
Contents

---

1 Terminology

Definition 1 (Substructure)**.** *For a set U of elements of a τ-structure $\mathbb{A}$, the* substructure *of $\mathbb{A}$ induced by U is the τ-structure $\mathbb{A}'$ defined as follows and for each $R \in \tau$,*

- *the universe $\tilde{U}$ of $\mathbb{A}$ is $U \cup \{c^{\mathbb{A}} \mid c$ is a constant symbol$\}$,*
- *each constant symbol is interpreted as the same element as in $\mathbb{A}$,*
- *for each predicate $R \in \tau$, R is interpreted as $R^{\mathbb{A}} \cap \tilde{U}^{\mathsf{ar}(R)}$.*

Remark: Throughout the rest, we assume that the *original* τ consists of relational predicates only unless stated otherwise, i.e. it does not contain any constant symbol. The reason is because, when we consider r-neighborhood at $\vec{a}$ we consider the substructure as something induced by the r-ball at $\vec{a}$ with each entry of $\vec{a}$ defining a new constant, by expanding the vocabulary τ to $\tau \cup \{c_1, \ldots, c_\ell$, where $\ell = |\vec{a}|$. Carrying the original constants in each substructure is quite cumbersome especially we are introducing new constants along the way. But all presentation henceforth can be extended to the case when τ contains some constant symbols, in the manner of Definition 1.

Definition 2. *For a τ-structure $\mathbb{A}$, the* Gaifman graph *$G(\mathbb{A})$ of $\mathbb{A}$ is a graph whose vertex set is the universe of $\mathbb{A}$, and there is an edge (without any orientation) between a pair of distinct elements $u, v \in \mathbb{A}$ if and only if there is some $R \in \tau$ such that u, v is related in the relation $R^{\mathbb{A}}$. That is, there is a tuple $\vec{a} \in R^{\mathbb{A}}$ such that $u = a_i$ and $v = a_j$ for some $1 \leq i, j \leq \mathsf{ar}(R)$.*

The *distance* between two elements u, v of $\mathbb{A}$ in $\mathbb{A}$ is defined as their distance in the Gaifman graph of $\mathbb{A}$, denoted as $\mathsf{dist}_{\mathbb{A}}(u, v)$. For a tuple $\vec{u} = (u_1, \ldots, u_\ell$ of elements of $\mathbb{A}$, the distance $\mathsf{dist}_{\mathbb{A}}(\vec{u}, v)$ between $\vec{u}$ and v in $\mathbb{A}$ is $\min\{\mathsf{dist}_{\mathbb{A}}(u_i, v) \mid i \in [\ell]\}$.

Definition 3 (r-ball, r-neighborhood)**.** *For a τ-structure $\mathbb{A}$ and an element $a \in \mathbb{A}$, the* r-ball *$B_r^{\mathbb{A}}(a)$ at a in $\mathbb{A}$ is the set of elements of $\mathbb{A}$ whose distance to a is at most r in the Gaifman graph $G(\mathbb{A})$ of $\mathbb{A}$. For a tuple $\vec{a} = (a_1, \ldots, a_\ell)$ of $\mathbb{A}$, the r-ball $B_r^{\mathbb{A}}(\vec{a})$ at $\vec{a}$ in $\mathbb{A}$ is defined the same way: $B_r^{\mathbb{A}}(\vec{a}) := \{v \in \mathbb{A} \mid$*

$\mathsf{dist}_{\mathbb{A}}(\vec{a}, v) \le r\}$. *The* r-neighborhood of $\vec{a}$ in $\mathbb{A}$ *is a* τ_ℓ*-structure, where* $\tau_\ell = \tau \cup \{c_1, \ldots, c_\ell\}$*, which the substructure of* $\mathbb{A}$ *induced by the* r*-ball at* $\vec{a}$ *when you look at only the relational predicates, and each* c_i *is interpreted as* a_i*. It is denoted by* $N_r^{\mathbb{A}}(\vec{a})$

Definition 4. *(ℓ-queries) Given an integer* $\ell \ge 0$*, an* ℓ-query *on* τ*-structures is a map* Q *such that*

- $Q(\mathbb{A}) \subseteq A^\ell$ *for any* τ*-struture* $\mathbb{A}$*; and*
- *it is closed under isomorphism, that is, if two* τ*-structures* $\mathbb{A}, \mathbb{B}$ *are isomorphic by an isomorphism* $h : A \to B$*, then* $Q(\mathbb{B}) = h(Q(\mathbb{A}))$*. Here,*

$$h(Q(\mathbb{A})) = \{(h(a_1), \ldots, h(a_\ell)) : (a_1, \ldots, a_\ell) \in Q(\mathbb{A})\}.$$

In particular, we consider A^0 *as a singleton set so every* 0*-query is exactly a property on* σ*-structures.*

Definition 5. *(Definable ℓ-queries) Given an* ℓ*-query* Q *on* τ*-structures and a logic* $\mathcal{L}$*,* Q *is* definable in $\mathcal{L}$ *if there is a formula* $\varphi(x_1, \ldots, x_\ell)$ *of* $\mathcal{L}$ *in* τ *such that*

$$Q(\mathbb{A}) = \{(a_1, \ldots, a_\ell) \in A^\ell : \mathbb{A} \vDash \varphi(a_1, \ldots, a_\ell)\}$$

for every τ*-structure* $\mathbb{A}$*.*

For example, the set of vertex pairs of distance exactly two is a 2-query. It is also FO-definable using the formula $\varphi(x, y) := \exists z\ \mathsf{edge}(x, z) \wedge \mathsf{edge}(z, y) \wedge \neg(x, y)$.

The isomorphism between two structures over the same vocabulary is defined in the usual way.

Definition 6 (Isomorphism between two structures)**.** *Let* $\mathbb{A}$ *and* $\mathbb{B}$ *be two* τ*-structures. A mapping* $\iota : \mathbb{A} \to \mathbb{B}$ *is an isomorphism between* $\mathbb{A}$ *and* $\mathbb{B}$ *if*

- ι *is a bijection,*
- *for every constant symbol* $c \in \tau$ *and for every* $i \le \ell$*,* $\iota(c^{\mathbb{A}}) = c^{\mathbb{B}}$*,*
- *for every predicate* $R \in \tau$ *with* $\mathsf{ar}(R) = k$ *and for every* k*-tuple* $(a_1, \ldots, a_k) \in \mathbb{A}^k$*,* $R(a_1, \ldots, a_k)$ *if and only if* $R(\iota(a_1), \ldots, \iota(a_k))$*.*

We write $\mathbb{A} \cong \mathbb{B}$ *when there is* $\mathbb{A}$ *is isomorphic to* $\mathbb{B}$*.*

Note that a property $\mathcal{P}$ of τ-structures is (defined so that) closed under isomorphism. That is, if $\mathbb{A}$ has the property $\mathcal{P}$ and $\mathbb{B}$ is isomorphic to $\mathbb{A}$ then $\mathbb{B}$ also has the property $\mathcal{P}$. Note also that ℓ-query, a generalization of a property, defined so as to be closed under isomorphism.

Intuitively, an ℓ-query Q on τ-structures is Hanf local if the query is closed under the isomorphism of the r-neighborhood. Hanf locality is not guaranteed, and it is rather an anomaly. However, it turns out that an FO-definable ℓ-query is Hanf local.

Definition 7 (r-local isomorphism between two relational structures)**.** *Let* $\mathbb{A}$ *and* $\mathbb{B}$ *be two* τ*-structures, where* τ *consists of relational predicates only (no constant symbols). We write* $\mathbb{A} \leftrightarrows_r \mathbb{B}$ *if there is a bijective mapping* $\iota : \mathbb{A} \to \mathbb{B}$ *(not necessarily isomorphism) such that for every* $a \in \mathbb{A}$*, it holds that* $N_r^{\mathbb{A}}(a) \cong \mathbb{N}_r^{\mathbb{B}}(\iota(a))$*.*

When we have $\mathbb{A} \leftrightarrows_r \mathbb{B}$, they have the same cardinality, they may not be isomorphic but 'locally' they are isomorphic everywhere. Note that in the r-local isomorphism, ι creates an element-to-element mapping designating 'which r-neighborhood to examine'. However, ι is not necessarily the isomorphism which witnesses $N_r^{\mathbb{A}}(a) \cong N_r^{\mathbb{B}}(\iota(a))$.

2 Hanf locality of FO: the case of boolean query

Definition 8 (Hanf locality of boolean query)**.** *Let $\ell > 0$. A property $\mathcal{P}$ on τ-structures is* Hanf local *if there is some integer $r \geq 0$ such that whenever two τ-structures $\mathbb{A}$ and $\mathbb{B}$ satisfy $\mathbb{A} \leftrightarrows_r \mathbb{B}$, then $\mathbb{A} \in \mathcal{P}$ of and only if $\mathbb{B} \in \mathcal{P}$. The smallest such integer r is called the* Hanf locality rank*, $\mathsf{hlr}(\mathcal{P})$ in short.*

Hanf locality is a very useful tool when you want to prove inexpressibility in FO of a property on τ-structures. Here, the recipe for using Hanf locality for establishing inexpressibility of a property $\mathcal{P}$ is the following, for every $r \geq 0$.

1. Choose two τ-structures $\mathbb{A} \in \mathcal{P}$ and $\mathbb{B} \notin \mathcal{P}$.

2. Show that $\mathbb{A} \leftrightarrow_r \mathbb{B}$.

Beware we use the bijection ι from $\mathbb{A}$ to $\mathbb{B}$ (present in the definition of $\leftrightarrow_r$) to introduce the new constant $a \in \mathbb{A}$ and $\iota(a) \in \mathbb{B}$ whenever the r-neighborhood is constructed. So, in order to ensure that $N_r^{\mathbb{A}}(a) \cong N_r^{\mathbb{B}}(\iota(a))$ in the second stage of the recipe, the isomorphism between $N_r^{\mathbb{A}}(a)$ and $N_r^{\mathbb{B}}(\iota(a))$ is not necessarily the restriction of ι (in fact, it cannot), but the isomorphism maps a to $\iota(a)$ as they are the interpretations of the same constant symbol of τ_1, i.e. the expanded vocabulary.

Example 9. The property CONNECTED. *Consider the vocabulary $\{\mathsf{edge}\}$. We want to show that the property* CONNECTED *is not Hanf local. Suppose that it is, with the Hanf locality rank r. The idea is to demonstrate two graphs G_1 and G_2 of the same size (vertex count), (i) which are indistinguishable when you look at any r-neighborhood of G_1 and the corresponding r-neighborhood, and (ii) one is connected whereas the other is not.*

Take G_1 as the disjoint union of two cycles, each of length $2r + 2$. Let G_2 be the cycle of length $4r + 4$. Let ι be an arbitrary bijection from G_1 to G_2. For any vertex v of G_1 or G_2, r-neighborhood at v in G_i is a path of length $2r$ whose two endpoints are non-adjacent. (We chose the length of each cycle of G_1 as the minimum integer so as to satisfy this property.) Therefore, $G_1 \leftrightarrows_r G_2$. However, G_1 is not connected and G_2 is connected. Therefore, CONNECTED *is not Hanf local!*

Theorem 10. *A* FO*-definable property is Hanf local.*

An immediate corollary of Theorem 10, together with the observation in Example 9 that CONNECTED is not Hanf local, means that CONNECTED is not FO-definable.

Inexpressibility in $\exists$ MSO: the example of ACYCLICITY. One can use Hanf locality to prove that some property $\mathcal{P}$ is not expressible in existential MSO. Let's consider the example of ACYCLICITY, the property of (undirected) graphs consisting of graphs without cycles. The vocabulary is over the usual one $\{\mathsf{edge}\}$.

First, let us express the complementary property, CYCLIC, consisting of graphs with some cycle(s). This property can be expressed in an existential MSO. (How?) Therefore, the negation of it yields a universal MSO-expression for ACYCLICITY. (Why?) Does there exist an existential MSO-expression for ACYCLIC-

ITY? It turns out not, and we can prove this using Hanf locality.

Suppose it is expressible in $\exists$MSO, with the sentence $\exists X_1 \cdots \exists X_\ell \varphi(X_1, \ldots, X_\ell)$. Here φ is an FO-sentence over the expanded vocabulary

$$\tau' := \{\mathsf{edge}\} \cup \{X_1, \ldots, X_\ell\},$$

where each X_i is a unary predicate. Therefore, thanks to Theorem 10, the property $\mathcal{P}$ of τ'-structures consisting of all ℓ-*colored* graphs (each unary predicate interpreted as a color class) satisfying φ. Let $d := \mathsf{hlr}(\mathcal{P})$. That is, if $(G_1, S_1, \ldots, S_\ell) \leftrightarrows_d (G_2, T_1, \ldots, T_\ell)$, then $(G_1, S_1, \ldots, S_\ell) \models \varphi$ if and only if $(G_2, T_1, \ldots, T_\ell) \models \varphi$. Especially, G_1 is acyclic if and only if G_2 is acyclic.

We want to design a pair of ℓ-colored graphs which is equivalent under $\cong_d$ whereas one is cyclic and the other is not. This shows that the initial assumption is wrong, that is ACYCLICITY is not expressible in existential MSO.

How do we build such a pair of ℓ-colored graphs? We use a similar construction as in case of TREES (inexpressibility in FO using G_1 as a long path, and G_2 as a disjoint union of a path and a cycle). But in order to establish the bijection between two ℓ-colored graphs with $\leftrightarrows_d$ one needs to be extra careful (and increase the length of paths and cycles quite long). Consider G_1 is an ℓ-colored path of length sufficiently long. The key observation is that there are bounded number of r-*neighborhood isomorphism types* in a ℓ-colored graphs of bounded degree. Therefore, one can find two vertices u, v on the path G_1 such that

- u is before v (fix the starting and endpoint s and t of the path G_1),
- the distance between u, v on G is at least $2r + 2$,
- the distance between s, u as well as v, t is at least r,
- and mostly importantly, the r-*neighborhood isomorphism type* around u and v are identical, even after considering the orientation from left-to-right.

One can prove that as there are bounded (by a function of r and ℓ) number of r-isomorphism types, there must exist such two vertices, sufficiently far from each other. Now, we create a new graph G_2 by

- connecting the predecessor of u with v, to make the former as the predecessor of v, and
- connecting the predecessor of v with u, creating a cycle.

The bijection from G_1 to G_2 is canonical; one can easily see that this certifies that $(G_1, S_1, \ldots, S_\ell) \leftrightarrows_d (G_2, T_1, \ldots, T_\ell)$. Therefore, by Theorem 10 the ℓ-colored graph G_1 is in $\mathcal{P}$ and only if G_2 as the ℓ-colored graph does. That is, the original existential MSO-sentence $\exists X_1 \cdots \exists X_\ell \varphi(X_1, \ldots, X_\ell)$ is satisfied by (uncolored) G_1 if and only if it is satisfied by G_2. However, one is acyclic while the other is not, and the sentence does not express ACYCLICITY.

3 Locality of FO: the case of (general) ℓ-query

We can extend the notion for a property $\mathcal{P}$ being r-local in the sense of Definition 8 to ℓ-queries. Again, we assume that the initial vocabulary τ consists of relational predicates only.

Intuitively, a query Q is *local* if one can decide if an element $v \in \mathbb{A}$ is in the query $Q(\mathbb{A})$, i.e. whether the element satisfies the property in $\mathbb{A}$, only by looking at the neighborhood of v. Let us see some examples of queries. If Q is an ℓ-query in general, we look at the neighborhood of a ℓ-tuple $\vec{v}$ of vertices. Which one seems local and which ones are not?

- 1-query Q consisting of all vertices of a graph contained in some cycle.
- 1-query Q consisting of all vertices of a graph contained in a triangle.
- 1-query Q consisting of all vertices of a graph contained in a cycle of length at most $2d$ for some fixed d.
- 2-query Q consisting of all vertex pairs with at least three common neighbors
- 2-query Q consisting of all vertex pairs contained in a cycle of length at most $2d$
- 2-query Q consisting of all vertex pairs contained in a cycle of length at least $2d$
- 2-query Q consisting of all vertex pairs (u, v) such that there are two vertex disjoint triangles, one containing u the other containing v.

Definition 11 (Hanf locality of ℓ-query)**.** *Let $\ell > 0$. An ℓ-query Q on τ-structures is* Hanf local *if there is some integer $r \geq 0$ such that the following holds:*

$$\text{for any } \tau\text{-structures } \mathbb{A}, \mathbb{B} \text{ and two } \ell\text{-tuples } \vec{a} \in \mathbb{A}^\ell \text{ and } \vec{b} \in \mathbb{B}^\ell$$
$$\text{if } N_r^{\mathbb{A}}(\vec{a}) \cong N_r^{\mathbb{B}}(\vec{b}) \text{ then } \vec{a} \in Q(\mathbb{A}) \text{ if and only if } \vec{b} \in Q(\mathbb{B}).$$

The smallest such integer r is called the (Hanf) locality rank*, or simply locality rank,* $\mathsf{lr}(Q)$ *in short.*

Again, beware that the isomorphism which witnesses $N_r^{\mathbb{A}}(\vec{a}) \cong N_r^{\mathbb{B}}(\vec{b})$ must map a_i of $\vec{a}$ to b_i of $\vec{b}$ as $N_r^{\mathbb{A}}(\vec{a})$, respectively, $N_r^{\mathbb{B}}(\vec{b})$ is the τ_ℓ-structure, where τ_ℓ is the expansion of τ by ℓ constants $c_1, \ldots, c_\ell$. And the r-neighborhood $N_r^{\mathbb{A}}(\vec{a})$ interprets each i-th constant symbol c_i as a_i.

Definition 12 (Gaifman locality of ℓ-query)**.** *Let $\ell > 0$. An ℓ-query Q on τ-structures is* Gaifman local *if there is some integer $r \geq 0$ such that the following holds:*

$$\text{for any } \tau\text{-structures } \mathbb{A} \text{ and two } \ell\text{-tuples } \vec{a}, \vec{b} \in \mathbb{A}^\ell,$$
$$\text{if } N_r^{\mathbb{A}}(\vec{a}) \cong N_r^{\mathbb{A}}(\vec{b}) \text{ then } \vec{a} \in Q(\mathbb{A}) \text{ if and only if } \vec{b} \in Q(\mathbb{A}).$$

The smallest such integer r is called the (Gaifman) locality rank*, or simply locality rank,* $\mathsf{lr}(Q)$ *in short.*

Example 13. Transitive Closure. *Consider the 2-query* TC *on directed graphs, a digraph seen as a relational structure over the vocabulary $\tau = \{\mathsf{edge}\}$ interpreted as the arcs of the directed graph.*

$$\mathsf{TC}(G) = \{(u, v) \mid \text{there is a directed } (u, v)\text{-path in } G\}$$

We want to argue that TC *is not Gaifman local. Suppose it is, with $\mathsf{lr}(\mathsf{TC}) = r$. Consider a directed graph G, sufficiently long ($\geq 4r + 2$ suffices). Then consider two vertices u and v, u precedes v on the path with* $\mathsf{dist}_G(u, v) \geq 2r + 2$*,* $\mathsf{dist}_G(\mathit{first}, u) \geq r$ *and* $\mathsf{dist}_G(v, \mathit{last}) \geq r$*. Then*

- *Each of $N_r(u, v)$ and $N_r(v, u)$ is the disjoint union of two directed paths, one centered at u the other centered at v,*

- *therefore,* $N_r(u,v) \cong N_r(v,u)$*,*
- *by definition of Gaifman locality, we have* $(u,v) \in \mathsf{TC}$ *if and only if* $(v,u) \in \mathsf{TC}$,
- *but we know that this is not the case,*
- *Conclusion:* TC *is not Gaifman local.*

Theorem 14. *Any* FO*-definable query* Q *is Hanf local. Moreover, if* Q *is defined by an* FO*-formula of quantifier rank at most* q*, then the Hanf locality rank of* Q *is at most* $\frac{3^q-1}{2}$.

Lemma 15. *Let* Q *be a non-boolean*[1] *query on* τ*-structures. If* Q *has Hanf locality rank at most* k*, then it has Gaifman locality rank at most* $3k+1$.

Corollary 16. *Any* FO*-definable query* Q *is Gaifman local. Moreover, if* Q *is defined by an* FO*-formula of quantifier rank at most* q*, then the Gaifman locality rank of* Q *is at most* $\frac{3^{q+1}-1}{2}$.

4 Gaifman's Locality Theorem

Note that whether $u \in \mathbb{A}$ is in the r-ball around $\vec{a} \in \mathbb{A}^\ell$ can be verified with an FO-formula. First, the 2-query Q which consists of pairs (u,v) of elements of any τ-structure $\mathbb{A}$ whose distance in the Gaifman graph of $\mathbb{A}$ is exactly 1 can be FO-expressed easily:

$$\begin{aligned} \mathsf{dist}_1(x,y) := & \bigvee_{\substack{R\in\tau;\\ k:=\mathsf{ar}(R)\geq 2}} \exists z_1 \cdots \exists z_k \bigvee_{1\leq i<j\leq \mathsf{ar}(R)} R(z_1,\ldots,z_{i-1},x,z_{i+1},\ldots,z_{j-1},y,z_{j+1},\ldots,z_k) \\ & \vee \bigvee_{\substack{R\in\tau;\\ k:=\mathsf{ar}(R)\geq 2}} \exists z_1 \cdots \exists z_k \bigvee_{1\leq i<j\leq \mathsf{ar}(R)} R(z_1,\ldots,z_{i-1},y,z_{i+1},\ldots,z_{j-1},x,z_{j+1},\ldots,z_k). \end{aligned}$$

One can FO-define the 2-query which consists of two elements of distance exactly k in the Gaifman graph.

$$\mathsf{dist}_k(x,y) := \exists z_1 \cdots \exists z_{k-1}\ \mathsf{dist}_1(x,z_1) \wedge \mathsf{dist}_1(z_1,z_2) \wedge \cdots \wedge \mathsf{dist}_1(z_{k-2},z_{k-1}) \wedge \mathsf{dist}_1(z_{k-1},y)$$

Using the above formula checking if two elements have distance k in the Gaifman graph, one can check if an element is in the r-ball around some ℓ-tuple $\vec{a}$ in $\mathbb{A}$.

$$\mathsf{dist}_{\leq r}(y,\vec{x}) := \bigvee_{i=1}^{\ell} \big(y = x_i \vee \bigvee_{k=1}^{r} \mathsf{dist}_k(y,x_i)\big).$$

We write $y \in B_r(\vec{x})$ as a shorthand for the formula $\mathsf{dist}_{\leq r}(y,\vec{x})$. Intuitively, we say that an FO-formula is local (r-local) if we restrict the interpretation of all quantified variables in the r-ball around (the interpretation of) the free variables.

Definition 17 (Locality of a formula)**.** *Let* $\mathsf{dist}_{>r}(x,y) := \neg\mathsf{dist}_{\leq r}(x,y)$.

- *An* FO*-formula* $\varphi(\vec{x})$ *is said to be* r-local *around* $\vec{x}$ *if every quantifier in* φ *is in the form* $\exists y \in B_r(\vec{x})$.

[1] We defined Gaifman locality only for ℓ-queries with $\ell > 0$.

- *An* FO*-sentence* Φ *is said to be* basic r-local *if it is in the form*

$$\Phi = \exists x_1 \cdots \exists x_\ell \; \Big(\bigwedge_{i=1}^{\ell} \alpha^{(r)}(x_i) \;\wedge\; \bigwedge_{i\neq j} \mathsf{dist}_{>2r}(x_i, x_j) \Big).$$

Here $\alpha^{(r)}(x)$ *is a* r*-local formula around the single variable* x*.*

Note that $\alpha^{(r)}(x)$ above can be replaced by $\alpha^{(r')}(x)$ for any $r' \leq r$ because one can add $\mathsf{dist}_{\leq r'}(y)$ for any quantified variable y occurring in $\alpha^{(r)}(x)$.

Let us contemplate on Definition 17. When a sentence if basic r-local, it asks if there exists a set of elements ($\exists x_1 \cdots \exists x_\ell$) which is pairwise faraway ($\bigwedge_{i\neq j} \mathsf{dist}_{>2r}(x_i, x_j)$), and each element satisfies some r-local formula ($\alpha^{(r)}(x)$). Especially, the last part checks if the r-neighborhood around the element has certain property, expressible in FO.

Theorem 18 (Gaifman Locality Theorem)**.** *Every* FO*-sentence is equivalent to a boolean combination of basic local sentences.*

Some examples. k-INDEPENDENT SET, the property of graph which has an independent set of size k, can be expressed with a basic r-local sentence for some fixed r.

$$k\mathsf{IND} := \exists x_1 \cdots \exists x_k \bigwedge_{i\neq j} \mathsf{dist}_{>1}(x_i, x_j).$$

How about the property CONTAINMENT OF C_{2d+1} of graphs having a cycle of length exactly $2d+1$? Consider the 1-query which says that a vertex contained in a cycle of length $2d+1$. This can be expressed in FO, and in particular with a d-local formula. Indeed, for a vertex v on a cycle of length $2d+1$, all vertices of the cycle is in the d-ball around v, and one can check if there is a cycle of length $2d+1$ in the d-ball around v:

$$\alpha(x) := \exists z_1 \in B_d(x) \cdots \exists z_{2d} \in B_d(x) \; \big(\, \mathsf{edge}(x, z_1) \wedge \mathsf{edge}(z_1, z_2) \wedge \cdots \wedge \mathsf{edge}(z_{2d}, x) \, \big)$$

Certainly, $\alpha(x)$ is a d-local formula around x. Now $\Phi := \exists x \; \alpha(x)$ defines the graph property of having a cycle of length $2d+1$ with a basic d-local sentence.

Let TWO DISJOINT C_{2d+1} be the graph property of having two disjoint cycles of length exactly $2d+1$ each. Is this FO-definable? Let us express this property as a boolean combination of basic d-local sentences. There are two possibilities, if there exists two disjoint copies C_1, C_2 of length $2d+1$ in a graph G. One case is when there are two cycles whose distance is at least $4d+1$. Then one can certify the existence of two such cycles with a basic $2d$-local sentence

$$\Phi_1 := \exists x_1 \exists x_2 \big(\, \alpha(x_1) \wedge \alpha(x_2) \wedge \mathsf{dist}_{>4d}(x_1, x_2) \, \big),$$

where $\alpha(x)$ is defined as above. (Recall that d-local formula is $2d$-local formula as well.) Notice that if $\alpha(v_1)$ and $\alpha(v_2)$ are satisfied, then the witnessing cycles (an interpretation of variables in $\alpha(v_i)$) must be disjoint as each cycle is contained in the d-neighborhood of v_i whereas the distance between v_1 and v_2 is sufficiently far apart.

The other case when two copies are of distance at most $4d$ can be expressed with a $2d$-local formula which verifies if there is a connected subgraph, within $2d$-neighborhood of some vertex, consisting of two disjoint cycles connected by a path of length at most $4d$. Whenever such two copies exist, a central vertex in the connecting path, say v, sees all vertices in the path as well as the two cycles in $N_{2d}(v)$. One can write (tediously long, but trivial) a $2d$-local FO-formula $\beta(x)$ which defines the 1-query that there is such a connected graph in $2d$-neighborhood of v. Now,

$$\Phi := \Phi_1 \vee \exists x\, \beta(x)$$

defines the property TWO DISJOINT C_{2d+1}.

How about k-DOMINATING SET? It is the property of graphs which has a dominating set[2] of size k, can be expressed by an FO-sentence.

$$\Phi_1 := \exists x_1 \cdots \exists x_k\, \forall y \bigwedge_{i=1}^{k} \big(x_i = y \vee \mathsf{edge}(x_i, y)\big).$$

Directly obtaining an equivalent FO-sentence as a boolean combination of basic local sentences is nontrivial, even though such a sentence exists due to Gaifman Locality Theorem. Let us consider the case of $k = 1$. The foremost challenge is to get rid of the universal quantifier. We first want to exclude the obvious case when at least two vertices are necessary for dominating the entire vertex set; that is, when you have a vertex pair whose distance is at least three. This can be expressed by a (negation of) basic 1-local sentence $\Phi_2 := \neg\exists x_1 \exists x_2\ \mathsf{dist}_{>2}(x_1, x_2)$. Notice that omitting a local formula of the form $\alpha^{(1)}(x_i)$ is equivalent to placing a trivially satisfiable 1-local formula.

Once a graph satisfies Φ_1, every pair of vertices have distance at most two. Here, we can use Φ_1 for $k = 1$ where every quantified variable z is follows by the extra locality checking formula $B_r(z)$ for $r = 2$! To summarize,

$$\mathsf{1DOM} := \neg\exists x_1 \exists x_2\ \mathsf{dist}_{>2}(x_1, x_2)\ \wedge\ \exists x\big(\Phi_1(2)(x)\big)$$

where $\Phi_1(2)(x)$ is obtained from Φ_1 by substituting $\exists x_i$ with $\exists x_i \in B_2(x)$ and $\forall y\, \varphi$ with $\forall y\ \big(y \in B_2(x) \to \varphi\big)$.

References

[2] A set D of vertices in G is a dominating set if every vertex of G is either in D or has a neighbor in D.