Discrete Mathematics 49 (1984) 291–302
North-Holland

ON THE DEFINABILITY OF PROPERTIES OF FINITE GRAPHS

György TURÁN

Research Group on the Theory of Automata, Hungarian Academy of Sciences, 6720 Szeged, Somogyi u. 7., Hungary

Received 18 October 1982
Revised 24 June 1983

This paper considers the definability of graph-properties by restricted second-order and first-order sentences. For example, it is shown that the class of Hamiltonian graphs cannot be defined by monadic second-order sentences (i.e., if quantification over the subsets of vertices is allowed); any first-order sentence that defines Hamiltonian graphs on n vertices must contain at least $\frac{1}{2}n$ quantifiers. The proofs use Fraïssé–Ehrenfeucht games and ultraproducts.

Introduction

We consider problems of the following type: given a property P of finite graphs and a class C of logical formulas what can be said about the definability of P in C? As the difficulty of defining or expressing a property corresponds intuitively to some notion of complexity, in this way we get problems related to those of complexity theory.

As basic examples for connections existing between definability classes and other classes of languages relevant in other fields of computer science, we mention the theorem of Fagin [5] characterizing languages in NP, and the theorem of Büchi [12] characterizing regular languages as a definability class.

In Section 3 we consider cases when C is a class of second-order formulas. The motivation for this is twofold: firstly Fagin characterizes NP as a second-order definability class, secondly most important graph-properties are not first-order definable but their standard definition is a simple second-order definition using quantification over subsets of vertices or subsets of edges (e.g., connectedness, k-colorability, existence of perfect matching, etc.). We obtain the exact relationship between some of these natural subclasses of second-order definable properties.

If C is the class of first-order sentences and P is not first-order definable one can do the following: let P_n be the class of graphs on n vertices belonging to P. Then of course P_n is finite and so there exists a first-order sentence φ_n such that if G is a graph on n vertices then $G \in P_n \Leftrightarrow G \models \varphi_n$. In this way we obtain a sequence $(\varphi_1, \varphi_2, \ldots)$ of first-order sentences defining P. The complexity of P can be measured by the complexity of the sequence $(\varphi_1, \varphi_2, \ldots)$ (with an appropriate notion of complexity of sentences). This is the approach taken by Immerman [9]. Bounds for this measure were obtained independently in [16], these are described in Section 4.

1. Some definitions

Graphs considered in this paper are finite, undirected, without loops and multiple edges. The *language of graphs* contains one binary relation $R(x, y)$ (and equality). Notions like a graph G being a model of a first-order sentence in the language of graphs are assumed to be known.

A sentence is *second-order* if its variables are first-order variables or relations.

Definition 1.1. A second-order sentence is *monadic* if its second-order variables are relations of arity 1 (i.e., correspond to subsets of vertices), *dyadic* if its second-order variables are relations of arity $\leqslant 2$.

Definition 1.2. A dyadic second-order sentence is *weakly dyadic* if its second order quantifiers over binary relations are of the form

$$\exists R_1(\forall x, y(R_1(x, y) \to R(x, y)) \wedge \cdots)$$
$$\forall R_1(\forall x, y(R_1(x, y) \to R(x, y)) \to \cdots)$$

(i.e., correspond to subsets of edges).

Definition 1.3. A second-order sentence is *existential* if it is of the form

$$\exists R_1 \cdots \exists R_k \Phi,$$

where Φ is a first-order sentence written in the language containing symbols for relations $R_1, \ldots, R_k$ and R.

A property P of graphs is second-order (resp. monadic, dyadic, weakly dyadic, existential weakly dyadic) definable if there exists a second-order (resp. monadic, dyadic, weakly dyadic, existential weakly dyadic) sentence Φ such that $G \models \Phi$ iff $G \in P$. A weakly dyadic property is also called *subgraph-definable* as in this case quantification is allowed over subgraphs of the graph given.

Notation. We use the following notation:
$\mathcal{D}$ = class of dyadic properties,
$\mathcal{S}$ = class of subgraph-definable properties,
$\exists\mathcal{S}$ = class of existential subgraph-definable properties,
$\mathcal{M}$ = class of monadic properties.

Definition 1.4. The *quantifier-rank* QR of a formula is defined inductively as follows: the quantifier-rank of an atomic formula is 0, $QR(\varphi \vee \psi) = QR(\varphi \wedge \psi) = \max(QR(\varphi), QR(\psi))$, $QR(\neg\varphi) = QR(\varphi)$, $QR(\exists x\varphi) = QR(\forall x\varphi) = QR(\varphi) + 1$. The *quantifier-number* QN of a formula φ is the number of quantifiers in φ.

2. Fraïssé–Ehrenfeucht games

In this section we define Fraïssé–Ehrenfeucht games and state some of their basic properties. These games are useful in obtaining negative results for definability and positive results for decidability as well (see e.g. Fagin [4], Ferrante–Rackoff [6]).

Definition 2.1. Let us consider two graphs G_1 and G_2, two players (I and II) and m, the number of moves. The m-move *first-order* Fraïssé–Ehrenfeucht game on G_1 and G_2 is the following (the term 'first-order' is explained by the Theorem below). In Move 1 player I selects one of the graphs and chooses a vertex v_1 of the graph selected, player II selects a vertex w_1 of the other graph. In Move 2 player I selects one of the graphs again (independently of his previous selection) and chooses a vertex v_2 of the graph selected, while player II selects a vertex w_2 of the other graph. (A vertex already selected can be chosen again.) This is repeated m times, finally we have $k \leq m$ points in each graph with a natural correspondence between the two sets ($v_i \leftrightarrow w_i$). Player II wins if this correspondence is an isomorphism between the subgraphs spanned by the points selected in G_1 and G_2, otherwise player I is the winner. G_1 and G_2 are *m-equivalent* in the first-order game ($G_1 \sim_m G_2$) iff player II has a winning strategy. (It can be shown that this is an equivalence relation indeed.)

Thus the role of player I in this game is to find differences between G_1 and G_2, while player II tries to utilize similarities. Equivalence can be considered as a measure of similarity (isomorphic graphs are obviously equivalent for every m).

The fundamental properties of the first-order game are summarized in the following theorem due to Fraïssé and Ehrenfeucht.

Theorem. *$G_1 \sim_m G_2$ if and only if for every first-order sentence φ of quantifier-rank $\leq m$ the following holds*

$$G_1 \models \varphi \Leftrightarrow G_2 \models \varphi.$$

Thus $G_1 \sim_m G_2$ iff no first-order sentence of quantifier-rank $\leq m$ can distinguish G_1 and G_2.

We shall use the following special case: if $G_1 \sim_m G_2$ and φ is a first-order sentence in prenex form with m quantifiers then $G_1 \models \varphi \Leftrightarrow G_2 \models \varphi$ holds. (To prove this one has to check that a distinguishing formula determines a winning strategy for player I, contradicting $G_1 \sim_m G_2$.)

Now we define second-order variants of the game corresponding to the second-order quantifiers introduced in Section 1.

Definition 2.2. The m-move *monadic second-order game* is defined similarly as the first-order game with the only difference that here at each move player I has two possibilities: either to make a *vertex-move* (i.e., select a graph and choose a vertex of it as in the first-order case), or to make a *set-move*: select a graph and choose a subset of its vertices, i.e., to add a new relation R_i of arity one to the structure. In any case player II must answer with a same kind of choice in the other graph. At the end of the game we consider isomorphism of the substructures spanned by the vertices chosen (edges *and* new relations) to decide the winner. The m-move *dyadic second-order game* is defined analogously only here players have *graph-moves* as well, i.e., in one step new relations of arity 2 can also be chosen. In the *second-order subgraph-game* graph-moves are restricted to the choice of subgraphs of the graphs given.

In the dyadic and subgraph-games we do not require the relations chosen to be symmetric. To each variant of the game there belongs a corresponding variant of equivalence (we do not use different notation for the different variants as hopefully it will be clear from the context what kind of equivalence is in question).

An important point for us is that the above theorem holds (with exactly the same proof) for the second-order variants of the game as well using the corresponding versions of sentences and equivalence, e.g. for the monadic case we have the following

Theorem: *$G_1 \sim_m G_2$ in the monadic second-order game if and only if for every monadic second-order sentence of quantifier-rank $\leq m$ the following holds*

$$G_1 \models \varphi \Leftrightarrow G_2 \models \varphi.$$

An important corollary of these theorems is that for any of the above games *the number of m-equivalence classes is finite* for every m as the set of sentences of quantifier-rank $\leq m$ is finite (modulo logical equivalence). We shall return to this point in Section 5.

3. Second-order definable graph-properties

In the first part of this section we give several examples of natural graph-properties belonging to the definability classes defined in Section 1.

Example 3.1. Let CONN $= \{G$: G is connected$\}$. Then CONN $\in \mathcal{M}$ holds. The defining sentence is the following:

$$\Phi \Leftrightarrow \forall R_1[((R_1 \neq \emptyset) \wedge (\overline{R_1} \neq \emptyset)) \rightarrow \exists x, y(R_1(x) \wedge \neg R_1(y) \wedge R(x, y))].$$

The sentence states that for every nonempty proper subset of the vertices there is an edge connecting this subset and it's complement. This is clearly equivalent to connectivity. We note that Fagin [4] and Hajek [8] showed that CONN$\notin\exists\mathcal{M}$.

Example 3.2. Let MATCH$=\{G$: G contains a perfect matching$\}$. Then MATCH$\in\exists\mathcal{S}$ holds. This is obvious.

Example 3.3. Let HAM$=\{G$: G contains a Hamiltonian cycle$\}$. Then HAM$\in\mathcal{S}$ holds. The defining sentence is the following:

$$\Phi \Leftrightarrow \exists R_1 \ (R_1 \text{ is a connected subgraph with all degrees} = 2).$$

We remark that here due to the above definition of connectivity this does *not* show that HAM is in $\exists\mathcal{S}$ (and as we shall see in fact HAM$\notin\exists\mathcal{S}$!).

Example 3.4. Let AUT$=\{G$: G has a proper automorphism$\}$. Then AUT$\in\mathcal{D}$ holds. The defining formula is the obvious transcription of the standard definition.

A further example is contained in the following proposition.

Proposition 3.1. *Let* PLA$=\{G$: *G is planar*$\}$. *Then* PLA$\in\mathcal{M}$ *holds.*

Proof. To show this we use Kuratowski's theorem. The property of containing a topological K_5 can be expressed as follows: there are 5 vertices $v_1, \ldots, v_5$ and 10 subsets $P_{1,2}, P_{1,3}, \ldots, P_{4,5}$ in G such that $\{v_i, v_j\} \subseteq P_{i,j}$, otherwise the $P_{i,j}$ are disjoint and connected for every $1 \leq i, j \leq 5$, $i \neq j$. A similar statement holds for $K_{3,3}$. □

If we add a further relation R' of arity 1 to the language of graphs (i.e., we consider graphs with a given subset V' of the vertices as basic structure), then the class

$$\{(G, V'):\ V' \text{ is an independent subset of maximal cardinality}\}$$

is evidently in $\mathcal{D}$. However, using König's theorem it can be shown to be in $\mathcal{S}$ as well.

It is evident that there are certain inclusions between the definability classes $\mathcal{M}$, $\exists\mathcal{S}$, $\mathcal{S}$ and $\mathcal{D}$. These inclusions are summarized in Fig. 1.

The question arising is whether any of these inclusions are proper? Theorem 3.1 below answers this question.

Theorem 3.1. *All inclusions of Fig.* 1. *are proper, i.e.*

(a) *$\mathcal{M}$ and $\exists\mathcal{S}$ are incomparable,*

(b) $\mathcal{M} \cup \exists\mathcal{S} \subsetneq \mathcal{S}$,

(c) $\mathcal{S} \subsetneq \mathcal{D}$.

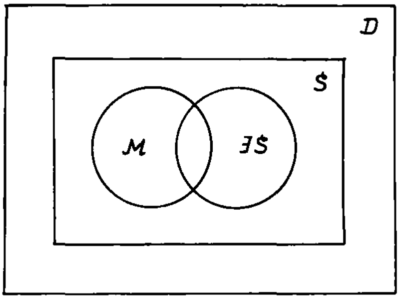


Fig. 1.

Proof. (a) Consider first the property MATCH. We show that MATCH$\notin\mathcal{M}$. Suppose that there is a monadic formula Φ defining MATCH. We can assume for simplicity that Φ is in prenex form with n quantifiers. Now consider the monadic n-move Fraïssé–Ehrenfeucht game, and let the number of equivalence classes be r. Taking the complete graphs $K_1, K_2, \ldots, K_{r+1}$, we can find i, j ($1 \leq i$, $j \leq r+1$, $i \neq j$) such that

$$K_i \sim_n K_j$$

(where $\sim$ means monadic equivalence). Let

$$G_1 = K_i \cup K_i, \qquad G_2 = K_i \cup K_j$$

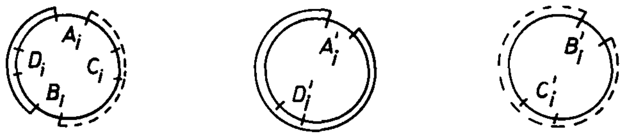
($\cup$ means union of two disjoint copies without adding any new vertices). Then $G_1 \sim_n G_2$ since player II can correspond the first K_i in G_1 to K_i in G_2, the second K_i in G_1 to K_j in G_2, and then win the game using the trivial strategy on the first pair and the (existing) winning strategy on the second pair. But then $\overline{G_1} \sim_n \overline{G_2}$, since complementing does not change the isomorphism of the graphs during the game. (*Note*: This does not hold for the weak dyadic game—as it is shown by this example.)

Now we have $\overline{G_1} = K_{i,i}$, $\overline{G_2} = K_{i,j}$, i.e., both are complete bipartite graphs, and as $i \neq j$, $G_1 \in$ MATCH, $G_2 \notin$ MATCH. But $G_1 \models \Phi \Leftrightarrow G_2 \models \Phi$, contradiction. Hence MATCH$\in \exists\mathcal{S} - \mathcal{M}$.

To show that $\mathcal{M} - \exists\mathcal{S}$ is nonempty, we consider the property CONN. Suppose CONN$\in\exists\mathcal{S}$ and Φ is an existential weak dyadic sentence defining CONN in prenex form

$$\exists R_1 \cdots \exists R_k \Psi$$

where $R_{s_1}, \ldots, R_{s_m}$ are variables of arity 1, $R_{t_1}, \ldots, R_{t_n}$ are variables of arity 2. Consider the sequence of graphs G_1^1, G_2^1, $G_3^1, \ldots$, where G_i^1 is a cycle of length i. Then $G_i^1 \models \Phi$ holds for every i. Let H_i^1 be the structure obtained by adding relations $R_1, \ldots, R_k$ so that $H_i^1 \models \Psi$ (such relations exist by the definition of Φ). A simple computation shows that there exists a constant c (depending on k) such that if $j = \lfloor c \log_2 i \rfloor$, than there will be 4 disjoint isomorphic arcs of length j on G_i^1



(a) (b)

Fig. 2

(i.e., isomorphic when we consider the restrictions of $R_1, \ldots, R_k$ on them). Choose 2 non-neighbouring ones, let them be A_i, B_i and the other two be C_i, D_i (see Fig. 2(a)). Now form graph G_i^2 as shown on Fig. 2(b): G_i^2 consists of two cycles, each of length $\geqslant 2j$ by definition. Define relations $R_1, \ldots, R_k$ on G_i^2 as on G_i^1 and denote the structure obtained this way by H_i^2. Now take the two ultraproducts (for definitions see Chang–Keisler [3])

$$\mathscr{H}_1 = \mathop{\mathsf{X}_u}_{i=1}^{\infty} H_i^1 \quad \text{and} \quad \mathscr{H}_2 = \mathop{\mathsf{X}_u}_{i=1}^{\infty} H_i^2.$$

Then using the properties of ultraproducts it can be shown that $\mathscr{H}_1$ is isomorphic to $\mathscr{H}_2$. (A detailed proof is given in [14]). We have $\mathscr{H}_1 \models \Psi$, so $\mathscr{H}_2 \models \Psi$ also holds. On the other hand as G_i^2 is disconnected, $G_i^2 \not\models \Phi$, hence $G_i^2 \models \forall R_1 \cdots \forall R_k \neg\Psi$, so $H_i^2 \models \neg\Psi$, and this gives $\mathscr{H}_2 \models \neg\Psi$, a contradiction.

(b) We show $\mathrm{HAM} \in \mathscr{S} - \mathscr{M} \cup \exists\mathscr{S}$. $\mathrm{HAM} \in \mathscr{S}$ is shown in Example 3.3. $\mathrm{HAM} \notin \mathscr{M}$ comes from the proof of the first part of (a) observing that $K_{i,i}$ is Hamiltonian while $K_{i,j}$ $(i \neq j)$ is not. $\mathrm{HAM} \notin \exists\mathscr{S}$ follows from the proof of the second part of (a) observing that G_i^1 is Hamiltonian while G_i^2 is not.

(c) We show $\mathrm{AUT} \in \mathscr{D} - \mathscr{S}$. $\mathrm{AUT} \in \mathscr{D}$ is evident (see Example 3.4). To prove $\mathrm{AUT} \notin \mathscr{S}$ consider the weakly dyadic Fraïssé–Ehrenfeucht game, and suppose Φ is a weakly dyadic sentence in prenex form with n quantifiers defining AUT. Let the number of equivalence classes in this kind of game of $n+1$ moves be r. Let P_i be a path of length i and consider $P_1, \ldots, P_{r+1}$. Then as in (a) we have some i, j $(1 \leqslant i, j \leqslant r+1,\ i \neq j)$ s.t. $P_i \sim_{n+1} P_j$. Now form G_1 and G_2 as in Fig. 3. Then $G_1 \in \mathrm{AUT}$, $G_2 \notin \mathrm{AUT}$, so $G_1 \models \Phi$, $G_2 \not\models \Phi$.

Thus if we show that $G_1 \sim_n G_2$ we are ready. This follows by the same kind of

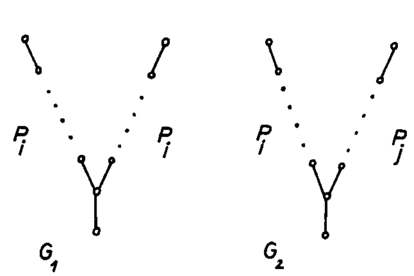


Fig. 3.

argument as in (a) with the only technical modification, that player II considers the two vertices of degree 3 to be selected in a fictive move 0. □

Remark. The second part of (a) can be proved with the method of Fagin [4] as well. This proof uses the proof of Lovász [13] of the theorem of Fagin.

4. The complexity of first-order definitions

As indicated in the introduction we are going to define a complexity measure related to the difficulty of defining a property on its instances on n vertices.

Definition 4.1. Let P be a class of graphs, let P_n be the class of graphs on n vertices belonging to P. For $n = 1, 2, \ldots$ let φ_n be first-order sentences s.t. if G is a graph on n vertices then $G \in P_n \Leftrightarrow G \models \varphi_n$. Suppose further that φ_n has minimal QN among all sentences satisfying the same condition. Then QN(P), the *quantifier-number complexity* of P is the sequence $(\mathrm{QN}_1(P), \mathrm{QN}_2(P), \ldots)$ s.t.

$$\mathrm{QN}_n(P) = \mathrm{QN}(\varphi_n).$$

The *quantifier-rank complexity* QR(P) is defined analogously.

Note that φ_n with minimal QN (or QR) always exists. We give some lower and upper bounds for the QN and QR complexity of some properties. Fraïssé–Ehrenfeucht games give lower bounds for QR, this translates trivially to lower bounds for QN. The first example is the standard example of short defining formulas (see e.g. Fischer–Rabin [7]).

Example 4.1. $\mathrm{QN}_n(\mathrm{CONN}) \leq 3\lceil \log_2(n-1) \rceil + 2$.

A graph G on n vertices is connected iff any two vertices can be connected by a path of length $\leq n-1$. Let $\varphi_k(x, y)$ be a sentence that holds iff x and y can be connected by a path of length $\leq 2^k$. Thus $\varphi_0(x, y) \Leftrightarrow R(x, y)$ and

$$\varphi_{k+1}(x, y) \Leftrightarrow \exists z(\varphi_k(x, z) \wedge \varphi_k(z, y)) \Leftrightarrow \exists z\, \forall u, v(((u = x \wedge v = z) \\ \vee (u = z \wedge v = y)) \rightarrow \varphi_k(u, v)).$$

In this sentence we use φ_k only once, so the result follows.

Corollary 4.1. $\mathrm{QN}_n(\mathrm{CONN}) \leq \frac{5}{2}\lceil \log_2(n-1) \rceil + 2$.

Proof. Let $\varphi_l^s(x, y)$ mean that x and y can be connected by a path of length $\leq s^l$, then

$$\varphi_{l+1}^s(x, y) \Leftrightarrow \exists z_1, \ldots, z_{s-1}\, \forall u, v[((u = x \wedge v = z_1) \vee \cdots) \rightarrow \varphi_l^s(u, v)].$$

Thus to describe connectivity we need $\lceil \log_s(n-1) \rceil (s+1) + 2$ quantifiers. The corollary follows with $s = 4$. □

The following example is an application of the above 'abbreviation trick.'

Theorem 4.1. *Let* BIP $=\{G$: G is bipartite$\}$. *Then holds*

$$\mathrm{QN}_n(\mathrm{BIP}) \le 4\lceil \log_2 n \rceil + 1.$$

Proof. A closed walk is a sequence $(x_1, e_1, x_2, e_2, \ldots, e_k, x_{k+1})$ of vertices and edges s.t. $e_i = (x_i, x_{i+1})$ and $x_{k+1} = x_1$, its length is k. Then a graph G on n vertices is not bipartite iff it contains a closed walk of length $n-1$ if n is even and of length n if n is odd. This is true because the length of a walk can always be increased by 2 repeating an edge in two directions. In the other direction, an odd closed walk must contain an odd cycle.

Let $\varphi_k(x, y)$ be a sentence that holds iff x and y can be connected be a walk of length k. Then

$$\varphi_k(x, y) \Leftrightarrow \begin{cases} \exists z(\varphi_{k/2}(x, z) \wedge \varphi_{k/2}(z, y)) & \text{if } k \text{ is even,} \\ \exists z, u(R(x, z) \wedge \varphi_{(k-1)/2}(z, u) \wedge \varphi_{(k-1)/2}(u, y)) & \text{if } k \text{ is odd.} \end{cases}$$

Thus $\varphi_n(x, y)$ can be written with $4\lceil \log_2 n \rceil$ quantifiers as in Example 4.1 and as we have seen

$$G \in \mathrm{BIP} \Leftrightarrow \begin{cases} \forall x(\neg\varphi_{n-1}(x, x)) & \text{if } n \text{ is even,} \\ \forall x(\neg\varphi_n(x, x)) & \text{if } n \text{ is odd.} \end{cases} \quad \square$$

A lower bound to QN(CONN) is given in the following theorem showing that the upper bound is optimal within a constant factor. The theorem was proved by Pósa [14] with different methods and independently by Immerman [9].

Theorem 4.2. $\mathrm{QR}_n(\mathrm{CONN}) \ge \lfloor \log_2 n \rfloor - 1$.

Proof (sketch). Let G_1 be a cycle of length n, G_2 be two cycles of length $\lfloor \frac{1}{2}n \rfloor$ and $\lceil \frac{1}{2}n \rceil$. Then obviously G_1 is connected, G_2 is not, so we have to show $G_1 \sim_m G_2$ in the first-order game where $m = \lfloor \log_2 n \rfloor - 2$. The winning strategy of player II can be given explicitly on an inductive way. Player II tries to maintain the following assumptions:

(a) after s moves the vertices chosen on G_1 are $P_1, \ldots, P_k$, $Q_1, \ldots, Q_l$ in circular order $(k+l=s)$, the corresponding points on the first cycle of G_2 are $P'_1, \ldots, P'_k$ in the same circular order, and $Q'_1, \ldots, Q'_l$ on the second cycle of G_2 in the same circular order;

(b) considering the arcs P_iP_{i+1}, $P'_iP'_{i+1}$, Q_jQ_{j+1}, $Q'_jQ'_{j+1}$

$$\begin{aligned} P_iP_{i+1} &\sim_{m_s} P'_iP'_{i+1} & i &= 1, \ldots, k-1 \\ Q_jQ_{j+1} &\sim_{m_s} Q'_jQ'_{j+1} & j &= 1, \ldots, k-1 \end{aligned}$$

where $m_s = \lfloor \log_2 n \rfloor - s - 2$;

(c) the length of the arcs P_kQ_1, QP_1, $P'_kP'_1$, $Q'Q'_1$ is $\ge \lfloor n/2^s \rfloor$.

It can be shown that these assumptions can be maintained independently of the choice of player I. Thus the theorem follows. □

Corollary 4.2. $\mathrm{QN}_n(\mathrm{CONN}) \geq \lfloor \log_2 n \rfloor - 1$.

It is easy to see that for any graph-property P $\mathrm{QN}(P) \leq n$ holds because with n variables we can simply describe the adjacencies of every graph on vertices belonging to P (in general this will be a very long sentence). Thus the maximal lower bound that can be obtained for QN is linear. The next theorem shows that such lower bounds hold for some natural properties.

Theorem 4.3. $\mathrm{QR}_n(\mathrm{HAM}) \geq \lfloor \frac{1}{2} n \rfloor$.

Proof. Suppose first n even. Then let

$$G_1 = K_{n/2,n/2}, \qquad G_2 = K_{n/2-1,n/2+1}.$$

Clearly $G_1 \in \mathrm{HAM}$. On the other hand $G_2 \notin \mathrm{HAM}$ (a Hamilton cycle would alternate between the two colour-classes). Thus it is enough to show that

$$G_1 \sim_{n/2-1} G_2$$

holds. The winning strategy of player II is to correspond the first class of G_1 to the $\frac{1}{2}n-1$-class of G_2, the second class of G_1 to the $\frac{1}{2}n+1$-class of G_2, and then to select corresponding vertices from the corresponding classes. As the spanned subgraphs will always be complete bipartite graphs of the same size he wins until the smallest class is not exhausted. This cannot happen in $\frac{1}{2}n-1$ moves.

If n is odd, a small modification is needed. Construct G_1 as follows: take $K_{(n-1)/2,(n-1)/2}$ and join a new vertex to all the other ones. G_2 is obtained from $K_{(n-3)/2,(n+1)/2}$ in the same way. Then G_1 is Hamiltonian, G_2 is not ($K_{(n-3)/2,(n+1)/2}$ does not contain a Hamiltonian path), and

$$G_1 \sim_{(n-3)/2} G_2$$

holds similarly as in the even case, only player II now lets the new vertices correspond. □

Corollary 4.3. $\mathrm{QR}_n(\mathrm{MATCH}) \geq n/2$ *if n is even.*

The same proof can be applied to this case as well.

It can be remarked that while HAM is computationally harder than MATCH, this difference is not reflected in their QN.

5. Further remarks

In this section we mention some related problems that may merit further investigation or where partial results have been obtained.

(1) The number of equivalence classes in the first-order Fraïssé–Ehrenfeucht game (or in other games respectively) indicates the power of first-order formulas so it may be of some interest to give bounds for these numbers. It is noted in [17] that the function $F(m)$ where $F(m)$ is the number of m-equivalence classes is not recursive (this follows from the undecidability of the theory of graphs). The upper bound that can be obtained following Ladner is a non-elementary function of m (in the sense of Kalmár). One can show a non-elementary lower bound as well. The idea is that high-order formulas can count very large sets and this can be coded back to graphs that correspond to the iterated subsets construction.

(2) Problems related to those investigated in this paper emerged in the theory of relational data-bases (see e.g. Chandra–Harel [2], Vardi [18], Immerman [11]). An interesting definability tool used there is the operation of transitive closure and minimal fixpoint. I.e., one is allowed to take the transitive closure of a definable relation, and the minimal fixpoint of a monoton formula when interpreted as a monoton operator on k-ary relations (e.g. the standard definition of transitive closure). One can show that the class of bipartite graphs can be defined using transitive closure.

(3) Consider graphs with four distinguished vertices a_1, b_1, a_2, b_2. Take the class of structures

$$\{(G, a_1, b_1, a_2, b_2)\colon \text{ there are disjoint paths from } a_1 \text{ to } b_1, \text{ and from } a_2 \text{ to } b_2\}.$$

What is the QR-complexity of this property? The results of Seymour [15] may be relevant. This question is related to the QR-complexity of planarity as well.

(4) A variant of the problems in Section 4 is the following: let be given two graphs G_1 and G_2 on $\leq n$ vertices. Find a first-order formula s.t. $G_1 \models \varphi$ and $G_2 \not\models \varphi$. What can be said about the QR-complexity of φ? It seems reasonable to assume that G_1 and G_2 are connected, with all degrees $\leq k$. What is the maximum of the QR of the minimal φ for all such G_1, G_2? One should guess $O(c_k \log_2 n)$, but from Immerman [9] a lower bound of $2^{\sqrt{\log_2 n}}$ follows. For trees one can show $O(\log^2 n)$ and good bounds can be given in general for graphs with small separators but these are only sufficient conditions.

(5) Immerman [10] shows interesting relations between QN-complexity and classical notions of complexity with an important assumption: one has a *linear order* on the vertices. This makes it possible to simulate general computational devices by formulas. Is there a structured model for computation on graphs (in the sense of Borodin [1]) that is related on a similar way to formulas without the linear order?

Acknowledgement

I am grateful to Prof. L. Lovász for his remarks.

References

[1] A. Borodin, Structured vs. general models in computational complexity, Logic and algorithmic International Symposium in honour of E. Specker (University of Geneva, 1982) 47–66.
[2] A.K. Chandra and D. Harel, Structure and complexity of relational queries, IEEE FOCS (1980) 333–347.
[3] C.C. Chang and H.J. Keisler, Model Theory (North-Holland, Amsterdam, 1973).
[4] R. Fagin, Monadic generalized spectra, Z. Math. Logik (1975) 89–96.
[5] R. Fagin, Generalized first-order spectra and polynomial-time recognizable sets, in: R. Karp, ed., Complexity of Computation, SIAM-AMS Proc. 7 (1974) 43–73.
[6] J. Ferrante and C.W. Rackoff, The Computational Complexity of Logical Theories, Lecture Notes in Math. 718 (1979).
[7] M.J. Fischer and M.O. Rabin, Super-exponential complexity of Presburger-arithmetic, in: R. Karp, ed., Complexity of Computation SIAM-AMS Proc. 7 (1974) 27–42.
[8] P. Hajek and T. Havránek, Mechanizing Hypothesis Formation (Springer, New York, 1978)
[9] N. Immerman, Number of quantifiers is better than number of tape cells, JCSS 22 (1981), 384–406.
[10] N. Immerman, Upper and lower bounds for first-order expressibility, IEEE FOCS (1980) 74–82.
[11] N. Immerman, Relational queries computable in polynomial time, ACM STOC (1982) 147–152.
[12] R.E. Ladner, Application of model theoretic games to discrete linear orders and finite automata, Inf. and Cont. (1977), 261–303.
[13] L. Lovász, unpublished.
[14] L. Pósa, unpublished.
[15] P.D. Seymour, Disjoint paths in graphs, Discrete Math. 29 (1980) 293–309.
[16] Gy. Turán, Gráf-tulajdonságok lokális felismerhetöségéröl és definiálhatóságáról, Thesis, 1980. (Univ. of Szeged).
[17] Gy. Turán, Cellular graph-automata and second-order definable graph-properties, in: F. Gécseg, ed., Fundamentals of Computing, Lecture Notes Computer Sci. 117 (1981) 384–393.
[18] M.Y. Vardi, The complexity of relational query languages, ACM STOC (1982) 137–146.