Algorithmic Model Theory
SS 2016

Prof. Dr. Erich Grädel and Dr. Wied Pakusa

Mathematische Grundlagen der Informatik
RWTH Aachen

Contents

3 Expressive Power of First-Order Logic

In the whole chapter we restrict ourselves to *finite* and *relational* vocabularies τ.

3.1 Ehrenfeucht-Fraïssé Theorem

Let $\mathfrak{A}$ and $\mathfrak{B}$ be τ-structures with $\overline{a} \in A^k$ and $\overline{b} \in B^k$ for some $k \geq 0$. Recall that we write $\mathfrak{A}, \overline{a} \equiv \mathfrak{B}, \overline{b}$ if no FO-formula can distinguish between $(\mathfrak{A}, \overline{a})$ and $(\mathfrak{B}, \overline{b})$, that is if for all $\varphi(\overline{x}) \in \mathrm{FO}(\tau)$ we have

$$\mathfrak{A} \models \varphi(\overline{a}) \Leftrightarrow \mathfrak{B} \models \varphi(\overline{b}).$$

For $m \geq 0$ we write $\mathfrak{A}, \overline{a} \equiv_m \mathfrak{B}, \overline{b}$ if the same holds for all $\mathrm{FO}(\tau)$-formulas of quantifier rank at most m. We aim to develop an algebraic characterisation of $\equiv_m$ via *back-and-forth systems* and a game-theoretic characterisation via *Ehrenfeucht-Fraïssé games*.

Back-and-forth systems. A *partial isomorphism* between τ-structures $\mathfrak{A}$ and $\mathfrak{B}$ is a bijective function p with *finite* domain $\mathrm{dom}(p) \subseteq A$ and range $\mathrm{rg}(p) \subseteq B$ such that p is an isomorphism between the substructures of $\mathfrak{A}$ and $\mathfrak{B}$ induced on $\mathrm{dom}(p)$ and $\mathrm{rg}(p)$, respectively, that is

$$p : \mathfrak{A} \restriction \mathrm{dom}(p) \cong \mathfrak{B} \restriction \mathrm{rg}(p).$$

$\mathrm{Part}(\mathfrak{A}, \mathfrak{B})$ denotes the set of partial isomorphism between $\mathfrak{A}$ and $\mathfrak{B}$. For all $\mathfrak{A}$ and $\mathfrak{B}$ we have $\emptyset \in \mathrm{Part}(\mathfrak{A}, \mathfrak{B})$. For $p \in \mathrm{Part}(\mathfrak{A}, \mathfrak{B})$ we write $p = \overline{a} \to \overline{b}$ for $\overline{a} \in A^k$ and $\overline{b} \in B^k$ if $\mathrm{dom}(p) = \{a_1, \ldots, a_k\}$ and $\mathrm{rg}(p) = \{b_1, \ldots, b_k\}$ and if $p(a_i) = b_i$ for $1 \leq i \leq k$.

Definition 3.1. Let $I \subseteq \mathrm{Part}(\mathfrak{A}, \mathfrak{B})$ and $p \in \mathrm{Part}(\mathfrak{A}, \mathfrak{B})$. Then p has *back-and-forth extensions* in I if

$$\forall\, a \in A\, \exists\, b \in B : p \cup \{(a,b)\} \in I \qquad \text{(forth)}$$
$$\forall\, b \in B\, \exists\, a \in A : p \cup \{(a,b)\} \in I \qquad \text{(back)}$$

Accordingly, for $I, J \subseteq \mathrm{Part}(\mathfrak{A}, \mathfrak{B})$ we say that I has *back-and-forth extensions* in J, if every $p \in I$ has back-and-forth extensions in J.

Definition 3.2. Let $m \geq 0$. A *back-and-forth system* for m-equivalence of $(\mathfrak{A}, \overline{a})$ and $(\mathfrak{B}, \overline{b})$ is a sequence $(I_i)_{i \leq m}$ of sets of partial isomorphisms $I_i \subseteq \mathrm{Part}(\mathfrak{A}, \mathfrak{B})$ such that

- $\overline{a} \to \overline{b} \in I_m$, and
- for all $0 < i \leq m$, I_i has back-and-forth extensions in I_{i-1}.

If such a system $(I_i)_{i \leq m}$ for $(\mathfrak{A}, \overline{a})$ and $(\mathfrak{B}, \overline{b})$ exists, then we write

$$(I_i)_{i \leq m} : (\mathfrak{A}, \overline{a}) \simeq_m (\mathfrak{B}, \overline{b}),$$

and we say that $(\mathfrak{A}, \overline{a})$ and $(\mathfrak{B}, \overline{b})$ are *m-isomorphic*.

Lemma 3.3. For every $m \geq 0$, every τ-structure $\mathfrak{A}$ and every $\overline{a} \in A^k$, there exists an $\mathrm{FO}(\tau)$-formula $\chi^m_{\mathfrak{A},\overline{a}}(x_1, \ldots, x_k)$ of quantifier rank m such that for all $\mathfrak{B}$ and $\overline{b} \in B^k$ we have

$$\mathfrak{B} \models \chi^m_{\mathfrak{A},\overline{a}}(\overline{b}) \Leftrightarrow \mathfrak{A}, \overline{a} \simeq_m \mathfrak{B}, \overline{b}.$$

Moreover the number of different formulas $\chi^m_{\mathfrak{A},\overline{a}}$ only depends on m, τ, and k, and not on $\mathfrak{A}$ or $\overline{a}$ (up to logical equivalence).

Proof. The construction is by induction on $m \geq 0$ (for all $k \geq 0$, $\mathfrak{A}$, and $\overline{a} \in A^k$ at the same time).

$$\chi^0_{\mathfrak{A},\overline{a}}(x_1, \ldots, x_k) = \bigwedge \{\varphi(x_1, \ldots, x_k) : \varphi \text{ is an atomic or negated atomic } \mathrm{FO}(\tau)\text{-formula with } \mathfrak{A} \models \varphi(x_1, \ldots, x_k)\}$$

We have that $\mathfrak{A}, \overline{a} \simeq_0 \mathfrak{B}, \overline{b}$ if, and only if, $\overline{a} \to \overline{b} \in \mathrm{Part}(\mathfrak{A}, \mathfrak{B})$ which means that $(\mathfrak{A}, \overline{a})$ and $(\mathfrak{B}, \overline{b})$ satisfy the same atomic formulas. Note that the number of different atomic formulas in k variables only depends on the vocabulary τ and on $k \geq 0$.

Now let $m > 0$. Then we set $\chi^m_{\mathfrak{A},\overline{a}}(x_1, \ldots, x_k) =$

$$\bigwedge_{a' \in A} \exists x\, \chi^{m-1}_{\mathfrak{A},\overline{a},a'}(x_1, \ldots, x_k, x) \wedge \forall x \bigvee_{a' \in A} \chi^{m-1}_{\mathfrak{A},\overline{a},a'}(x_1, \ldots, x_k, x).$$

Since the number of different formulas $\chi^{m-1}_{\mathfrak{A},\overline{a},a'}$ (up to equivalence) only depends on $m-1$ and $k+1$ (by the induction hypothesis), also the number of different formulas $\chi^m_{\mathfrak{A},\overline{a}}$ only depends on m and k (up to equivalence) and not on $\mathfrak{A}$ or $\overline{a}$. This is of particular importance if one of the structures is infinite, because it guarantees that the conjunction and the disjunction in $\chi^m_{\mathfrak{A},\overline{a}}$ are finite. It holds

$$\begin{aligned}
& (\mathfrak{A}, \overline{a}) \simeq_m (\mathfrak{B}, \overline{b}) \\
\Longleftrightarrow\quad & \begin{cases} \forall a' \in A\, \exists b' \in B : (\mathfrak{A}, \overline{a}, a') \simeq_{m-1} (\mathfrak{B}, \overline{b}, b') \\ \forall b' \in B\, \exists a' \in A : (\mathfrak{A}, \overline{a}, a') \simeq_{m-1} (\mathfrak{B}, \overline{b}, b') \end{cases} \\
\Longleftrightarrow \text{ (by (IH))}\quad & \begin{cases} \forall a' \in A\, \exists b' \in B : \mathfrak{B} \models \chi^{m-1}_{\mathfrak{A},\overline{a},a'}(\overline{b}, b') \\ \forall b' \in B\, \exists a' \in A : \mathfrak{B} \models \chi^{m-1}_{\mathfrak{A},\overline{a},a'}(\overline{b}, b') \end{cases} \\
\Longleftrightarrow\quad & \mathfrak{B} \models \chi^m_{\mathfrak{A},\overline{a}}(\overline{b}).
\end{aligned}$$

Q.E.D.

Ehrenfeucht-Fraïssé games. The Ehrenfeucht-Fraïssé game $G_m(\mathfrak{A}, \overline{a}, \mathfrak{B}, \overline{b})$ is played by two players according to the following rules.

The *arena* consists of the structures $\mathfrak{A}$ and $\mathfrak{B}$. We assume that $A \cap B = \emptyset$. The players are called *Spoiler* and *Duplicator*, and a play of $G_m(\mathfrak{A}, \overline{a}, \mathfrak{B}, \overline{b})$ consists of m moves.

The initial position is $G_m(\mathfrak{A}, \overline{a}, \mathfrak{B}, \overline{b})$. In the *i-th move*, $1 \leq i \leq m$, the play proceeds from the position

$$G_{m-i+1}(\mathfrak{A}, \overline{a}, c_1, \ldots, c_{i-1}, \mathfrak{B}, \overline{b}, d_1, \ldots, d_{i-1}).$$

Spoiler either chooses an element $c_i \in A$ or an element $d_i \in B$. Duplicator answers by choosing an element $c_i \in A$ or $d_i \in B$ in the other structure. The new position is $G_{m-i}(\mathfrak{A}, \overline{a}, c_1, \ldots, c_i, \mathfrak{B}, \overline{b}, d_1, \ldots, d_i)$. After m moves, elements $c_1, \ldots, c_m$ from $\mathfrak{A}$ and $d_1, \ldots, d_m$ from $\mathfrak{B}$ are chosen. Duplicator wins at a final position $G_0(\mathfrak{A}, \overline{a}, c_1, \ldots, c_m, \mathfrak{B}, \overline{b}, d_1, \ldots, d_m)$ if $\mathfrak{A}, \overline{a}, \overline{c} \equiv_0 \mathfrak{B}, \overline{b}, \overline{d}$. Otherwise Spoiler wins.

A *winning strategy* of Spoiler is a function which determines, for every reachable position, a move such that Spoiler wins each play which is consistent with this strategy, no matter how Duplicator plays. Winning strategies for Duplicator are defined analogously. We say that *Spoiler (respectively, Duplicator) wins the game* $G_m(\mathfrak{A}, \overline{a}, \mathfrak{B}, \overline{b})$ if this player has a winning strategy for $G_m(\mathfrak{A}, \overline{a}, \mathfrak{B}, \overline{b})$. By induction on the number of moves it is easy to show that for every (sub)game exactly one of the two players has a winning strategy.

Theorem 3.4 (Ehrenfeucht, Fraïssé)**.** Let $\mathfrak{A}, \mathfrak{B}$ be τ-structures (recall, τ is finite and relational), let $\overline{a} \in A^k$ and $\overline{b} \in B^k$ and let $m \geq 0$. Then the following statements are equivalent:

(i) $\mathfrak{A}, \overline{a} \equiv_m \mathfrak{B}, \overline{b}$.
(ii) $\mathfrak{A}, \overline{a} \simeq_m \mathfrak{B}, \overline{b}$.
(iii) $\mathfrak{B} \models \chi^m_{\mathfrak{A},\overline{a}}(\overline{b})$.
(iv) Duplicator wins $G_m(\mathfrak{A}, \overline{a}, \mathfrak{B}, \overline{b})$.

Proof. Since $\mathfrak{A} \models \chi^m_{\mathfrak{A},\overline{a}}(\overline{a})$ and since $\mathrm{qr}(\chi^m_{\mathfrak{A},\overline{a}}) \leq m$, we have that $(i) \Rightarrow (iii)$. By Lemma 3.3, $(ii) \Leftrightarrow (iii)$. Recall from the introductory course that $(iv) \Rightarrow (ii)$. The proof strategy was to show, by induction on the quantifier rank $m \geq 0$, that if a formula $\varphi(\overline{x})$ of quantifier rank m can distinguish between $\mathfrak{A}, \overline{a}$ and $\mathfrak{B}, \overline{b}$, then we can extract a winning strategy for Spoiler from this formula for the game $G_m(\mathfrak{A}, \overline{a}, \mathfrak{B}, \overline{b})$.

Hence, it suffices to prove $(ii) \Rightarrow (iv)$. Let $(I_i)_{i \leq m} : (\mathfrak{A}, \overline{a}) \simeq_m (\mathfrak{B}, \overline{b})$. For $m = 0$ the claim holds, since $\overline{a} \to \overline{b} \in I_m \subseteq \mathrm{Part}(\mathfrak{A}, \mathfrak{B})$. For $m > 0$ assume that the Spoiler at position $G_m(\mathfrak{A}, \overline{a}, \mathfrak{B}, \overline{b})$ picks an element $a' \in A$. By the forth property Duplicator can pick $b' \in B$ such that $(\overline{a}, a') \to (\overline{b}, b') \in I_{m-1}$. Hence, $(I_i)_{i \leq m-1} : (\mathfrak{A}, \overline{a}, a') \simeq_{m-1} (\mathfrak{B}, \overline{b}, b')$. By the induction hypothesis, Duplicator wins $G_{m-1}(\mathfrak{A}, \overline{a}, a', \mathfrak{B}, \overline{b}, b')$. If Spoiler picks an element $b' \in B$ the reasoning is analogous using the back property. Q.E.D.

Corollary 3.5. For all $k \geq 0$, the relation $\equiv_m$ induces an equivalence relation on pairs $(\mathfrak{A}, \overline{a})$ of τ-structures $\mathfrak{A}$ and $\overline{a} \in A^k$ of finite index.

Corollary 3.6. A class $\mathcal{K}$ of τ-structures is FO-definable if, and only if, there exists $m \geq 0$ such that for all τ-structures $\mathfrak{A}$ and $\mathfrak{B}$ with $\mathfrak{A} \equiv_m \mathfrak{B}$ it holds that $\mathfrak{A} \in \mathcal{K} \Leftrightarrow \mathfrak{B} \in \mathcal{K}$.

3.2 Hanf's technique

Describing winning strategies in Ehrenfeucht-Fraïssé games can be difficult. In this section we want to establish sufficient criteria for structures $\mathfrak{A}$ and $\mathfrak{B}$ which guarantee that Duplicator has a winning strategy in the game $G_m(\mathfrak{A}, \mathfrak{B})$. The following approach goes back to Hanf who gave a similar criterion to characterise $\equiv$ (equivalance in full first-order logic). However, since we are mainly interested in properties of *finite* structures, $\equiv$ is far too powerful (two finite structures $\mathfrak{A}, \mathfrak{B}$ are isomorphic if, and only if, $\mathfrak{A} \equiv \mathfrak{B}$).

Gaifman graph. Let $\mathfrak{A}$ be a τ-structure. The *Gaifman-graph* $\mathcal{G}(\mathfrak{A}) = (V^{\mathcal{G}(\mathfrak{A})}, E^{\mathcal{G}(\mathfrak{A})})$ of $\mathfrak{A}$ is defined as the undirected graph over the vertex set $V^{\mathcal{G}(\mathfrak{A})} = A$ with the edge relation

$$\begin{aligned} E^{\mathcal{G}(\mathfrak{A})} = \{(a,b) : a \neq b \text{ and the elements } a, b \text{ occur together} \\ \text{in some tuple } \overline{c} \in R^{\mathfrak{A}} \text{ for a relation } R \in \tau\}. \end{aligned}$$

The Gaifman graph allows us to define a notion of distance between the elements of the structure $\mathfrak{A}$: we define $d^{\mathfrak{A}} : A^2 \to \mathbb{N} \cup \{\infty\}$ as the usual distance function in the Gaifman graph $\mathcal{G}(\mathfrak{A})$ of $\mathfrak{A}$.

Let $r \geq 0$. The *r-neighbourhood* of an element $a \in A$ is the set $N^r_{\mathfrak{A}}(a) = N^r(a) = \{b \in A : d^{\mathfrak{A}}(a,b) \leq r\}$. In particular, $N^0(a) = \{a\}$. For a tuple $\overline{a} = (a_1, \ldots, a_k) \in A^k$ we set

$$N^r(\overline{a}) = \bigcup_{1 \leq i \leq k} N^r(a_i).$$

The *r-isomorphism type* of an element $a \in A$ is the isomorphism type ι of the structure $(\mathfrak{A} \restriction N^r(a), a)$ (that is of the substructure of $\mathfrak{A}$ induced on the r-neighbourhood of a extended by a new constant symbol to distinguish the element a). This means that for τ-structures $\mathfrak{A}, \mathfrak{B}$, two elements $a \in A$ and $b \in B$ have the same *r-isomorphism type* if there is an isomorphism $\pi : \mathfrak{A} \restriction N^r(a) \to \mathfrak{B} \restriction N^r(b)$ with $\pi(a) = b$.

Definition 3.7. Let $r \geq 0$ and $t \geq 0$. Two τ-structures $\mathfrak{A}$ and $\mathfrak{B}$ are (r,t)*-Hanf equivalent* if for all isomorphism types ι of structures $(\mathfrak{C}, c)$ (where $\mathfrak{C}$ is a τ-structure and $c \in C$ is a distinguished constant) the number of $a \in A$ with r-isomorphism type ι is the same as the number of $b \in B$ with r-isomorphism type ι or both numbers exceed the *threshold* t.

Remark 3.8. If $\mathfrak{A}$ and $\mathfrak{B}$ are (r,t)-Hanf equivalent, then they also are (r',t)-Hanf equivalent for all $r' \leq r$.

Theorem 3.9 (Hanf's Theorem)**.** Let $m \geq 0$ and let $\mathfrak{A}$ and $\mathfrak{B}$ be two τ-structures such that all 3^m-neighbourhoods in $\mathfrak{A}$ and $\mathfrak{B}$ have at most $e \geq 0$ many elements.

If $\mathfrak{A}$ and $\mathfrak{B}$ are $(3^m, m \cdot e)$-Hanf equivalent, then $\mathfrak{A} \equiv_m \mathfrak{B}$.

Proof. For $i \geq 0$ we obtain a back-and-forth system for m-equivalence of $\mathfrak{A}$ and $\mathfrak{B}$ by setting

$$I_{m-i} = \{\bar{a} \to \bar{b} \in \mathrm{Part}(\mathfrak{A}, \mathfrak{B}) : |\bar{a}| = |\bar{b}| = i, \\ \mathfrak{A} \restriction N^{3^{m-i}}(\bar{a}), \bar{a} \cong \mathfrak{B} \restriction N^{3^{m-i}}(\bar{b}), \bar{b}\}.$$

We have $I_m = \{\emptyset\}$, so let $i \geq 1$. Without loss of generality, it suffices to show that I_{m-i} has forth-extensions in I_{m-i-1}. Let $\bar{a} = (a_1, \ldots, a_i)$ and $\bar{b} = (b_1, \ldots, b_i)$ and ρ be such that $\rho : \mathfrak{A} \restriction N^{3^{m-i}}(\bar{a}), \bar{a} \cong \mathfrak{B} \restriction N^{3^{m-i}}(\bar{b}), \bar{b}$. Let $a \in A$. We have to find $b \in B$ such that $\mathfrak{A} \restriction N^{3^{m-i-1}}(\bar{a}, a), \bar{a}, a \cong \mathfrak{B} \restriction N^{3^{m-i-1}}(\bar{b}, b), \bar{b}, b$.

Case 1 (close to $\bar{a}$). If $a \in N^{2 \cdot 3^{m-i-1}}(\bar{a})$, then we choose $b = \rho(a) \in N^{2 \cdot 3^{m-i-1}}(\bar{b})$. This is a valid choice since we have $\rho : \mathfrak{A} \restriction N^{3^{m-i}}(\bar{a}), \bar{a}, a \cong \mathfrak{B} \restriction N^{3^{m-i}}(\bar{b}), \bar{b}, b$.

Case 2 (far from $\bar{a}$). If $a \notin N^{2 \cdot 3^{m-i-1}}(\bar{a})$, then $N^{3^{m-i-1}}(a) \cap N^{3^{m-i-1}}(a_j) = \emptyset$ for all $1 \leq j \leq i$. Hence, it suffices to find $b \in B$ with the same 3^{m-i-1}-isomorphism type as a (call this ι) and the property that $N^{3^{m-i-1}}(b) \cap N^{3^{m-i-1}}(b_j) = \emptyset$ for all $1 \leq j \leq i$.

We know that in $\mathfrak{A}$ and $\mathfrak{B}$ there are the same numbers of realisations of ι or more than $m \cdot e$ many. By our assumption, we know that in

$N^{2 \cdot 3^{m-i-1}}(\bar{a})$ there are at most $m \cdot e$ realisations, and the same number of realisations can be found in $N^{2 \cdot 3^{m-i-1}}(\bar{b})$ (because of ρ). Hence, we can find a $b \in B$ as claimed. Q.E.D.

Corollary 3.10. Let $m \geq 0$ and let $\mathfrak{A}$ and $\mathfrak{B}$ be τ-structures such that the maximal degree in the Gaifman graphs $\mathcal{G}(\mathfrak{A})$ and $\mathcal{G}(\mathfrak{B})$ is $d \geq 0$. If $\mathfrak{A}$ and $\mathfrak{B}$ are $(3^m, m \cdot d^{3^m})$ equivalent, then $\mathfrak{A} \equiv_m \mathfrak{B}$.

Corollary 3.11. Connectivity of finite graphs is not definable in first-order logic.

Proof. Let $\mathfrak{A}_n$ be a cycle of length $2n$ and let $\mathfrak{B}_n$ be the disjoint union of two cycles of length n. For m we can set $n = 3^{m+1}$. Then $\mathfrak{A}_n$ and $\mathfrak{B}_n$ are $(3^m, \infty)$-Hanf equivalent but $\mathfrak{A}_n$ is connected while $\mathfrak{B}_n$ is not. Q.E.D.

3.3 Gaifman's Theorem

Hanf's technique shows that first-order logic can essentially express local properties only: if two structures realise the same number of $f(m)$-neighbourhood types, then no first-order sentence with quantifier rank $\leq m$ can distinguish between both structures. Gaifman's Theorem makes this observation more precise by showing that every FO-sentence is equivalent to an FO-sentence which only speaks about neighbourhoods of elements of a bounded radius (and this semantic property is guaranteed by the syntactic structure of the sentence). To formally introduce this *Gaifman normal form* for first-order logic we first have to introduce the notions of *local formulas* and *local sentences*.

First of all, for every $r \geq 0$ we can find an FO-formula $\vartheta_{\leq r}(x,y)$ which defines in each structure $\mathfrak{A}$ the pairs of elements $(a,b) \in A^2$ whose distance in the Gaifman graph $\mathcal{G}(\mathfrak{A})$ of $\mathfrak{A}$ is at most r, that is

$$\vartheta^{\mathfrak{A}}_{\leq r} = \{(a,b) : d^{\mathfrak{A}}(a,b) \leq r\}.$$

In formulas we will usually write $d(x,y) \leq r$ as a shorthand for $\vartheta_{\leq r}(x,y)$. Also we write $d(\bar{x},y) \leq r$ for a tuple of variables

$\overline{x} = (x_1, \dots, x_k)$ to abbreviate the formula

$$d(\overline{x}, y) \leq r = \bigvee_{1 \leq i \leq k} d(x_i, y) \leq r.$$

Local formulas. A formula $\varphi(\overline{x})$ is r-local if its evaluation in a structure $\mathfrak{A}$ with respect to a tuple $\overline{a} \in A^k$ only depends on the r-neighbourhood of $\overline{a}$. To capture this formally, we inductively define the *relativisation* $\varphi^{N^r(\overline{x})}(\overline{x}, \overline{y})$ of a formula $\varphi(\overline{x}, \overline{y})$ to the r-neighbourhood $N^r(\overline{x})$ of $\overline{x}$ (for the construction we assume that no variable in $\overline{x}$ is bound in φ):

$$\begin{aligned}
&\varphi^{N^r(\overline{x})} = \varphi \quad \text{for atomic formulas } \varphi \\
&\varphi^{N^r(\overline{x})} = \psi^{N^r(\overline{x})} \circ \vartheta^{N^r(\overline{x})} \quad \text{for } \varphi = \psi \circ \vartheta, \circ \in \{\wedge, \vee\} \\
&\varphi^{N^r(\overline{x})} = \neg\psi^{N^r(\overline{x})} \quad \text{for } \varphi = \neg\psi \\
&\varphi^{N^r(\overline{x})} = \exists z(d(\overline{x}, z) \leq r \wedge \psi^{N^r(\overline{x})}) \quad \text{for } \varphi = \exists z\psi \\
&\varphi^{N^r(\overline{x})} = \forall z(d(\overline{x}, z) \leq r \rightarrow \psi^{N^r(\overline{x})}) \quad \text{for } \varphi = \forall z\psi
\end{aligned}$$

Lemma 3.12. For all $r \geq 0$, $\mathfrak{A}$, $\overline{a} \in A^k$ and $\overline{b} \in (N^r(\overline{a}))^\ell$ we have

$$\mathfrak{A} \restriction N^r(\overline{a}) \models \varphi(\overline{a}, \overline{b}) \quad \Leftrightarrow \quad \mathfrak{A} \models \varphi^{N^r(\overline{x})}(\overline{a}, \overline{b}).$$

Definition 3.13. A formula $\varphi(\overline{x})$ is called *r-local* if $\varphi(\overline{x}) \equiv \varphi^{N^r(\overline{x})}(\overline{x})$, that is if for all $\mathfrak{A}$ and $\overline{a} \in A^k$ we have

$$\mathfrak{A} \models \varphi(\overline{a}) \Leftrightarrow \mathfrak{A} \models \varphi^{N^r(\overline{x})}(\overline{a}) \Leftrightarrow \mathfrak{A} \restriction N^r(\overline{a}) \models \varphi(\overline{a}).$$

Note that r-locality is a semantic property of formulas. However, it is easy to see that all formulas $\varphi^{N^r(\overline{x})}(\overline{x})$ are r-local (in other words, the syntatic transformations guarantee that we obtain a local formula, but of course there are local formulas which do not have this syntactic form). Moreover, it is not hard to verify that every formula $\varphi(\overline{x})$ which is r-local is also r'-local for all $r' \geq r$. For a formula $\varphi(\overline{x})$ we write $\varphi^r(\overline{x}) = \varphi^{N^r(\overline{x})}(\overline{x})$ to denote the r-local version of the formula $\varphi(\overline{x})$.

Local sentences. An ℓ-tuple of elements $\overline{a} = (a_1, \dots, a_\ell) \in A^\ell$ in a structure $\mathfrak{A}$ is called *r-scattered* if $d(a_i, a_j) > 2r$ for all a_i and a_j, $i \neq j$, that is if the r-neighbourhoods $N^r(a_i)$, $1 \leq i \leq \ell$, are pairwise disjoint. A *basic local sentence* of *Gaifman rank* (r, m, ℓ) is a sentence of the form

$$\exists x_1 \cdots \exists x_\ell \left(\bigwedge_{i \neq j} d(x_i, x_j) > 2r \wedge \bigwedge_i \psi^r(x_i) \right),$$

where $\mathrm{qr}(\psi) = m$, which expresses the existence of an r-scattered tuple of length ℓ such that every point in this tuple satisfies an r-local property which is specified by a formula ψ of quantifier-rank m. A *local sentence* is Boolean combination of basic local sentences.

Theorem 3.14 (Gaifman)**.** Every first-order sentence is equivalent to a local sentence.

To prove Gaifman's Theorem it suffices to show the following lemma.

Lemma 3.15. If $\mathfrak{A}$ and $\mathfrak{B}$ satisfy the same basic local sentences, then $\mathfrak{A} \equiv \mathfrak{B}$.

Proof (of Gaifman's Theorem using the preceeding lemma). Let Φ denote the set of all basic local sentences. Let φ be an FO-sentence and let $\mathcal{K} = \mathrm{Mod}(\varphi)$ be the class of models of φ. For $\mathfrak{A} \in \mathcal{K}$ we define

$$\Phi(\mathfrak{A}) = \{\varphi : \varphi \in \Phi, \mathfrak{A} \models \varphi\} \cup \{\neg\varphi : \varphi \in \Phi, \mathfrak{A} \models \neg\varphi\}$$

Then for all $\mathfrak{A} \in \mathcal{K}$ we have $\Phi(\mathfrak{A}) \models \varphi$, because if $\mathfrak{B} \models \Phi(\mathfrak{A})$, then $\mathfrak{A}$ and $\mathfrak{B}$ agree on all sentences from Φ and thus, by the preceeding lemma, we have that $\mathfrak{A} \equiv \mathfrak{B}$. By the compactness theorem, we can find finite sets $\Phi_0(\mathfrak{A}) \subseteq \Phi(\mathfrak{A})$ such that $\Phi_0(\mathfrak{A}) \models \varphi$ for all $\mathfrak{A} \in \mathcal{K}$.

We claim that for a finite subclass $\mathcal{K}_0 \subseteq \mathcal{K}$, the sentence φ is equivalent to $\bigvee_{\mathfrak{A} \in \mathcal{K}_0} \bigwedge \Phi_0(\mathfrak{A})$ (which is a local sentence). We know that $\bigvee_{\mathfrak{A} \in \mathcal{K}_0} \bigwedge \Phi_0(\mathfrak{A}) \models \varphi$, so assume that for every finite subclass of structures $\mathcal{K}_0 \subseteq \mathcal{K}$ the set $\{\varphi\} \cup \{\neg \bigwedge \Phi_0(\mathfrak{A}) : \mathfrak{A} \in \mathcal{K}_0\}$ would be satisfiable. Then, by compactness, also $\{\varphi\} \cup \{\neg \bigwedge \Phi_0(\mathfrak{A}) : \mathfrak{A} \in \mathcal{K}\}$ would be satisfiable which is impossible since $\mathfrak{A} \models \bigwedge \Phi_0(\mathfrak{A})$ for all $\mathfrak{A} \in \mathcal{K}$. Q.E.D.

Proof (of Lemma 3.15). For all $m \geq 0$, we prove by induction on $j = m, \ldots, 0$ that one can find values $g(0), g(1), \ldots, g(m)$ such that

$$I_j = \{\overline{a} \to \overline{b} : |\overline{a}| = |\overline{b}| = m - j, (\mathfrak{A} \restriction N^{7^j}(\overline{a}), \overline{a}) \equiv_{g(j)} (\mathfrak{B} \restriction N^{7^j}(\overline{b}), \overline{b})\}$$

defines a back-and-forth system for m-equivalence of $\mathfrak{A}$ and $\mathfrak{B}$. Sufficient criteria for the values $g(0), \ldots, g(m)$ are collected in the course of the proof (and it will be obvious that we can find values which satisfy all contraints). Note that $I_m = \{\emptyset\}$.

Let $0 \leq j < m$ and let $\overline{a} \to \overline{b} \in I_{j+1}$. Then we know that

$$(\mathfrak{A} \restriction N^{7^{j+1}}(\overline{a}), \overline{a}) \equiv_{g(j+1)} (\mathfrak{B} \restriction N^{7^{j+1}}(\overline{b}), \overline{b}).$$

By symmetry, it suffices to show that $\overline{a} \to \overline{b}$ has a forth-extension in I_j. Let $a \in A$. We have to find $b \in B$ such that

$$(\mathfrak{A} \restriction N^{7^j}(\overline{a}a), \overline{a}a) \equiv_{g(j)} (\mathfrak{B} \restriction N^{7^j}(\overline{b}b), \overline{b}b).$$

To this end we consider the $g(j)$-types of the 7^j-neighbourhoods of tuples in $\mathfrak{A}$ and $\mathfrak{B}$. Recall from Lemma 3.3 that we can describe these types by a first-order formula. More precisely, for a structure $\mathfrak{D}$ and a tuple $\overline{d}$ in $\mathfrak{D}$ we set

$$\psi^j_{\overline{d}}(\overline{x}) = \left[\chi^{g(j)}_{(\mathfrak{D}\restriction N^{7^j}(\overline{d}),\overline{d})}(\overline{x})\right]^{N^{7^j}(\overline{x})}.$$

Then $\psi^j_{\overline{d}}(\overline{x})$ is a 7^j-local formula such that $\mathfrak{C} \models \psi^j_{\overline{d}}(\overline{c})$ if the 7^j-neighbourhood of $\overline{c}$ in $\mathfrak{C}$ (with distinguished tuple $\overline{c}$) is $g(j)$-equivalent to the 7^j-neighbourhood of $\overline{d}$ in $\mathfrak{D}$ (with distinguished tuple $\overline{d}$). To find an appropriate $b \in B$ we distinguish between the following cases.

Case 1 (a is close to $\overline{a}$). Assume that $a \in N^{2\cdot 7^j}(\overline{a})$. Then

$$(\mathfrak{A} \restriction N^{7^{j+1}}(\overline{a}), \overline{a}) \models \exists z (d(\overline{a}, z) \leq 2 \cdot 7^j \wedge \psi^j_{\overline{a}a}(\overline{a}, z)).$$

We assume that the quantifier rank of this formula, which only depends on j and $g(j)$, is at most $g(j+1)$ (this gives a first condition on $g(j+1)$).

But then, by our precondition, we can find $b \in N^{2\cdot 7^j}(\overline{b})$ such that

$$(\mathfrak{B} \restriction N^{7^j}(\overline{b})) \models \psi^j_{\overline{a}a}(\overline{b}, b),$$

which implies that $\overline{a}a \to \overline{b}b \in I_j$.

Case 2 (a is far from $\overline{a}$). Assume that $a \notin N^{2\cdot 7^j}(\overline{a})$. Then the 7^j-neighbourhoods of a and $\overline{a}$ are disjoint, i.e. $N^{7^j}(\overline{a}) \cap N^{7^j}(a) = \emptyset$. Hence it suffices to find a $b \in B$ whose 7^j-neighbourhood is disjoint with the 7^j-neighbourhood of $\overline{b}$ and such that the 7^j-neighbourhood of a in $\mathfrak{A}$ and of b in $\mathfrak{B}$ have the same $g(j)$-type. Formally the requirements for $b \in B$ are:

$$N^{7^j}(\overline{b}) \cap N^{7^j}(b) = \emptyset$$
$$\mathfrak{B} \restriction N^{7^j}(b) \models \psi^j_a(b).$$

For $s \geq 1$ we define a formula $\delta_s(x_1, \ldots, x_s)$ which expresses the existence of a $2 \cdot 7^j$-scattered tuple of elements whose 7^j-neighbourhood has the same $g(j)$-type as the 7^j-neighbourhood of a in $\mathfrak{A}$:

$$\delta_s(x_1, \ldots, x_s) = \bigwedge_{\ell \neq k} d(x_\ell, x_k) > 4 \cdot 7^j \wedge \bigwedge_k \psi^j_a(x_k).$$

We now determine the maximal lenght e of such tuples which are realised in $\mathfrak{A}$ and the maximal lenght i of such tuples which are realised in $\mathfrak{A} \restriction N^{2\cdot 7^j}(\overline{a})$, that is i and e are determined such that

$$(\mathfrak{A} \restriction N^{7^{j+1}}, \overline{a}) \models \exists x_1 \cdots \exists x_i \, (\bigwedge_k d(\overline{a}, x_k) \leq 2 \cdot 7^j \wedge \delta_i) \tag{3.1}$$

$$(\mathfrak{A} \restriction N^{7^{j+1}}, \overline{a}) \not\models \exists x_1 \cdots \exists x_{i+1} \, (\bigwedge_k d(\overline{a}, x_k) \leq 2 \cdot 7^j \wedge \delta_{i+1}) \tag{3.2}$$

$$\mathfrak{A} \models \exists x_1 \cdots \exists x_e \, \delta_e \tag{3.3}$$

$$\mathfrak{A} \not\models \exists x_1 \cdots \exists x_{e+1} \, \delta_{e+1}. \tag{3.4}$$

Of course, $i \leq e$. Moreover, $i \leq m - j = |\overline{a}| = |\overline{b}|$. We claim that the corresponding values determined in $\mathfrak{B}$ are the same. For 3.1 and 3.2 we guarantee this by choosing $g(j+1)$ large enough. Note that the quantifier rank of the formulas in 3.1 and 3.2 only depends on m (because i is

bounded by m), j and $g(j)$ (we obtain a second condition on $g(j+1)$). For 3.3 and 3.4 this follows since these are basic local sentences and $\mathfrak{A}$ and $\mathfrak{B}$ satisfy the same basic local sentences by our assumption.
Case 2.1 ($i = e$). Then we claim that *all* $c \in A$ whose 7^j-neighbourhood has the same $g(j)$-type as a are contained in $N^{6\cdot 7^j}(\bar{a})$. Indeed, we could extend each $2 \cdot 7^j$-scattered tuple of such elements in $N^{2\cdot 7^j}(\bar{a})$ by each such element $c \in A$ with $d(\bar{a}, c) > 6 \cdot 7^j$. Since $a \notin N^{2\cdot 7^j}(\bar{a})$ we have

$$(\mathfrak{A} \upharpoonright N^{7^{j+1}}(\bar{a}), \bar{a}) \models \exists z\, (2 \cdot 7^j < d(\bar{a}, z) \leq 6 \cdot 7^j \,\wedge\, \psi_a^j(z) \,\wedge\, \psi_{\bar{a}}^j(\bar{a})).$$

We assume that $g(j+1)$ is larger than the quantifier rank of this formula (this gives a third condition on $g(j+1)$). Then by our assumption we have that

$$(\mathfrak{B} \upharpoonright N^{7^{j+1}}(\bar{b}), \bar{b}) \models \exists z\, (2 \cdot 7^j < d(\bar{b}, z) \leq 6 \cdot 7^j \,\wedge\, \psi_a^j(z) \,\wedge\, \psi_{\bar{a}}^j(\bar{b})).$$

This in turn shows that we can find an appropriate $b \in B$.

Case 2.2 ($i < e$). In this case we know that $\mathfrak{B} \models \exists x_1 \cdots \exists x_{i+1} \delta_{i+1}$ which implies that we can find $b \in B$ such that $N^{7^j}(\bar{b}) \cap N^{7^j}(b) = \emptyset$ and such that $\mathfrak{B} \models \psi_a^j(b)$. Q.E.D.

3.4 Lower bound for the size of local sentences

Gaifman's Theorem states that for every FO-sentence there is an equivalent local one. In the following we show that the local sentence can be much longer than the original one, as captured by

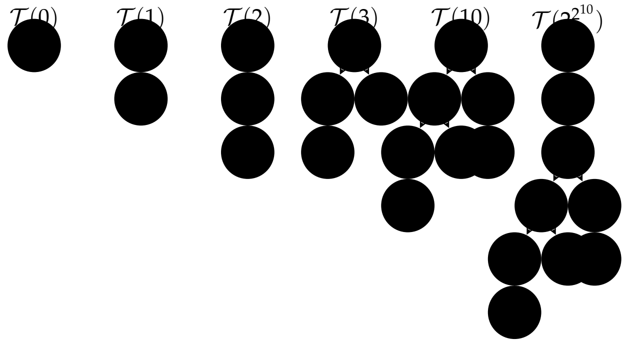
Theorem 3.16. For every $h \geq 1$ there is an FO(E)-sentence $\varphi_h \in \mathcal{O}(h^4)$ such that every FO(E)-sentence in Gaifman normal form, i.e. every local sentence, that is equivalent to φ_h has size at least $Tower(h)$.

Here, $Tower\colon \mathbb{N} \to \mathbb{N}$ is the function defined by $Tower(0) := 1$ and $Tower(n) := 2^{Tower(n-1)}$ for $n > 0$. In order to prove this theorem we first introduce and analyse an encoding of natural numbers by trees.

Definition 3.17. For natural numbers i, n we write $bit(i,n)$ to denote the i-th bit in the binary representation of n, i.e., $\mathrm{bit}(i,n) = 0$ if $\lfloor \frac{n}{2^i} \rfloor$ is even, and $\mathrm{bit}(i,n) = 1$ if $\lfloor \frac{n}{2^i} \rfloor$ is odd. Inductively we define a directed and rooted tree $\mathcal{T}(n)$ for each natural number n as follows:

- $\mathcal{T}(0)$ is the one-node tree.
- For $n > 0$ the tree $\mathcal{T}(n)$ is obtained by creating a new root and attaching to it all trees $\mathcal{T}(i)$ for all i such that $\mathrm{bit}(i,n) = 1$.

  The following figure illustrates these trees.



It is straightforward to see that

$$\text{for all } h, n \geq 0, \quad \mathrm{height}(\mathcal{T}(n)) \leq h \iff n < Tower(h).$$

Recall that the height of a tree is the length of its longest path.

For a graph $G = (V, E)$ and some node $v \in V$, let G_v be the subgraph induced on the set of nodes reachable from v. Now, we show that important properties of these tree encodings of natural numbers can be expressed by small FO(E)-formulas in the sense of the following three Lemmata.

Lemma 3.18. For each $h \geq 0$ there is a formula $eq_h(x,y) \in$ FO(E) of length $\mathcal{O}(h)$ such that for all graphs $G = (V, E)$ we have that: if there are $u, v \in V$ and $m, n < Tower(h)$ with $G_u \cong \mathcal{T}(n)$ and $G_v \cong \mathcal{T}(m)$, then $G \models eq_h(u,v) \Leftrightarrow n = m$.

Proof. • If $h = 0$, set $eq_h(x,y) := true$.

- If $h > 0$, $eq_h(x,y)$ has to be equivalent to

$$\forall z(Exz \to \exists w(Eyw \land eq_{h-1}(z,w)))\land$$
$$\forall w(Eyw \to \exists z(Exz \land eq_{h-1}(z,w))).$$

The length of the formula we get by this recursive definition would be exponential in h. However, we can rewrite it as follows:

$$\begin{aligned} eq_h(x,y) :=& (\exists z Exz \leftrightarrow \exists w Eyw)\land \\ & \forall z(Exz \to \exists w(Eyw \land \forall w'(Eyw' \to \exists z'(Exz'\land \\ & \forall u \forall v((u = z \land v = w) \lor (u = z' \land v = w') \to \\ & eq_{h-1}(u,v)))))). \end{aligned}$$

Q.E.D.

Lemma 3.19. For $h \geq 0$ there is a formula $code_h(x) \in \mathrm{FO}(E)$ of length $\mathcal{O}(h^2)$ such that for all graphs $G = (V,E)$ and $v \in V$:

$$G \models code_h(v) \iff G_v \cong \mathcal{T}(i) \text{ for some } i < Tower(h).$$

Proof.
- If $h = 0$, set $code_h(x) := \neg\exists y Exy$.
- If $h > 0$, set

$$\begin{aligned} code_h(x) :=& \forall y(Exy \to code_{h-1}(y))\land \\ & \forall y_1 \forall y_2(Exy_1 \land Exy_2 \land eq_{h-1}(y_1,y_2) \to y_1 = y_2). \end{aligned}$$

Observe that

$$\begin{aligned} \|code_h(x)\| &= \|code_{h-1}(x)\| + \|eq_{h-1}(x,y)\| + \mathcal{O}(1) \\ &\leq c \cdot (1 + 2 + \cdots + h) \text{ for some } c \geq 1, \end{aligned}$$

implying that $\|code_h(x)\| \in \mathcal{O}(h^2)$.

Q.E.D.

Lemma 3.20. For $h \geq 0$ there are formulas

(1) $bit_h(x,y)$ of length $\mathcal{O}(h)$,
(2) $less_h(x,y)$ of length $\mathcal{O}(h^2)$,

(3) $min(x)$ of length $\mathcal{O}(1)$,
(4) $succ_h(x,y)$ of length $\mathcal{O}(h^3)$,
(5) $max_h(x)$ of length $\mathcal{O}(h^4)$,

such that for all $G = (V,E)$ and nodes $u,v \in V$ with $G_u \cong \mathcal{T}(m)$ and $G_v \cong \mathcal{T}(n)$, where $m,n < Tower(h)$:

(1) $G \models bit_h(u,v) \iff bit(m,n) = 1$,
(2) $G \models less_h(u,v) \iff m < n$,
(3) $G \models min(u) \iff m = 0$,
(4) $G \models succ_h(u,v) \iff m + 1 = n$,
(5) $G \models max_h(u) \iff m = Tower(h) - 1$.

Proof. (1) $bit_h(x,y) := \exists z(Eyz \land eq_h(x,z))$,

(2)
- If $h = 0$, set $less_h(x,y) := false$.
- If $h > 0$, set

$$\begin{aligned} less_h(x,y) :=& \exists y'(Eyy' \land \forall x'(Exx' \to \neg eq_{h-1}(x',y'))\land \\ & \forall x''(Exx'' \land less_{h-1}(y',x'') \to \\ & \exists y''(Eyy'' \land eq_{h-1}(y'',x''))) \end{aligned}$$

(3) $min(x) := \neg\exists y Exy$.

(4)
- If $h = 0$, set $succ_h(x,y) := false$.
- If $h > 0$, set

$$\begin{aligned} succ_h(x,y) =& \exists y'(Eyy'\land \\ & \forall y''(Eyy'' \land y' \neq y'' \to less_{h-1}(y',y''))\land \\ & \forall x'(Exx' \to \neg eq_{h-1}(x',y')\land \\ & \forall y''(Eyy'' \land less_{h-1}(y',y'') \to \\ & \exists x''(Exx'' \land eq_{h-1}(y'',x'')))\land \\ & \forall x''(Exx'' \land less_{h-1}(y',x'') \to \\ & \exists y''(Eyy'' \land eq_{h-1}(y'',x'')))\land \\ & \neg min(y') \to (\exists x'(Exx' \land min(x'))\land \\ & \forall x'(Exx' \land less_{h-1}(x',y') \to \\ & \exists z(succ_{h-1}(x',z) \land (z = y' \lor Exz)))). \end{aligned}$$

(5) • If $h = 0$, set $max_h(x) := \neg\exists y Exy$.
• If $h > 0$, set

$$\begin{aligned} max_h(x) :=& \exists y(Exy \wedge min(y)) \wedge \forall y(Exy \rightarrow \\ & (max_{h-1}(y) \vee \exists z(Exz \wedge succ_{h-1}(y,z))). \end{aligned}$$

This formula is correct since $x = Tower(h) - 1 = 2^{Tower(h-1)} - 1$ implies that $\mathcal{T}(Tower(h) - 1)$ has a subtree $\mathcal{T}(i)$ for any $i \leq Tower(h-1) - 1$.

Q.E.D.

Finally, we use these three lemmata to prove a last lemma of which Theorem 3.16 is a corollary.

Lemma 3.21. For all $h \geq 1$ there is a formula $\varphi_h \in \mathrm{FO}(E)$ with $\|\varphi_h\| \in \mathcal{O}(h^4)$ such that every local sentence ψ which is equivalent to φ_h on the class of forests of height less or equal to h has size $\|\psi\| \geq Tower(h)$.

Proof. Let F_h be the forest consisting of all trees $\mathcal{T}(i)$ with $0 \leq i < Tower(h)$ and let F_h^{-i} be the forest F_h without the tree $\mathcal{T}(i)$ for some $0 \leq i < Tower(h)$. Furthermore, $root(x) := \neg\exists y Eyx$. Now, define

$$\begin{aligned} \varphi_h :=& \exists x(root(x) \wedge min(x)) \wedge \\ & \forall x(root(x) \wedge \neg max_h(x) \rightarrow \exists y(root(y) \wedge succ_h(x,y))). \end{aligned}$$

Observe that $\|\varphi_h\| \in \mathcal{O}(h^4)$ and $F_h \models \varphi_h$ as well as $F_h^{-i} \not\models \varphi_h$ for each $0 \leq i < Tower(h)$.

Let ψ be a local sentence which is equivalent to φ_h on the class of all forests of height less or equal to h. We want to show that $\|\psi\| \geq Tower(h)$.

ψ is a Boolean combination of basic local sentences $\chi_1, \ldots, \chi_L$ with

$$\chi_\ell = \exists x_1 \ldots \exists x_{k_\ell}(\bigwedge_{i \neq j} d(x_i, x_j) > 2 \cdot r_\ell \wedge \bigwedge_i \psi_\ell^{r_\ell}(x_i)).$$

W.l.o.g. there is some $m \leq L$ such that $F_h \models \chi_\ell$ for all $\ell \leq m$ and $F_h \not\models \chi_\ell$ for all $m < \ell \leq L$. Hence we can find for all $\ell \leq m$ nodes $u_{\ell,1}, \ldots, u_{\ell,k_\ell}$ in F_h such that $F_h \models d(u_{\ell,i}, u_{\ell,j}) > 2 \cdot r_\ell \wedge \psi_\ell^{r_\ell}(u_{\ell,i})$ for all $i \neq j$. The set U

consisting of all these nodes contains at most $k_1 + \cdots + k_m \leq \|\psi\|$ many nodes.

Towards a contradiction assume that $\|\psi\| < Tower(h)$. Since F_h contains $Tower(h)$ many disjoint trees, there is at least one $j < Tower(h)$ such that $\mathcal{T}(j)$ in F_h contains no U-node. We claim that $F_h^{-j} \models \psi$ (which would yield the desired contradiction).

- $F_h^{-j} \models \chi_\ell$ where $l \leq m$: the local properties around the nodes $u_{\ell,1}, \ldots, u_{\ell,k_\ell}$ also hold in F_h^{-j} since the neighbourhoods are not changed by removing the tree $T(j)$.
- $F_h^{-j} \models \chi_\ell$ where $m < \ell \leq L$: clear, since F_h^{-j} is a substructure of F_h.

Q.E.D.