A \approx B$ if and only if $A \cong B$; thus Hanf equivalence is not a useful tool for analyzing the expressive power of logical languages over structures which are actually finite, as opposed to merely locally finite. Fagin et al. (1995) proved that classes of structures defined by single first-order sentences θ over \mathfrak{F}_{τ} are Hanf r -local, for large enough r , depending only on the quantifier rank of θ .

Theorem 3.2 (Fagin, Stockmeyer, & Vardi). *For every τ and for every first-order sentence θ with signature τ , there is an $r \in \omega$ such that for all $A, B \in \mathfrak{F}_{\tau}$, if $A \approx_r B$, then $A \equiv_{\theta} B$.*

Theorem 3.2 provides a valuable technique for establishing that many well-known combinatorial properties of finite structures are not first-order definable. Fagin et al. (1995) derived Theorem 3.2 from the following threshold locality theorem for first-order logic.

Theorem 3.3 (Fagin, Stockmeyer, & Vardi). *For every τ and for every first-order sentence θ with signature τ , there is an $r \in \omega$ such that for all $d \in \omega$, there is a $t \in \omega$, such that for all $A, B \in \mathfrak{F}_\tau^d$, if $A \approx_{r,t} B$, then $A \equiv_\theta B$.*

Here too, the locality radius may be chosen to depend only on the quantifier rank of θ , whereas the threshold t depends also on the degree bound d on the Gaifman graphs of the structures. It is easy to verify that the argument of Fagin et al. (1995) suffices to establish Theorems 3.3 for arbitrary structures of bounded degree, not just finite structures, that is, for \mathfrak{K}_τ^d in place of \mathfrak{F}_τ^d . Libkin (2004) presents a direct proof of Theorem 3.2, with generalizations to more powerful logical languages, and much illuminating discussion of locality and its applications.

4 Hanf Locality of Queries Invariantly Definable over Locally Finite Structures

In logic, a mathematical object, such as a simple graph, is typically identified with an abstract relational structure: in the case of a simple graph, this consists of a set of vertices paired with an irreflexive and symmetric edge relation between them. But this identification is hardly universal throughout mathematics and computer science. For example, a graph is often presented as a drawing of nodes in the plane together with arcs joining them to indicate the edges, or as an adjacency matrix whose axes are labelled by the vertices in some arbitrary order and whose entries indicate the presence or absence of an edge. In this section, we will focus attention on the connection between an abstract structure and such presentations of it. In particular, we will investigate invariant elementary definability over classes of finite and of locally finite structures with respect to various such schemes of presentation.

A special case of presentation invariant definability has been studied in the context of database theory. The “Data Independence Principle” (see Abiteboul et al. (1995)) states that the results of a database query must be independent of (that is, invariant with respect to) the arbitrary order via which the entries in the database are encoded. The importance of this condition has given rise to sustained investigation of the properties of queries that are (linear) order-invariant definable over finite structures in various logical languages that underly database query languages. Linear orders are a very special kind

of presentation of a structure that involves only the underlying universe of the structure while ignoring its relational component. Presentations of this general character are often called “built-in numerical relations,” and locality of queries invariantly definable with respect to such presentations has been studied intensively in finite model theory (see, for example, Grohe and Schwentick (2000) and Anderson et al. (2012)). In this section we will introduce more general schemes of presentation.

Definition 4.1. *Let $A \in \mathfrak{K}_\tau$ and R be a relation symbol (of some fixed but unspecified arity).*

- (a) *An R -presentation of A is an expansion of A to a structure A^* of similarity type $\tau \cup \{R\}$.*
- (b) *An R -presentation scheme \mathbb{P} for a class of structures $\mathfrak{K} \subseteq \mathfrak{K}_\tau$ is a collection of R -presentations of members of \mathfrak{K} that contains at least one presentation of each member of \mathfrak{K} . We sometimes write $A^* \mathbb{P} A$ to abbreviate $A^* \in \mathbb{P}$ and A^* is an R -presentation of A .*
- (c) *An R -presentation scheme \mathbb{P} of a collection of structures $\mathfrak{K} \subseteq \mathfrak{K}_\tau$ is an elementary presentation scheme if and only if there is a first order sentence θ such that for every $A \in \mathfrak{K}$ and every R -presentation A^* of A ,*

$$A^* \in \mathbb{P} \iff A^* \models \theta.$$

Example 4.2. *Let \mathfrak{G} be the collection of finite simple graphs.*

- (a) *For every $G \in \mathfrak{G}$, $\langle G, < \rangle \in \mathbb{L}(\mathfrak{G})$ if and only if $<$ linearly orders the vertices of G . We call this the linear presentation of finite graphs.*
- (b) *For every $G \in \mathfrak{G}$, $\langle G, < \rangle \in \mathbb{T}(\mathfrak{G})$ if and only if $<$ is a traversal of G . (Recall that $<$ is a traversal of a graph G just in case it is a linear ordering of the vertices of G with the additional property that the connected components of G occupy disjoint intervals, and that in each such interval, every node but the first has a preceding neighbor. It is well-known (see Corneil and Krueger (2008)) that $<$ is a traversal of a graph G if and only if it satisfies the following first-order condition: for all $a, b, c \in G$, if $a < b < c$ and Eac , then there is a $d \in G$ such that $d < b$ and Edb .) We call this the traversal presentation of finite graphs.*
- (c) *A ternary relation O on the vertex set of a graph G locally orders G if and only if for every $a \in G$, the binary relation $\{\langle b, c \rangle \mid Oabc\}$ is a strict linear ordering of the G -neighbors of a . For every $G \in \mathfrak{G}$, $\langle G, O \rangle \in \mathbb{O}(\mathfrak{G})$ if and only if O locally orders G . We call this the local order presentation of finite graphs.*

It is easy to see that all these examples are elementary presentation schemes. Moreover, each of these examples can be extended to collections of finite τ -structures via their Gaifman graphs. We next explore the power of presentations to reveal information about the structures they present. We characterize this power in terms of presentation-invariant elementary definability.

Definition 4.3. *Let \mathbb{P} be an elementary R -presentation scheme for a class of structures $\mathfrak{K} \subseteq \mathfrak{K}_\tau$.*

- (a) *A first order sentence θ in the signature $\tau \cup \{R\}$ is \mathbb{P} -invariant over \mathfrak{K} if and only if for every $A \in \mathfrak{K}$, $(\exists A^*)(A^*\mathbb{P}A \wedge A^* \models \theta)$ if and only if $(\forall A^*)(A^*\mathbb{P}A \rightarrow A^* \models \theta)$.*
- (b) *$\mathfrak{J} \subseteq \mathfrak{K}$ is \mathbb{P} -invariant elementary on \mathfrak{K} if and only if there is a first order sentence θ that is \mathbb{P} -invariant on \mathfrak{K} and for every $A \in \mathfrak{K}$*

$$A \in \mathfrak{J} \iff (\exists A^*)(A^*\mathbb{P}A \wedge A^* \models \theta).$$

It follows from the the Beth Definability Theorem that a boolean query is \mathbb{P} -invariant elementary over the class of arbitrary structures if and only if it is elementary, for any elementary presentation scheme \mathbb{P} . On the other hand, a well-known example due to Gurevich (first exposed in Exercise 17.27 of Abiteboul et al. (1995)), namely the collection of *finite* boolean algebras with an even number of atoms, shows that there are queries \mathbb{L} -invariant elementary over the collection of finite structures that are not elementary. Observe that this collection is also \mathbb{O} -invariant elementary. It is easy to see that connectivity is \mathbb{T} -invariant elementary over the the class of finite simple graphs. On the other hand, Gurevich (1984) shows that connectivity is not \mathbb{L} -invariant elementary.

As the example of traversals demonstrates, we will need to consider “tame” classes of elementary presentations in order to establish locality of invariantly definable boolean queries. The first tameness condition we introduce, neighborhood boundedness, excludes presentations which themselves have “global reach”, that is, neighbors of a point in the (Gaifman graph) of the presentation of a structure cannot be arbitrarily far way from that point in the (Gaifman graph) of the structure itself.

Definition 4.4. *Suppose \mathbb{P} is a presentation scheme for the class of all locally finite τ -structures $\mathfrak{K}_\tau^{<\omega}$.*

- (a) *We say \mathbb{P} is neighborhood bounded if and only if there is a constant ν (called the neighborhood expansion factor for \mathbb{P}) such that for every $A \in \mathfrak{K}_\tau^{<\omega}$, for every A^* such that $A^*\mathbb{P}A$, and for every $a \in A$, $B_1(G_{A^*}, a) \subseteq B_\nu(G_A, a)$.*

(b) We say \mathbb{P} is degree bounded if and only if there is a function $g : \omega \mapsto \omega$ (called a degree bound for \mathbb{P}) such that for every d, A, A^* , if $A \in \mathfrak{K}_\tau^d$ and $A^* \mathbb{P} A$, then $A^* \in \mathfrak{K}_{\tau \cup R}^{g(d)}$.

Observe that if \mathbb{P} is neighborhood bounded with neighborhood expansion factor ν then for every $A \in \mathfrak{K}_\tau^{<\omega}$, every A^* such that $A^* \mathbb{P} A$, every $a \in A$, and every $r \in \omega$, $B_r(G_{A^*}, a) \subseteq B_{\nu \cdot r}(G_A, a)$. Note also, that if \mathbb{P} is neighborhood bounded, then for every $A \in \mathfrak{K}_\tau^{<\omega}$ and for every A^* such that $A^* \mathbb{P} A$, $A^* \in \mathfrak{K}_{\tau \cup \{R\}}^{<\omega}$.

Note \mathbb{O} , the presentation scheme of locally ordered finite graphs, is neighborhood bounded (with neighborhood expansion factor $\nu = 2$), while \mathbb{L} , the presentation scheme of linearly ordered finite graphs, is not (the Gaifman graph of a linearly ordered set is a clique). As the results of Grohe and Schwentick (2000) and Anderson et al. (2012) demonstrate, the failure of a presentation to be neighborhood bounded does not preclude locality of boolean queries invariantly definable with respect to it.

Remark 4.5. *Observe that every neighborhood bounded presentation \mathbb{P} is degree bounded. In particular, suppose that ν is the neighborhood expansion factor of \mathbb{P} , and that $A \in \mathfrak{K}_\tau^d$ and $A^* \mathbb{P} A$. For any point $a \in A^*$, the degree of a is just the size of the neighborhood $B_1(A^*, a) - 1$. But, this is bounded by the size of the neighborhood $B_\nu(A, a) - 1$ which is bounded by d^ν .*

A crucial tool in the proof of Theorem 4.6 will be the use of the *local type* of a point, that is, the collection of formulas with all quantifiers bounded to some r -neighborhood of the point. Note that for any finite relational signature τ , there is a first-order formula $\delta_\tau^r(x, y)$ such that for every τ -structure A and $a, b \in A$, $A \models \delta_\tau^r[a, b]$ if and only if the distance between a and b in G_A is at most r . If $\varphi(x_1, \dots, x_k)$ is a formula, we write $\varphi^r(x_1, \dots, x_k)$ for the formula which results from $\varphi(x_1, \dots, x_k)$ by relativizing all its quantifiers (Qy) to the formula $\delta_\tau^r(x_1, y) \vee \dots \vee \delta_\tau^r(x_k, y)$. We call such a formula an *r -local formula*. If \bar{a} is a k -tuple of elements of A , the *r -local type* of \bar{a} in A is the set of formulas $\varphi^r(\bar{x})$ such that $A(\bar{a}) \models \varphi^r(\bar{a})$. Note that for structures $A \in \mathfrak{K}_\tau^{<\omega}$ and $a \in A$, the r -local type of a in A is determined by a single r -local formula which characterizes the *finite* structure $A^r(a)$ up to isomorphism. We write $\sigma^r(a, A)(x)$ for the r -local type of a in A . The *local type* of \bar{a} in A is the union of the r -local types of \bar{a} in A , for all $r \in \omega$. We write $\sigma^\infty(a, A)$ for the local type of a in A . It is easy to see that the connected component of a point determines its local-type, since any quantifiers in an r -local formula range only over the connected component; Theorem 4.6 exploits the fact that the converse holds in locally finite structures.

The next theorem establishes that the original form of Hanf's Theorem extends to boolean queries that are first-order invariantly definable with respect to neighborhood

bounded presentations, that is, if two structures are Hanf equivalent for every finite radius up to countable threshold, then they agree on every such query.

Theorem 4.6. *Let \mathbb{P} be an elementary neighborhood bounded presentation and $\mathfrak{J} \subseteq \mathfrak{K}_r$ a \mathbb{P} -invariant elementary boolean query on $\mathfrak{K}_r^{\leq \omega}$. If $A, B \in \mathfrak{K}_r^{\leq \omega}$ and $A \approx B$, then $A \equiv_{\mathfrak{J}} B$.*

Proof. To show this, we need to observe that there is a “universal” structure U in the Hanf-class of A , that is, the class of structures C with $A \approx C$. Note that, in a locally finite structure A , the local type of a point a determines its connected component (in G_A , the Gaifman graph of A). Indeed, for each $r \in \omega$, $\sigma^r(a, A)$ determines the isomorphism type of the finite structure $A^r(a)$. Thus, if A and B are locally finite, with $\sigma^\infty(A, a) = \sigma^\infty(B, b)$, there is a finitely branching tree of isomorphisms from the neighborhoods $A^r(a)$ onto the neighborhoods $B^r(b)$, and thence, by the König Infinity Lemma, an isomorphism from the pointed component $A^\infty(a)$ onto the pointed component $B^\infty(b)$. Therefore, in particular, if $A \approx B$ and $a \in A$ is a point such that, for some r , $A^r(a)$ appears finitely many times in A , then B must contain connected components isomorphic to the connected component of a with exactly the same multiplicity as they occur in A .

Consider the connected components of A ; we may divide them into two kinds: those components which appear finitely often, and those which appear infinitely often. We define the universal structure U to consist of each of the components of A which appears finitely many times in A (and the same number of each such component as in A) together with countably many of each component such that every finite-radius neighborhood of this component appears infinitely often in A . (Note that such a component might not, itself, appear in A .) Note that U would be the same if we constructed it based on B instead. Therefore, by the assumption that \mathfrak{J} is a \mathbb{P} -invariant elementary boolean query, it suffices to show that, given a presentation $A^*\mathbb{P}A$, there is a presentation $U^*\mathbb{P}U$ so that $A^* \approx U^*$.

We construct U^* as follows. First, we simply copy each component of A^* to an identical component of U^* (which can be done since there is an embedding of A into U). Next, consider some component C of U which has not been assigned a presentation. It suffices to assign a \mathbb{P} -presentation C^* to C such that each r -neighborhood in the presentation is present infinitely often in A^* . Take any point $u \in C$ and let ν be the neighborhood expansion factor of \mathbb{P} . Note first that for every r , $C^{\nu \cdot r}(u)$ appears infinitely often in A , and, since A^* is locally finite, there are, up to isomorphism, only finitely many presentations of $C^{\nu \cdot r}(u)$, so at least one of these presentations appears infinitely often in A^* . It follows at once that we may inductively choose a sequence of \mathbb{P} -presentations $C^{*r}(u)$ such that for every r , $C^{*r}(u) \subseteq C^{*(r+1)}(u)$ and $C^{*r}(u)$ occurs,

up to isomorphism, infinitely often in A^* . Finally, we let $C^* = \bigcup_{r \in \omega} C^{*r}(u)$. \square

Theorem 4.6 is only of interest in connection with *infinite*, locally finite structures, such as arise in some approaches to software and hardware verification via model-checking; for finite A and B , and “logical” presentations \mathbb{P} , if $A \approx B$, then $A \cong B$, thence if $A^* \mathbb{P} A$ and $B^* \mathbb{P} B$, then $A^* \cong B^*$. We do not know whether a Hanf locality result with finite threshold bound holds for boolean queries which are \mathbb{P} -invariant elementarily definable with respect to neighborhood bounded presentations in general. The next definition introduces a very natural class of *local* presentation schemes, which allows us to establish a non-uniform Hanf threshold locality result for boolean queries invariantly definable both over finite, and over locally finite, structures.

Definition 4.7. Suppose $\mathfrak{K} \subseteq \mathfrak{K}_\tau$ is closed under substructures and \mathbb{P} is a presentation scheme for \mathfrak{K} . We say \mathbb{P} is *local* if and only if

- (a) for all $A \in \mathfrak{K}$, all A^* with $A^* \mathbb{P} A$, and all finite $B \subseteq A$, $(A^* \upharpoonright B) \mathbb{P} B$, and
- (b) for all $A \in \mathfrak{K}$, all $B, C \subseteq A$, and all finite B^*, C^* , if $B \cap C = \emptyset$ and $B^* \mathbb{P} B$ and $C^* \mathbb{P} C$, then there is an A^* such that $A^* \mathbb{P} A$ and $(A^* \upharpoonright B) = B^*$ and $(A^* \upharpoonright C) = C^*$.

We call condition (a) of Definition 4.7 *localization*: if we restrict a presentation of a structure to a finite substructure, it is a presentation of that substructure. We call condition (b) of Definition 4.7 *disjoint local amalgamation*: presentations of disjoint finite pieces of a structure may be combined in a presentation of the entire structure. It is easy to verify that both \mathbb{L} and \mathbb{O} are local. On the other hand, it is clear that \mathbb{T} , the presentation scheme of traversals of finite graphs, has neither localization nor disjoint local amalgamation. By modifying \mathbb{O} to the (elementary) presentation scheme of “local circular successor” on locally finite graphs, we have an example of a presentation with disjoint local amalgamation that lacks localization. On the other hand, the (elementary) presentation “component coloring”, which requires each component to lie within, or be disjoint from, a given unary predicate, satisfies localization, but violates disjoint amalgamation. Note that both “local circular successor” and “component coloring” are neighborhood bounded.

The next result shows that boolean queries invariantly definable with respect to local, neighborhood-bounded elementary presentations are Hanf threshold local on bounded degree structures. The result is non-uniform in the sense that the radius of locality, as well as the multiplicity threshold, depends on the degree of the structures over which the query is evaluated.

Theorem 4.8. Suppose $\mathfrak{K} \subseteq \mathfrak{K}_\tau$ is closed under substructures and \mathbb{P} is a local, neighborhood-bounded elementary R -presentation scheme for \mathfrak{K} . If $\mathfrak{J} \subseteq \mathfrak{K}_\tau$ is a \mathbb{P} -invariant elementary

boolean query on \mathfrak{K} , then for all d , there are $r, t \in \omega$ such that \mathfrak{J} is Hanf r, t -threshold local on $\mathfrak{K} \cap \mathfrak{K}_T^d$.

Note that Theorem 4.8 is non-uniform in the degree bound d , that is, both the locality radius and the threshold multiplicity depend on d . This is in contrast to the Threshold Locality Theorem of Fagin et al. (1995) where the locality radius depends only on the quantifier rank of an elementary boolean query and not on the degree bound on the Gaifman graphs of a class of structures over which it is evaluated.

The bound on degree enters into the proof of Theorem 4.8 in part via the fact that the class of structures of degree bounded by d is wide, in the sense of Atserias et al. (2008).

Definition 4.9. Let $\mathfrak{K} \subseteq \mathfrak{K}_T$ and $A \in \mathfrak{K}$.

- (a) $X \subseteq A$ is r -scattered (away from $Y \subseteq A$) if and only if for all $a, a' \in X$, $A^r(a) \cap A^r(a') = \emptyset$ (and for all $a' \in Y$, $A^r(a) \cap A^r(a') = \emptyset$).
- (b) \mathfrak{K} is p, q -wide if and only if for every $r \in \omega$, there is a function $\zeta_r : \omega \mapsto \omega$, called the p, q -width growth factor for \mathfrak{K} with respect to r , such that for every $A \in \mathfrak{K}$, $X_1, \dots, X_p \subseteq A$, every $Z \subseteq A$ with $|Z| \leq q$, and $m \in \omega$, if $|X_i| \geq \zeta_r(m)$, for $1 \leq i \leq p$, then there are sets $Y_i \subseteq X_i$ with $|Y_i| \geq m$, for $1 \leq i \leq p$, and $\bigcup_{1 \leq i \leq p} Y_i$ is r -scattered away from Z .

The concept of a wide class, as defined in Atserias et al. (2008), corresponds to $1, 0$ -wide. This notion was introduced, though not named, in Atserias et al. (2006), where it was used to establish the homomorphism preservation theorem over classes of finite structures of bounded degree, which are, in addition, closed under substructures and disjoint unions. The notion, and its generalizations, almost wide and quasi-wide, were exploited in Atserias et al. (2008) and Dawar (2010) to establish both homomorphism preservation theorems and preservation with respect to extensions theorems over several “tame” classes of finite structures. Important connections between these notions of width and the theory of sparse graphs are expounded in Nesetril and de Mendez (2012) where, among other things, it is established that a substructure closed class is wide if and only if it has bounded degree.

The proof of Theorem 4.8 makes use of the following combinatorial lemma which is implicit in Atserias et al. (2008).

Lemma 4.10. Suppose \mathfrak{K} has bounded degree, that is, there is a d such that the maximum degree of G_A for every $A \in \mathfrak{K}$ is at most d . Then for every n, q , \mathfrak{K} is n, q -wide.

Proof of Theorem 4.8. Let $\mathfrak{K} \subseteq \mathfrak{K}_\tau$ be closed under substructures; we write \mathfrak{K}^d for $\mathfrak{K} \cap \mathfrak{K}_\tau^d$. Let \mathbb{P} be a local, neighborhood-bounded elementary presentation scheme for \mathfrak{K} , with neighborhood expansion factor ν , and $\mathfrak{J} \subseteq \mathfrak{K}_\tau$ be a \mathbb{P} -invariant elementary boolean query on \mathfrak{K} . By Remark 4.5, we may also suppose that \mathbb{P} is degree bounded with degree bound d^ν .

By Definition 4.3, we may fix θ to be a first-order sentence \mathbb{P} -invariant on \mathfrak{K} such that for every $A \in \mathfrak{K}$

$$A \in \mathfrak{J} \iff (\exists A^*)(A^* \mathbb{P} A \wedge A^* \models \theta).$$

We need to show that for all $d \in \omega$, there are $r, t \in \omega$ such that for all $A, B \in \mathfrak{K}^d$, if $A \approx_{r,t} B$, then $A \equiv_{\mathfrak{J}} B$. Fix d , and let $d^* = d^\nu$. By Theorem 3.3, we may choose $r_\theta \in \omega$ and $t_{d^*} \in \omega$, such that for all $A^*, B^* \in \mathfrak{K}_{\tau \cup R}^{d^*}$, if $A^* \approx_{r_\theta, t_{d^*}} B^*$, then $A^* \equiv_\theta B^*$. (Indeed, as we've emphasized above, r_θ depends only on (the quantifier rank) of θ , and not on the degree bound d^* , but this observation does *not* translate to uniformity of our result.)

Suppose that for every $r, t \in \omega$, there are structures $A_{r,t}, B_{r,t} \in \mathfrak{K}^d$ such that $A_{r,t} \approx_{r,t} B_{r,t}$, but $A_{r,t} \in \mathfrak{J}$ and $B_{r,t} \notin \mathfrak{J}$; we derive a contradiction from this hypothesis, by showing that for large enough r^* and t^* , if $A_{r^*,t^*} \approx_{r^*,t^*} B_{r^*,t^*}$, then there are presentations $A_{r^*,t^*}^* \mathbb{P} A_{r^*,t^*}$ and $B_{r^*,t^*}^* \mathbb{P} B_{r^*,t^*}$ such that $A_{r^*,t^*}^* \approx_{r_\theta, t_{d^*}} B_{r^*,t^*}^*$.

Let $s = \nu \cdot r_\theta$. Let ι_1, \dots, ι_k be a list of all the isomorphism types of pointed s -neighborhoods that occur in any $A_{r,t}$ (equivalently, $B_{r,t}$). For each $1 \leq j \leq k$, let $A_{r,t}^j$ be the set of $a \in A_{r,t}$ such that $A_{r,t}^s(a)$ is of isomorphism type ι_j and in like manner define $B_{r,t}^j$. It is easy to construct sequences $A_n = A_{r_n, t_n}$ and $B_n = B_{r_n, t_n}$ with $r_n < r_{n+1}$ and $t_n < t_{n+1}$, for all $n \in \omega$ with the following property:

(1). *there is a partition of $\{1, \dots, k\}$ into a pair of sets U and V such that*

- (a) *for every $j \in U$ there is an $m_j < t_1$ such that for all $n \geq 1$, $|A_n^j| = m_j = |B_n^j|$ and*
- (b) *for all $j \in V$ and for all $n \geq 1$, $|A_n^j|, |B_n^j| > n$.*

We call the isomorphism type ι_j *rare*, if it falls under (1)(a), and in this case, we also call points in A_n^j and B_n^j *rare*; we call the isomorphism type ι_j *common*, if it falls under (1)(b), and in this case, we also call points in A_n^j and B_n^j *common*. First, we choose r' large enough so that at most a fixed number t_{rare} of disjoint r' -neighborhoods suffices to cover all rare points in A_n and B_n for all n . It is easy to see that any $r' > 2s \cdot \sum_{j \in U} m_j$ suffices to guarantee this. Moreover, recalling that for all n , $A_n \approx_{r_n, t_n} B_n$ if $r_{n'} \geq r'$ and $t_{n'} \geq t_{rare}$, the disjoint r' -neighborhoods $A_{n'}^{r'}(a_{n'1}), \dots, A_{n'}^{r'}(a_{n't_A})$ and

$B_{n'}^{r'}(a_{n'1}), \dots, B_{n'}^{r'}(a_{n't_B})$, may be chosen so that $t_A = t_B$, and for every $1 \leq i \leq t_A$, $A_{n'}^{r'}(a_{n'i}) \cong B_{n'}^{r'}(b_{n'i})$. It follows immediately, by the fact that \mathbb{P} is a local presentation scheme, that there are for each such n' , presentations $A_{n'}^* \mathbb{P} A_{n'}$ and $B_{n'}^* \mathbb{P} B_{n'}$ such that for every $1 \leq i \leq t_A$, $(A_{n'}^* \upharpoonright A_{n'}^{r'}(a_{n'i})) \cong (B_{n'}^* \upharpoonright B_{n'}^{r'}(b_{n'i}))$. We let $r^* = 2s \cdot \sum_{j \in U} m_j$, and q^* be the maximum size of an r^* -neighborhood in a graph of degree d . Then $q = q^* \cdot t_{\text{rare}}$ is the maximum size of a cover of the s -neighborhoods generated by U by a system of disjoint r^* -neighborhoods. For all large enough n , we select $\alpha_n \subseteq A_n$ and $\beta_n \subseteq B_n$ to be such covers. We now turn to deal with common points.

First, observe that for all $j \in V$, there are finitely many, say j^* , isomorphism types $\iota_{j1}, \dots, \iota_{jj^*}$ such that every presentation of an s -neighborhood of isomorphism type ι_j is of type ι_{ji} , for some $1 \leq i \leq j^*$. Let $j^{\max} = \max\{j^* \mid j \in V\}$. Let $m = q + t_{d^*} \cdot j^{\max}$. We are now ready to apply Lemma 4.10. First, let $p = |V|$ and ζ_s be the p, q -width growth factor for $\{G_C \mid C \in \mathfrak{K}^d\}$ with respect to s . It follows, that if $|A_n^j|, |B_n^j| > \zeta_s(m)$, then there are systems of points $Y_n^A = \{a_{ji} \in A_n^j \mid j \in V \text{ and } 1 \leq i \leq t_{d^*} \cdot j^*\}$ s -scattered away from α_n and $Y_n^B = \{b_{ji} \in B_n^j \mid j \in V \text{ and } 1 \leq i \leq t_{d^*} \cdot j^*\}$ s -scattered away from β_n . Now choose $n \geq \zeta_s(m)$, so that for every $j \in V$, $|A_n^j|, |B_n^j| > \zeta_s(m)$ and thus Y_n^A is s -scattered away from α_n and Y_n^B is s -scattered away from β_n . We may now conclude, using disjoint local amalgamation for \mathbb{P} that there are $A^* \mathbb{P} A_n$ and $B^* \mathbb{P} B_n$ such that multiplicity of isomorphism types of presentations of each rare type in A^* and B^* are identical, and are all covered by α_n and β_n respectively, and the multiplicity of isomorphism types of presentations of each common type in both A^* and B^* exceeds the threshold t_{d^*} . Thus, $A^* \approx_{r_\theta, t_{d^*}} B^*$, which contradicts our hypothesis. \square

The following theorem establishes non-uniform Hanf threshold locality for boolean queries invariantly first-order definable over bounded degree finite structures (as opposed to a arbitrary bounded degree structures) with respect to local, neighborhood-bounded elementary presentation schemes. We omit its proof, since it is virtually identical to the proof of Theorem 4.8.

Theorem 4.11. *Suppose $\mathfrak{K} \subseteq \mathfrak{F}_\tau$ is closed under substructures and \mathbb{P} is a local, neighborhood-bounded elementary R -presentation scheme for \mathfrak{K} . If $\mathfrak{J} \subseteq \mathfrak{F}_\tau$ is a \mathbb{P} -invariant elementary boolean query on \mathfrak{K} , then for all d , there are $r, t \in \omega$ such that \mathfrak{J} is Hanf r, t -threshold local on $\mathfrak{K} \cap \mathfrak{F}_\tau^d$.*

5 Conclusion

In this paper we establish three Hanf locality results for invariantly definable boolean queries, Theorems 4.6, 4.8, and 4.11. These results suggest interesting avenues for further investigation. We list a few below.

- (a) Does Theorem 4.8 (and 4.11) hold uniformly with respect to degree? We conjecture that the answer is no, that is, for some substructure closed class $\mathfrak{K} \subseteq \mathfrak{K}_\tau(\mathfrak{F}_\tau)$ and some \mathbb{P} local, neighborhood-bounded elementary presentation scheme for \mathfrak{K} , there is a \mathbb{P} -invariant elementary boolean query \mathfrak{J} on \mathfrak{K} such that then for all r , there is a d , such that for all t , there are $A, B \in \mathfrak{K}_\tau^d(\mathfrak{F}_\tau^d)$ such that $A \approx_{r,t} B$, but $A \not\equiv_{\mathfrak{J}} B$.
- (b) Can Theorem 4.11 be extended to other “tame” classes of structures, such as structures with Gaifman graphs of bounded tree-width. We are optimistic, insofar as application of generalizations of the notion of wide class, as deployed in the proof of Theorem 4.11, have been successfully applied to lift other model-theoretic results about degree-bounded structures to very general tame classes of finite structures (see Atserias et al. (2006, 2008); Dawar (2010)).
- (c) In connection with Theorem 4.6, it would be nice to have further examples of non-elementary queries that are elementarily invariantly definable via neighborhood bounded presentations over locally finite structures.

References

- Serge Abiteboul, Richard Hull, and Victor Vianu. *Foundations of Databases*. Addison-Wesley, 1995.
- Matthew Anderson, Dieter van Melkebeek, Nicole Schweikardt, and Luc Segoufin. Locality from circuit lower bounds. *SIAM J. Comput.*, 41(6):1481–1523, 2012.
- Albert Atserias, Anuj Dawar, and Phokion G. Kolaitis. On preservation under homomorphisms and unions of conjunctive queries. *J. ACM*, 53(2):208–237, 2006.
- Albert Atserias, Anuj Dawar, and Martin Grohe. Preservation under extensions on well-behaved finite structures. *SIAM J. Comput.*, 38(4):1364–1381, 2008.
- Michael Benedikt and Luc Segoufin. Towards a characterization of order-invariant queries over tame graphs. *J. Symb. Log.*, 74(1):168–186, 2009.

- Derek G. Corneil and Richard Krueger. A unified view of graph searching. *SIAM J. Discrete Math.*, 22(4):1259–1276, 2008.
- Anuj Dawar. Homomorphism preservation on quasi-wide classes. *J. Comput. Syst. Sci.*, 76(5):324–332, 2010.
- R. Fagin, L. Stockmeyer, and M. Y. Vardi. On monadic NP vs. monadic co-NP. *Information and Computation*, 120(1):78–92, 1995.
- H. Gaifman. On local and nonlocal properties. In J. Stern, editor, *Logic Colloquium '81*, pages 105–135. North Holland, 1982.
- Martin Grohe and Stephan Kreutzer. Methods for algorithmic meta theorems. In M. Grohe and J.A. Makowsky, editors, *Model Theoretic Methods in Finite Combinatorics*, pages 181–206. American Mathematical Society, 2011.
- Martin Grohe and Thomas Schwentick. Locality of order-invariant first-order formulas. *ACM Trans. Comput. Log.*, 1(1):112–130, 2000.
- Y. Gurevich. Toward logic tailored for computational complexity. In *Computation and Proof Theory*, pages 175–216. Springer Lecture Notes in Mathematics 1104, 1984.
- W. Hanf. Model-theoretic methods in the study of elementary logic. In J. Addison, L. Henkin, and A. Tarski, editors, *The Theory of Models*, pages 132–145. North Holland, 1965.
- Lauri Hella, Leonid Libkin, and Juha Nurmonen. Notions of locality and their logical characterizations over finite models. *J. Symb. Log.*, 64(4):1751–1773, 1999.
- Leonid Libkin. *Elements of Finite Model Theory*. Springer, 2004.
- Jaroslav Nesetril and Patrice Ossona de Mendez. *Sparsity - Graphs, Structures, and Algorithms*. Springer, 2012.
- Juha Nurmonen. On winning strategies with unary quantifiers. *J. Log. Comput.*, 6(6):779–798, 1996.