

MATHEMATICAL LOGIC FOR COMPUTER SCIENCE

(A.Y. 20/21)

HANDOUT N. 8

ABSTRACT. Definability, Axiomatizability, Expressibility. Queries over finite structures. Ehrenfeucht-Fraïssé games for proving non-expressibility of queries in predicate logic.

1. EHRENFUCHT-FRAÏSSÉ GAMES

We now characterize \equiv_k in terms of two-player games. This characterization is due to A. Ehrenfeucht and is useful in many applications.

The basic idea is to revisit our Back-and-Forth proof for *DLO* turning it into a two-players games: the first player, the Spoiler, picks an element in one of the structures \mathfrak{A} and \mathfrak{B} and the second player, the Duplicator, has to answer with an element in the other structure so that the sequence of choices defines an order-isomorphism. We relax the order in which the moves are done, allowing the Spoiler to pick, at each turn, an element from either \mathfrak{A} or \mathfrak{B} .

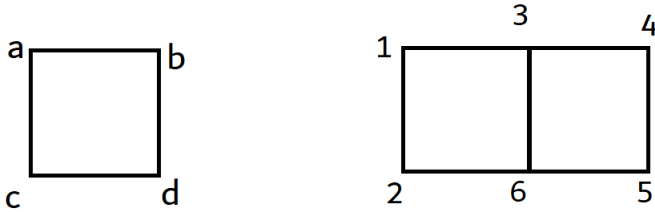
Let's start with a particular case. Consider two finite graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ (i.e., structures adequate for the finite relational language consisting of a single binary relation symbol $E(x, y)$). We describe a game played by two players, Spoiler and Duplicator, on G_1, G_2 . A round of the game takes place as follows.

- (1) Spoiler chooses whether to play on G_1 or on G_2 . Then he picks an element in the chosen graph.
- (2) Duplicator picks an element in the other graph.

So, k rounds of the game determine two sequences: (a_1, \dots, a_k) of elements of G_1 and (b_1, \dots, b_k) of elements of G_2 . The Duplicator wins the k -rounds game if the map $a_i \mapsto b_i$ preserves equality and the edge relations, i.e., $a_i = a_j$ iff $b_i = b_j$ and $E_1(a_i, a_j)$ iff $E_2(b_i, b_j)$. (A map with these properties is called an isomorphism from $\{a_1, \dots, a_k\}$ to $\{b_1, \dots, b_k\}$ and a partial isomorphism between G_1 and G_2 .)

Otherwise the Spoiler wins.

Example 1.1. Let's play the game on the following two graphs \mathfrak{A} (left) and \mathfrak{B} (right).



An example of 3-rounds game is the following:

Round 1: Spoiler play 3, Duplicator plays a

Round 2: S plays d , Duplicator plays 2

Round 3: Spoiler plays 6, Duplicator: ??

After Round 1 the winning conditions for Duplicator are satisfied: the sequences are (a) in A and (3) in B which obviously satisfy partial isomorphisms.

After Round 2: the sequences are (a, d) and $(3, 2)$ which satisfy partial isomorphism: $(a, d) \in E^{\mathfrak{A}}$ iff $(2, 3) \in E^{\mathfrak{B}}$, and $a \neq d$ just as $2 \neq 3$.

After the Spoiler's move at Round 3 Duplicator has no winning answer: if he plays an already played element, say d , he will violate the equality condition in the definition of partial isomorphism: $d = d$ while $2 \neq 6$. If he plays c he will violate the preservation of edge relations: while a and c are connected by an edge in the first graph, 2 and 6 are not in the second. Similarly if he choose b .

We have that the Duplicator does not win all 3-rounds games on these two graphs. On the other hand it is easy to prove that Duplicator can win all 2-rounds games on these two structures.

This has an interesting correspondence with the quantifier complexity of sentences distinguishing \mathfrak{A} from \mathfrak{B} . We already noticed that we can distinguish \mathfrak{A} and \mathfrak{B} with a sentence of quantifier-rank 3 (e.g. "There exist three disconnected nodes"). Thus $\mathfrak{A} \not\equiv_3 \mathfrak{B}$. On the other hand we observed that, apparently, the two structures cannot be distinguished by a sentence of smaller quantifier rank, i.e. $\mathfrak{A} \equiv_2 \mathfrak{B}$.

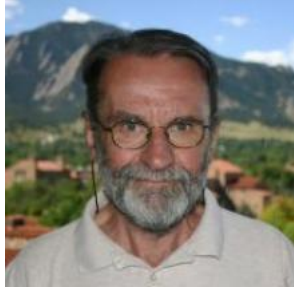
This correspondence between the number of rounds the Duplicator can win and the quantifier complexity of formula distinguishing the two structures holds in general and is the object of the main Theorem below, linking bounded elementary equivalence with Spoiler-Duplicator games.

The idea is generalized to arbitrary relational structures as follows.

Definition 1.2 (Partial Isomorphism). Let (a_1, \dots, a_n) be in A and (b_1, \dots, b_n) be in B . The map $a_1 \mapsto b_1, \dots, a_n \mapsto b_n$ is a *partial isomorphism* between \mathfrak{A} and \mathfrak{B} iff

- (1) For all $i, j \leq n$, $a_i = a_j$ iff $b_i = b_j$,
- (2) For all $i \leq n$, $a_i = c^{\mathfrak{A}}$ iff $b_i = c^{\mathfrak{B}}$,
- (3) For every relation symbol P , for all (i_1, \dots, i_k) in $[1, n]$, $(a_{i_1}, \dots, a_{i_k}) \in P^{\mathfrak{A}}$ iff $(b_{i_1}, \dots, b_{i_k}) \in P^{\mathfrak{B}}$.

FIGURE 1. Andrzej Ehrenfeucht



E-F Games:

Given $k \geq 0$, $t \geq 0$, structures \mathfrak{A} and \mathfrak{B} for \mathcal{L} , and t -ples ($t \geq 0$) (a_1, \dots, a_t) in A^t and (b_1, \dots, b_t) in B^t , the game G_k on (\mathfrak{A}, \vec{a}) and (\mathfrak{B}, \vec{b}) (the k -moves Ehrenfeucht-Fraïssé Game) is defined as follows:

Two players, **Duplicator** and **Spoiler**. A play of the game has k moves (or rounds). At move i ($1 \leq i \leq k$), **Spoiler** chooses an element in either A or B . **Duplicator** answers choosing an element in the other domain. After k moves the game determines k elements in A , let them be a'_1, \dots, a'_k and k corresponding elements in B , let them be b'_1, \dots, b'_k . **Duplicator** wins the game iff

$$(c_1^{\mathfrak{A}}, \dots, c_n^{\mathfrak{A}}, a_1, \dots, a_t, a'_1, \dots, a'_k) \mapsto (c_1^{\mathfrak{B}}, \dots, c_n^{\mathfrak{B}}, b_1, \dots, b_t, b'_1, \dots, b'_k)$$

where c_1, \dots, c_n are the constants of \mathcal{L} , is a partial isomorphism of \mathfrak{A} in \mathfrak{B} .

Duplicator wins the game if he wins all plays.

If \mathfrak{A} e \mathfrak{B} are isomorphic, then Duplicator wins the game (using the isomorphism and its inverse for choosing his answers).

If Duplicator wins the $G_m(\mathfrak{A}, \mathfrak{B})$ game for some m larger than the cardinality of A , then \mathfrak{A} and \mathfrak{B} are isomorphic. At least one play defines a partial isomorphism defined on the whole of A .

We will show that k -moves Ehrenfeucht-Fraïssé games characterize \equiv_k . For proving this the following Lemma is useful (compare with Fraïssé's Theorem).

Lemma 1.3. *For all $k \geq 1$, for all $t \geq 0$, for all $(a_1, \dots, a_t) \in A^t$, $(b_1, \dots, b_t) \in B^t$, the following are equivalent:*

- (1) *Duplicator wins game G_k on (\mathfrak{A}, \vec{a}) and (\mathfrak{B}, \vec{b}) .*
- (2) *The following two properties are satisfied.*
 - (2.1) *For all $a \in A$ there exists $b \in B$ such that Duplicator wins the game G_{k-1} on $(\mathfrak{A}, \vec{a}a)$ and $(\mathfrak{B}, \vec{b}b)$,*
 - (2.2) *For all $b \in B$ there exists $a \in A$ such that Duplicator wins the game G_{k-1} on $(\mathfrak{A}, \vec{a}a)$ and $(\mathfrak{B}, \vec{b}b)$.*

Proof. We prove (1) implies (2). This direction is almost obvious. Suppose Duplicator wins G_{k+1} on (\mathfrak{A}, \vec{a}) and (\mathfrak{B}, \vec{b}) . At the first move Spoiler can play any $a \in A$ or any $b \in B$. In both cases Duplicator must answer maintaining the partial isomorphism. Therefore $\vec{a}a$ and $\vec{b}b$ for any pair (a, b) chosen in that way is a partial isomorphism. Duplicator also has a strategy to answer the next k moves of the game, for each configuration of the first move. Thus for any pair (a, b) determined by the first move, Duplicator knows how to play k more moves of the game on $(\mathfrak{A}, \vec{a}a)$ and $(\mathfrak{B}, \vec{b}b)$. Thus (2.1) holds. (2.2) is proved analogously.

We prove (2) implies (1). Suppose the two properties (2.1) and (2.2) are satisfied for k . We prove that Duplicator wins the G_{k+1} game on (\mathfrak{A}, \vec{a}) and (\mathfrak{B}, \vec{b}) . At the first move, if Spoiler picks $a \in A$ then Duplicator can pick a $b \in B$ that satisfies condition (2.1). If Spoiler picks $b \in B$ then Duplicator can pick an $a \in A$ satisfying condition (2.2). In both cases Duplicator can win game G_k on $(\mathfrak{A}, \vec{a}a)$ and $(\mathfrak{B}, \vec{b}b)$. By definition of the game we are done. □

Theorem 1.4 (A. Ehrenfeucht). *For every $k \geq 0$, the following are equivalent.*

- (1) *Duplicator wins game G_k on $(\mathfrak{A}, \mathfrak{B})$.*
- (2) $\mathfrak{A} \equiv_k \mathfrak{B}$.

Proof. The base case: By definition Duplicator wins the 0-moves game on (\mathfrak{A}, \vec{a}) and (\mathfrak{B}, \vec{b}) iff the map $c_i^{\mathfrak{A}} \mapsto c_i^{\mathfrak{B}}$ and $\vec{a} \mapsto \vec{b}$ is a partial isomorphism of \mathfrak{A} in \mathfrak{B} . Consider the case in which \vec{a} and \vec{b} are empty. Then Duplicator wins the 0-moves game on \mathfrak{A} and \mathfrak{B} iff the map $\chi : c_i^{\mathfrak{A}} \mapsto c_i^{\mathfrak{B}}$ is a partial isomorphism. If the map χ is a partial isomorphism, then for each relation symbol R with k arguments in \mathcal{L} , for each choice of constants c_{i_1}, \dots, c_{i_k} in \mathcal{L} , we have $\mathfrak{A} \models R(c_{i_1}, \dots, c_{i_k})$ iff $\mathfrak{B} \models R(c_{i_1}, \dots, c_{i_k})$ since by definition of partial isomorphism we have $(c_{i_1}^{\mathfrak{A}}, \dots, c_{i_k}^{\mathfrak{A}}) \in R^{\mathfrak{A}}$ iff $(c_{i_1}^{\mathfrak{B}}, \dots, c_{i_k}^{\mathfrak{B}}) \in R^{\mathfrak{B}}$. Furthermore $\mathfrak{A} \models (c_i = c_j)$ iff $\mathfrak{B} \models (c_i = c_j)$. Thus \mathfrak{A} and \mathfrak{B} satisfy the same atomic sentences (the only atomic sentences are $R(c_{i_1}, \dots, c_{i_k})$ and $c_i = c_j$). Each sentence of degree ≤ 0 is a boolean combination of atomic sentences and thus $\mathfrak{A} \equiv_0 \mathfrak{B}$ hold. We have proved that:

$$\text{Duplicator wins the game } G_0 \text{ on } \mathfrak{A}, \mathfrak{B} \iff \mathfrak{A} \equiv_0 \mathfrak{B}.$$

Consider the case $k+1$. Prove (1) to (2). Suppose Duplicator wins game G_{k+1} on $(\mathfrak{A}, \mathfrak{B})$. By the Lemma

- For all $a \in A$ there exists $b \in B$ such that Duplicator wins G_k on (\mathfrak{A}, a) and (\mathfrak{B}, b) ,
- For all $b \in B$ there exists $a \in A$ such that Duplicator wins G_k on (\mathfrak{A}, a) and (\mathfrak{B}, b) .

By inductive hypothesis:

- For all $a \in A$ there exists $b \in B$ such that $(\mathfrak{A}, a) \equiv_k (\mathfrak{B}, b)$,
- For all $b \in B$ there exists $a \in A$ such that $(\mathfrak{A}, a) \equiv_k (\mathfrak{B}, b)$.

By Theorem ?? we have the thesis.

Suppose $\mathfrak{A} \equiv_{k+1} \mathfrak{B}$. If Spoiler moves $a \in A$, by the Forth condition we can find $b \in B$ such that $(\mathfrak{A}, a) \equiv_k (\mathfrak{B}, b)$. By inductive hypothesis Duplicator wins G_k on (\mathfrak{A}, a) and (\mathfrak{B}, b) . If Spoiler moves $b \in B$, by the Back condition we can find $a \in A$ such that $(\mathfrak{A}, a) \equiv_k (\mathfrak{B}, b)$. By inductive hypothesis Duplicator wins G_k on $((\mathfrak{A}, a), (\mathfrak{B}, b))$. By the Lemma, the two conditions just proved are equivalent to the fact that Duplicator wins G_{k+1} on \mathfrak{A} and \mathfrak{B} . \square

Putting it all together we have

$$\text{Duplicator wins } G_k(\mathfrak{A}, \mathfrak{B}) \iff \mathfrak{A} \equiv_k \mathfrak{B}.$$

2. NON-EXPRESSIBILITY RESULTS

Consider the following simple boolean query on the class of finite structures: $EC(\mathfrak{A}) = 1$ if the domain of \mathfrak{A} has even cardinality, and 0 otherwise.

We apply the just described method to show that the Even Cardinality query is not expressible over finite structures in predicate logic.

Theorem 2.1. *The Even Cardinality query is not expressible in the language of graphs over finite graphs.*

Proof. Let $k \geq 0$. We need to find two graphs \mathfrak{A} and \mathfrak{B} such that we can prove that $\mathfrak{A} \equiv_k \mathfrak{B}$ but such that they have a different parity of cardinality.

Let \mathfrak{A} and \mathfrak{B} be the totally disconnected graphs on $k+1$ and $k+2$ vertices respectively (but any choice of p even and q odd labove k would be equally fine).

It is very easy to observe that the Duplicator wins the k -rounds E-F game on such \mathfrak{A} and \mathfrak{B} . The only relation that needs be preserved is total disconnection, and there are enough elements in each structure to make the Duplicator able to answer for at least k rounds. \square

Note that the fact that the Even Cardinality Query is not expressible over finite graphs does not imply that it cannot be defined over another class of structures. Let's consider finite linear orders.

Theorem 2.2. ¹ *Let $k > 0$. If \mathfrak{A} and \mathfrak{B} are finite linear orders with more than 2^k elements then $\mathfrak{A} \equiv_k \mathfrak{B}$.*

Proof. We show that the Duplicator has a winning strategy for the game with k moves on \mathfrak{A} and \mathfrak{B} in a language extended with constants min and max that denote the first and the last element of the linear order.

We work with an inductive hypothesis stronger than the thesis.

Let a_{-1}, a_0 (respectively b_{-1}, b_0) be the interpretations in \mathfrak{A} (respectively in \mathfrak{B}) of min and max . We indicate with $d(x, y)$ the distance between x and y , i.e., $|x - y|$. If x, y are at distance below n we say that they are n -close. We also call $\{y \mid d(x, y) < n\}$ the **neighbourhood** of x of radius n .

Let a_1, \dots, a_i be the elements played in \mathfrak{A} after move i and b_1, \dots, b_i be the corresponding elements played in \mathfrak{B} .

We show that the Duplicator has a strategy such that after every move i of the game, for every $-1 \leq p, q \leq i$, the following three points are satisfied.

- (1) If a_q 2^{k-i} -close to a_p , then $d(b_p, b_q) = d(a_p, a_q)$,
- (2) If a_q it is not 2^{k-i} -close to a_p , then $d(b_p, b_q) \geq 2^{k-i}$,
- (3) $a_p \leq a_q$ if and only if $b_p \leq b_q$ and $a_p = a_q$ iff $b_p = b_q$.

¹The proofs in this Section are essentially drawn from Libkin's Elements of Finite Model Theory.

Note that only point (3) is necessary to get result, by definition of E-F games. The other two conditions are stronger than needed but will make the proof below work.

Intuitively:

- (1) Condition (1) says that “small” intervals are rigidly mapped to intervals of the same size;
- (2) Condition (2) says that “large” intervals are sent to large intervals.

After move i , the notion of “small” means being smaller than 2^{k-i} . To memorize the relevant value of the exponent it is convenient to think of it as the number of rounds remaining. So the notion of “small” is successively:

$$2^k, 2^{k-1}, 2^{k-2}, \dots, 2^1, 2^0.$$

(Case $i = 0$) $p, q \in \{-1, 0\}$. By hypothesis on the cardinality of L_1 we have $d(a_{-1}, a_0) \geq 2^k$, and by hypothesis on the cardinality of L_2 we have $d(b_{-1}, b_0) \geq 2^k$. Therefore the condition (2) is satisfied. (3) obviously is satisfied and (1) is not applied.

(Case $i > 0$) We suppose that the Spoiler plays in \mathfrak{A} . If Spoiler picks an already chosen element, there is nothing to prove.

The elements $(a_{-1}, a_0, a_1, \dots, a_i)$ already played in \mathfrak{A} partition L_1 in $i+1$ intervals. If the Spoiler plays an element a_{i+1} then this element falls necessarily in an interval $]a_j, a_\ell[$, for some $j, \ell \leq i$ (i.e., a_j, a_ℓ are already played and $a_j < a_\ell$) such that the interval does not contain other already played elements (this is where we need the assumption that a_{-1}, a_0 have been already played). We call such an interval a fresh interval.

Since $(a_{-1}, a_0, a_1, \dots, a_i)$ and $(b_{-1}, b_0, b_1, \dots, b_i)$ is an order isomorphism by property (3) of the inductive hypothesis, also $]b_j, b_\ell[$ is a fresh interval, i.e., there are no elements of \mathfrak{B} already played in the interval $]b_j, b_\ell[$.

We must show that, for every $p, q \in \{-1, \dots, i+1\}$ the three conditions are met.

We distinguish two cases depending on the size of the interval $[a_j, a_\ell]$ and subcases depending on whether a_{i+1} is closer to the left-end point or to the right end-point.

(Case 1) $d(a_j, a_\ell) < 2^{k-i}$. Point (1) of the inductive hypothesis gives $d(b_j, b_\ell) = d(a_j, a_\ell)$. Then we can choose b_{i+1} in the range of $]b_j, b_\ell[$ at distance $d(a_j, a_{i+1})$ from b_j and distance $d(a_{i+1}, a_\ell)$ from b_ℓ .

(Case 2) $d(a_j, a_\ell) \geq 2^{k-i}$. By point (2) of the inductive hypothesis we have $d(b_j, b_\ell) \geq 2^{k-i}$. The intervals $[a_j, a_\ell]$ and $[b_j, b_\ell]$ split in two intervals with at least $2^{k-(i+1)}$ elements. We distinguish between three cases depending on the distance of a_{i+1} from the endpoints of the interval $[a_j, a_\ell]$.

(Case 2.1) If $d(a_j, a_{i+1}) < 2^{k-(i+1)}$, then $d(a_{i+1}, a_\ell) \geq 2^{k-(i+1)}$. We can choose b_{i+1} in $[b_j, b_\ell]$ at distance $d(a_j, a_{i+1})$ from b_j . Such a b_{i+1} will satisfy $d(b_{i+1}, b_\ell) \geq 2^{k-(i+1)}$ as required.

(Case 2.2) If $d(a_{i+1}, a_\ell) < 2^{k-(i+1)}$, then $d(a_j, a_{i+1}) \geq 2^{k-(i+1)}$. We can choose b_{i+1} in $[b_j, b_\ell]$ at distance $d(a_{i+1}, a_\ell)$ from b_ℓ . Such a b_{i+1} will satisfy $d(b_j, b_{i+1}) \geq 2^{k-(i+1)}$ as required.

(Case 2.3) If $d(a_j, a_{i+1}) \geq 2^{k-(i+1)}$ and $d(a_{i+1}, a_\ell) \geq 2^{k-(i+1)}$, we can choose b_{i+1} at the center of the interval $[b_j, b_\ell]$, i.e., among the elements of L_2 in $[b_j, b_\ell]$ at a distance at least $2^{k-(i+1)}$ both from b_j that b_ℓ . There is at least one such element, because $d(b_j, b_\ell) \geq 2^{k-i}$.

(Exercise: prove that the choices made in the preceding cases are sufficient to prove the validity of the three points of the inductive hypothesis for $i+1$. Simply check that the three points are met for the pair of points a_{i+1} and a_h with $h \in \{-1, \dots, i\}$). \square

Corollary 2.3. *The Even Cardinality query is not first-order definable over finite models.*

Proof. Suppose by way of contradiction that the complexity of the sentence that defines parity is k . Just take two linear orders with more than 2^k elements and cardinality of different parity. \square

Corollary 2.4. *Connectivity is not first-order definable on finite graphs.*

Proof. We can reduce connectivity to the Even Cardinality query as follows. Associate to each linear order L a graph $G_L = (V_L, E_L)$ so that G_L is connected if and only if L has an odd number of elements. The association:

$$L \longrightarrow G_L$$

5

is obtained as follows. The set of vertices V_L is the set of elements of L . We define the edge relations E_L of the graph G_L as follows.

- We put an edge between each vertex x and the successor of his successor with respect to the linear order L .
- Put an edge between the last and the second element of L .
- Put an edge between between the penultimate and the first element of L .

Note that this description is **uniform** in L in the sense that it gives a recipe to define the edge-relation G_L given any linear order L .

It is easy to verify that:

A linear order L has **even cardinality** iff the graph G_L is **not connected**.

This gives rise to a “reduction”, in the following way; called a **first-order reduction**.

It can be observed that it is possible to write a first-order formula $F(x, y)$ with two free variables in the language $\{<\}$ that defines the construction of the edge relation of the graph associated to a linear order in the sense that:

For every linear order L , for all $a, b \in L$, we have $L \models F(x, y)[a, b]$ if and only if $(a, b) \in E_L$.

To write down this formula it is sufficient to translate the construction of G_L from L described above in the language of linear orders. Note that this is done so that it works uniformly for any linear order as above indicated.

To this aim it is sufficient to express the successor relation, the first and last elements and the successor of successor relation. This can be easily done in the language of orders. For example the formula

$$S(x, y) := (x < y) \wedge \forall z (\neg((x < z) \wedge (z < y)))$$

defines the successor relation, in the sense that, for any linear order L , for all $a, b \in L$: $L \models S(x, y)[a, b]$ iff b is the immediate successor of a with respect to the linear ordering $<^L$. The same can be done for the second-successor relation as well as for the first and last element of the ordering (Exercise).

Now, if there existed a sentence in the language of graphs that expresses connectivity over finite graphs, let it be C , we could conclude that even cardinality is expressible by a sentence in the language of linear orders, which we proved not to be possible.

In fact, let C be the sentence expressing connectivity. Replace in C each occurrence of the edge relation $E(x, y)$ with the formula $F(x, y)$ described above.

We obtain a sentence C^* in the language of linear orders such that the following holds: for any finite linear order L ,

$$L \models C^* \text{ if and only if } G_L \models C.$$

Therefore C^* expresses the odd cardinality query in the language of orders over finite linear orderings:

- (1) If $L \models C^*$ then $G_L \models C$ then G_L is connected hence L has odd cardinality domain.
- (2) If L has odd cardinality domain then G_L is connected and thus $G_L \models C$ and therefore $L \models C^*$.

Thus, $\neg C^*$ expresses the even cardinality query over finite linear orders. Contradiction. \square

Corollary 2.5. *The transitive closure is not first-order definable over finite models.*

Proof. Reduce transitive closure to connectivity as follows. A formula for the transitive closure in the language of graphs is satisfied by two elements v, w of a graph $G = (V, E)$ if and only if there exists a path between v and w in G . The universal closure of such a formula would define connectivity. \square

Other queries can be proved non-expressible with the above method: e.g., the Planarity, Acyclicity, 2-colorability, Hamiltonicity queries in the language of graphs. Another typical application at this level is to prove that the property of being a (directed rooted) tree is not first-order definable over finite models. Indeed, the above method is complete for proving inexpressibility results: if a query is inexpressible in first-order logic then it is possible to prove it by the method described above.