

THE MONADIC SECOND-ORDER LOGIC OF GRAPHS IV: DEFINABILITY PROPERTIES OF EQUATIONAL GRAPHS*

Bruno COURCELLE

*Université Bordeaux I, Laboratoire d'Informatique (Formation associée au CNRS), 351 Cours
de la Libération, 33405 Talence, France*

Communicated by Y. Gurevich

Received 25 July 1988; revised 28 November 1989

We establish that every equational graph can be characterized, up to isomorphism, by a formula of monadic second-order logic. It follows that the isomorphism of two equational graphs is decidable. We also establish that a graph specified in an equational graph by monadic second-order formulas is equational.

Introduction

This paper continues the study of the monadic second-order logic of countable graphs that has been initiated in Courcelle [11]. The aspects of this research concerning finite graphs are presented in Courcelle [9, 10, 12–15].

In order to introduce our results, we review a few basic definitions. A graph is *definable* if it is characterized, up to isomorphism, by a closed formula of monadic second-order logic. It is *equational* if it is a component of the least solution of a system of equations in graphs. A graph is *of finite width* if it is expressible by a finite or infinite graph expression constructed with finitely many of the graph operations introduced in Bauderon and Courcelle [4] (see also [15]). Not every infinite graph is of finite width, but every equational graph is. All these properties hold in a more or less trivial way, for finite graphs. We only consider countably infinite graphs in this paper.

It is proved in [11] that every definable graph of finite width is equational. The main result of this paper is that every equational graph is definable. Hence a graph is equational iff it is definable and is of finite width. (A graph is of finite width iff it is of finite tree-width, in the sense of Robertson and Seymour [21].)

Since the monadic second-order theory of an equational graph is decidable (Courcelle [11]), it follows that the isomorphism of two equational graphs is decidable.

* This work has been supported by the “Programme de Recherches Coordonnées: Mathématiques et Informatique”, and by the ESPRIT-Basic Research Action contract No. 3299, “Computing by graph transformations”.

Reprints can be obtained on request by electronic mail on UUCP network: courcell@geocub.greco-prog.fr.

We also introduce the *relative definability* of graphs. A graph G is definable in a graph H if, roughly speaking, its vertices and edges can be defined from those of H , by means of monadic second-order formulas. (A precise definition is given in Section 3.) The connected component of a designated vertex of a graph H is definable in H in this sense. We also establish that, if a graph is definable in an equational graph, then it is equational.

We now sketch the technique used to establish the main result. An equational graph G can be described by an infinite regular tree called its *syntactic tree*. This tree can be defined in G in the above sense. A formula can express that the defined tree is the desired syntactic tree. Another formula can “verify” that the considered graph is isomorphic to the “value graph” of the defined syntactic tree.

The paper is organized as follows. Section 1 gives preliminary definitions on graphs and graph operations. Section 2 introduces systems of graph equations and states a few preliminary results concerning them. Section 3 defines monadic second-order logic and definability notions concerning graphs. Sections 4 and 5 contain the proof of the main theorem. The result on the relative definability of equational graphs is proved in Section 6. An appendix contains some technical proofs that are omitted in Section 2.

1. Notations and definitions concerning graphs

We denote by \mathbb{N} the set of nonnegative integers, and by \mathbb{N}_+ the set of positive ones. We denote by $[n]$ the interval $\{1, 2, 3, \dots, n\}$ for $n \geq 0$ (with $[0] = \emptyset$).

For sets A and B , we denote by $A - B$ the set $\{a \in A \mid a \notin B\}$. By a *partition* of a set A , we mean an indexed set $(A_i)_{i \in I}$ of possibly empty, pairwise disjoint subsets of A , such that $A = \bigcup \{A_i \mid i \in I\}$.

The domain of a partial mapping $f: A \rightarrow B$ is denoted by $\mathbf{Dom}(f)$. The restriction of f to a subset A' of A is denoted by $f \upharpoonright A'$. The partial mapping with an empty domain is denoted by \emptyset , as the empty set. If two partial mappings $f: A \rightarrow B$ and $f': A' \rightarrow B$ coincide on $\mathbf{Dom}(f) \cap \mathbf{Dom}(f')$ we denote by $f \cup f'$ their common extension into a partial mapping: $A \cup A' \rightarrow B$, with domain $\mathbf{Dom}(f) \cup \mathbf{Dom}(f')$.

The cardinality of a set A is denoted by $\mathbf{Card}(A)$. The powerset of A is denoted by $\mathcal{P}(A)$.

A binary relation R on a set A is considered as a subset of $A \times A$. Hence, $x R y$ and $(x, y) \in R$ are equivalent notations. The transitive closure of R is denoted by R^+ , and its reflexive and transitive closure is denoted by R^* . The set of equivalence relations on A is denoted by $\mathbf{Eq}(A)$.

The set of nonempty sequences of elements of a set A is denoted by A^+ . The generic sequence is denoted by (a_1, \dots, a_n) with commas and parentheses. The empty sequence is denoted by $()$, and A^* is $A^+ \cup \{()\}$. When A is an alphabet, i.e., when its elements are letters, a sequence (a_1, \dots, a_n) in A^+ can be written

unambiguously $a_1 a_2 \cdots a_n$. The empty sequence is denoted by ε , a special symbol reserved for this purpose. The elements of A^* are called words. The length of a sequence μ is denoted by $|\mu|$.

We shall use $:=$ for “equal by definition”, i.e., for introducing a new notation. The notation $:\Leftrightarrow$ will be used similarly for defining logical conditions.

Graphs

As in [4, 10–14], we deal with labeled, directed hypergraphs equipped with a sequence of distinguished vertices called the sequence of sources. The labels are chosen in a ranked alphabet, i.e., in a set A , each element of which has an associated integer (in \mathbb{N}) that we call its *type*. The type mapping is $\tau: A \rightarrow \mathbb{N}$. The type of the label of a hyperedge must be equal to the length of its sequence of vertices. (This type may be 0, i.e., we allow hyperedges with no vertex.) In order to shorten the statements we shall simply call *graphs* these hypergraphs, and *edges* their hyperedges.

1.1. Definitions (Concrete and abstract graphs). Let A (and τ) be as above, let $n \in \mathbb{N}$. A *concrete n -graph* is a quintuple $G = \langle \mathbf{V}_G, \mathbf{E}_G, \mathbf{lab}_G, \mathbf{vert}_G, \mathbf{src}_G \rangle$ where:

- \mathbf{V}_G is the set of vertices of the graph,
- \mathbf{E}_G is its set of edges,
- $\mathbf{lab}_G: \mathbf{E}_G \rightarrow A$ is a total mapping that assigns to each edge of G a label in the alphabet A ,
- $\mathbf{vert}_G: \mathbf{E}_G \rightarrow \mathbf{V}_G^*$ is a total mapping that associates with an edge e of G , the sequence of its vertices (this sequence must be of length $\tau(e) := \tau(\mathbf{lab}_G(e))$ and its i th element is also denoted by $\mathbf{vert}_G(e, i)$), and finally
- \mathbf{src}_G is a sequence of length n in \mathbf{V}_G^* (or equivalently a mapping: $[n] \rightarrow \mathbf{V}_G$), called the *sequence of sources*.

We shall denote by $\mathbf{src}_G(i)$ the i th element of the sequence \mathbf{src}_G . (If $n = 0$, then G has no source.) “Source” is just an easy sounding word for “distinguished vertex”. There is no notion of flow involved. The integer n is the *type* of G .

Whenever we need to specify the alphabet A , we say that G is a *concrete n -graph over A* . A *concrete graph* is a concrete n -graph for some $n \geq 0$.

Let G and H be concrete graphs of the same type n . A *homomorphism* $G \rightarrow H$ is a pair of mappings $h = (h_{\mathbf{V}}, h_{\mathbf{E}})$ where $h_{\mathbf{V}}: \mathbf{V}_G \rightarrow \mathbf{V}_H$, $h_{\mathbf{E}}: \mathbf{E}_G \rightarrow \mathbf{E}_H$, and such that:

$$\begin{aligned} \mathbf{lab}_H \circ h_{\mathbf{E}} &= \mathbf{lab}_G, \\ h_{\mathbf{V}}(\mathbf{vert}_G(e, i)) &= \mathbf{vert}_H(h_{\mathbf{E}}(e), i) \quad \text{for all } i \in [\tau(e)], \text{ all } e \in \mathbf{E}_G, \\ h_{\mathbf{V}}(\mathbf{src}_G(i)) &= \mathbf{src}_H(i) \quad \text{for all } i \in [n]. \end{aligned}$$

If no ambiguity can arise, we denote $h_{\mathbf{V}}$ and $h_{\mathbf{E}}$ by h .

An *isomorphism* is a homomorphism such that $h_{\mathbf{V}}$ and $h_{\mathbf{E}}$ are bijective. The isomorphism class of a concrete graph is called an *abstract graph*, or simply a *graph* in the sequel.

A graph G is *finite* if \mathbf{V}_G and \mathbf{E}_G are finite. Otherwise, a graph has at most countably many vertices and edges.

We denote by $\mathbf{FCG}(A)_n$, by $\mathbf{FCG}(A)$, by $\mathbf{FG}(A)_n$, and by $\mathbf{FG}(A)$, the sets of finite concrete n -graphs, of finite concrete graphs, of finite n -graphs, and of finite graphs respectively, over A . The notations $\mathbf{CG}(A)_n$, $\mathbf{CG}(A)$, $\mathbf{G}(A)_n$ and $\mathbf{G}(A)$ are used similarly for the corresponding sets of finite and infinite graphs.

A vertex v belongs to an edge e , if $v = \mathbf{vert}_G(e, i)$ for some i . A vertex is *isolated* if it belongs to no edge. An *internal* vertex of G is a vertex that is not a source. The sources are called *external* vertices. We denote by \mathbf{IV}_G the set of internal vertices of G . A graph with no edge and no internal vertex is *degenerated*.

For every n in \mathbb{N} , we denote by \mathbf{n} the unique degenerated n -graph with n pairwise distinct sources. For every a in A of type n , we denote by a the n -graph with a single edge e labeled by a , no internal vertex, and a sequence of n pairwise distinct sources equal to the sequence of vertices of e .

1.2. Definition (Subgraphs). Let G be a concrete graph. A *subgraph* of G is a concrete graph H such that $\mathbf{V}_H \subseteq \mathbf{V}_G$, $\mathbf{E}_H \subseteq \mathbf{E}_G$, $\mathbf{lab}_H = \mathbf{lab}_G \upharpoonright \mathbf{E}_H$, $\mathbf{vert}_H = \mathbf{vert}_G \upharpoonright \mathbf{E}_H$ and \mathbf{src}_H is the sequence obtained from \mathbf{src}_G by the deletion of the vertices of $\mathbf{V}_G - \mathbf{V}_H$. (Hence $\tau(H) \leq \tau(G)$.) This is denoted by $H \subseteq G$.

If $V \subseteq \mathbf{V}_G$ and $E \subseteq \mathbf{E}_G$, we denote by $G \upharpoonright (V \cup E)$ the subgraph H of G such that $\mathbf{E}_H = E$ and $\mathbf{V}_H = V \cup \{v \in \mathbf{V}_G \mid v \text{ is a vertex belonging to an edge } e \text{ of } E\}$.

Two concrete graphs G and H are *disjoint* if $\mathbf{V}_G \cap \mathbf{V}_H = \emptyset$ and $\mathbf{E}_G \cap \mathbf{E}_H = \emptyset$. The *union* of two disjoint concrete 0-graphs G and H is the 0-graph K such that $\mathbf{V}_K = \mathbf{V}_G \cup \mathbf{V}_H$, $\mathbf{E}_K = \mathbf{E}_G \cup \mathbf{E}_H$, $G \subseteq K$, and $H \subseteq K$. This notion extends to finite or countable sets of pairwise distinct 0-graphs in an obvious way.

1.3. Definition (Quotient graphs). Let G be a concrete graph, let \approx be an equivalence relation on \mathbf{V}_G . We denote by $[v]$ the equivalence class of v w.r.t. \approx .

We denote by G/\approx the concrete graph H such that $\mathbf{V}_H = \mathbf{V}_G/\approx$, $\mathbf{E}_H = \mathbf{E}_G$, $\mathbf{lab}_H = \mathbf{lab}_G$, $\mathbf{vert}_H(e, i) = [\mathbf{vert}_G(e, i)]$ for all $e \in \mathbf{E}_H (= \mathbf{E}_G)$, all $i \in [\tau(e)]$, and $\mathbf{src}_H(i) = [\mathbf{src}_G(i)]$ for all $i \in [\tau(G)]$.

We call G/\approx the *quotient graph* of G by \approx . There is a canonical surjective homomorphism: $G \rightarrow G/\approx$.

If G is an abstract graph, then G/\approx is the isomorphism class of \tilde{G}/\approx where \tilde{G} is any concrete graph representing G .

We recall from Bauderon and Courcelle [4] the definition of graph substitutions. This notion is fundamental in the theory of systems of equations.

1.4. Definition (Graph substitutions). We first give a basic definition, and then we proceed by successive extensions.

Basic definition. Let H, G_1, \dots, G_n be pairwise disjoint concrete graphs. Let e_1, \dots, e_n be n distinct edges of H , such that $\tau(e_i) = \tau(G_i)$ for all $i = 1, \dots, n$. We denote by $G = H[G_1/e_1, \dots, G_n/e_n]$ the result of the *simultaneous substitution* in H , of G_1 for e_1, \dots , and of G_n for e_n . This graph will also be denoted by $H[G_1, \dots, G_n]$, when the sequence e_1, \dots, e_n is known from the context.

Formally, G is the $\tau(H)$ -graph defined as follows.

$G = K/\approx$ where:

$$\mathbf{V}_K = \mathbf{V}_H \cup \mathbf{V}_{G_1} \cup \dots \cup \mathbf{V}_{G_n},$$

$$\mathbf{E}_K = \mathbf{E}_H \cup \mathbf{E}_{G_1} \cup \dots \cup \mathbf{E}_{G_n} - \{e_1, \dots, e_n\},$$

$$\mathbf{lab}_K = (\mathbf{lab}_H \cup \mathbf{lab}_{G_1} \cup \dots \cup \mathbf{lab}_{G_n}) \upharpoonright \mathbf{E}_K,$$

$$\mathbf{vert}_K = (\mathbf{vert}_H \cup \mathbf{vert}_{G_1} \cup \dots \cup \mathbf{vert}_{G_n}) \upharpoonright \mathbf{E}_K,$$

$$\mathbf{src}_K(i) = f_{\mathbf{v}}(\mathbf{src}_H(i)), \text{ for all } i \text{ in } [\tau(H)], \text{ where}$$

$f_{\mathbf{v}}$ is the canonical surjection: $\mathbf{V}_K \rightarrow \mathbf{V}_G$, and

\approx is the equivalence relation on \mathbf{V}_K generated by the set of pairs of vertices:

$$\{(\mathbf{vert}_H(e_i, j), \mathbf{src}_{G_i}(j)) \mid i \in [n], j \in [\tau(e_i)]\}.$$

We let $f_{\mathbf{E}}$ be the identity: $\mathbf{E}_K \rightarrow \mathbf{E}_G$, hence, $f = (f_{\mathbf{v}}, f_{\mathbf{E}})$ is the canonical surjective graph homomorphism: $K \rightarrow K/\approx = G$.

We also let:

$$\bar{g}_i = \text{the inclusion homomorphism: } G_i^0 \rightarrow K,$$

(the notation G_i^0 is introduced in Definition 1.5 below) and we let:

$$g_i = \text{the homomorphism } f \circ \bar{g}_i: G_i^0 \rightarrow G \text{ for } i = 1, \dots, n.$$

We call (g_1, \dots, g_n) the *tuple of homomorphisms associated with the substitution* in H , of G_1 for e_1, \dots , and of G_n for e_n , (or more shortly *associated with* $H[G_1/e_1, \dots, G_n/e_n]$).

We also say that a vertex x of H , is *mapped into the vertex* $f_{\mathbf{v}}(x)$ of $H[G_1, \dots, G_n]$.

Let (G'_1, \dots, G'_n) be another n -tuple of graphs, of the respective types of G_1, \dots, G_n . Let $w_i: G_i \rightarrow G'_i$ be a homomorphism for each i . We shall denote by $H[w_1, \dots, w_n]$ the homomorphism $h: G = H[G_1, \dots, G_n] \rightarrow G' = H[G'_1, \dots, G'_n]$ defined as follows. (This notation assumes that the sequence of edges to which substitutions are made, is known from the context.) In order to define it, we let K' and f' be as K and f above, defined w.r.t. H and G'_1, \dots, G'_n . We let:

$$w_{\mathbf{v}}: \mathbf{V}_K \rightarrow \mathbf{V}_{K'} \text{ be } \mathbf{id}_{\mathbf{V}_H} \cup w_{1\mathbf{v}} \cup \dots \cup w_{n\mathbf{v}}.$$

and $w_{\mathbf{E}}: \mathbf{E}_K \rightarrow \mathbf{E}_{K'}$ be defined similarly. Hence w is a homomorphism: $K \rightarrow K'$.

There exists a unique homomorphism, denoted by $H[w_1, \dots, w_n]$, making the following diagram commutative:

$$\begin{array}{ccc} K & \xrightarrow{w} & K' \\ f \downarrow & & \downarrow f' \\ H[G_1, \dots, G_n] = G & \xrightarrow{H[w_1, \dots, w_n]} & G' = H[G'_1, \dots, G'_n] \end{array}$$

If (g_1, \dots, g_n) and (g'_1, \dots, g'_n) are the tuples of homomorphisms associated with $H[G_1, \dots, G_n]$ and $H[G'_1, \dots, G'_n]$, then one has the following commutative diagram:

$$\begin{array}{ccc} G_i^0 & \xrightarrow{w_i} & G_i'^0 \\ g_i \downarrow & & \downarrow g'_i \\ H[G_1, \dots, G_n] & \xrightarrow{H[w_1, \dots, w_n]} & H[G'_1, \dots, G'_n] \end{array}$$

We now consider several extensions.

First extension. The definition of $H[G_1/e_1, \dots, G_n/e_n]$ extends as follows to the case where the graphs H, G_1, \dots, G_n are not pairwise disjoint. One constructs n concrete graphs $\tilde{G}_1, \dots, \tilde{G}_n$, respectively isomorphic to G_1, \dots, G_n , and such that $H, \tilde{G}_1, \dots, \tilde{G}_n$ are pairwise disjoint, and one lets

$$H[G_1/e_1, \dots, G_n/e_n] := H[\tilde{G}_1/e_1, \dots, \tilde{G}_n/e_n].$$

Again we omit the mention of e_1, \dots, e_n if they are known from the context. The definitions of the tuple of homomorphisms, and of $H[w_1, \dots, w_n]$ extend immediately. (The graphs $\tilde{G}_1, \dots, \tilde{G}_n$ are constructed in some canonical way that we need not specify. Hence the result of the substitution is uniquely defined.)

Second extension. Let $\{e_i \mid i \in I\}$ be a possibly infinite indexed set of pairwise distinct edges of a graph H . Let G_i be a concrete graph of the type of e_i for each i in I . Then, one can define $H[G_i/e_i; i \in I]$ by substituting simultaneously G_i for e_i , for all i in I . The definitions given above for the case of finitely many edges extend immediately.

Third extension. Let U be a finite or infinite ranked alphabet, let H belong to $\mathbf{CG}(A \cup U)$ and G_u be a graph of the type of u for each u in U . Then, we denote by $H[G_u/u; u \in U]$ the result of the simultaneous substitution of G_u for all edges of H labelled by u , and for all symbols u . We shall mainly do this in cases where U is a finite set enumerated as $\{u_1, \dots, u_n\}$, and the graph to be substituted for u_i is G_i . We shall use in this case the notation $H[G_1/u_1, \dots, G_n/u_n]$ or the notation $H[G_1, \dots, G_n]$ if U is known from the context.

This definition also extends to the case of abstract graphs.

1.5. Definition (Graph operations). A *graph operation* is a mapping $f: \mathbf{CG}(A)_{n_1} \times \dots \times \mathbf{CG}(A)_{n_k} \rightarrow \mathbf{CG}(A)_n$ of the form:

$$f(G_1, \dots, G_k) := H[G_1/e_1, \dots, G_k/e_k]$$

for some fixed finite graph H of type n and some fixed edges e_1, \dots, e_k of H , of respective types n_1, \dots, n_k . We say that f is defined by the tuple (H, e_1, \dots, e_k) . It operates on abstract graphs in an obvious way.

We now recall the definitions of the basic graph operations introduced in [4]. The first operation is the *disjoint sum*, $\oplus_{n,m}$ of profile $n \times m \rightarrow n + m$, defined by (H, e, e') , where H is the graph with two edges e and e' , with set of vertices $[n + m]$, and such that:

$$\begin{aligned} \mathbf{vert}_H(e) &:= (1, \dots, n) && \text{with } n = \tau(e), \\ \mathbf{vert}_H(e') &:= (n + 1, \dots, n + m) && \text{with } m = \tau(e'), \\ \mathbf{src}_H &= (1, \dots, n + m). \end{aligned}$$

The second operation is the *source redefinition*, associated with a mapping α from $[p]$ to $[n]$. It is denoted by σ_α , has the profile $n \rightarrow p$, and is defined by the pair (H, e) where H consists of a single edge e of type n , has the set of vertices $[n]$, and is such that $\mathbf{vert}_H(e) := (1, \dots, n)$ and $\mathbf{src}_H := (\alpha(1), \dots, \alpha(p))$. (If G is as in Definition 1.1, then $\sigma_\alpha(G) = \langle \mathbf{V}_G, \mathbf{E}_G, \mathbf{lab}_G, \mathbf{vert}_G, \mathbf{src}_G \circ \alpha \rangle$). If $p = 0$, then α is necessarily the empty map (always denoted by \emptyset), and $\sigma_\alpha(G)$ is the 0-graph obtained from G by “forgetting” its sources. We also denote it by G^0 .

When p is small it is convenient to write $\sigma_{i_1, i_2, \dots, i_p}(G)$ instead of $\sigma_\alpha(G)$, with $i_j = \alpha(j)$ for $j = 1, \dots, p$.

The third operation is the *source fusion*. For every equivalence relation δ on $[n]$, the mapping $\theta_\delta: \mathbf{CG}(A)_n \rightarrow \mathbf{CG}(A)_n$ is associated with the pair (H, e) where H consists of a single edge e of type n , with vertices such that $\mathbf{vert}_H(e, i) = \mathbf{vert}_H(e, j)$ iff i and j are equivalent w.r.t. δ . Its sequence of sources is equal to $\mathbf{vert}_H(e)$. Intuitively, the graph $\theta_\delta(G)$ is obtained from G by fusing the i th and j th sources, whenever i and j are equivalent w.r.t. δ . If δ is the equivalence relation on $[n]$ generated by a single pair (i, j) , then we denote θ_δ by $\theta_{i,j}$. It is clear that if δ is the equivalence relation generated by a set of pairs $\{(i_1, j_1), \dots, (i_k, j_k)\}$, then:

$$\theta_\delta = \theta_{i_1, j_1} \circ \dots \circ \theta_{i_k, j_k}.$$

1.6. Definition (Paths). With a graph G we associate the set

$$\begin{aligned} \mathbf{P}(G) := \{ & (v, e, i, j, v') \mid v, v' \in \mathbf{V}_G, e \in \mathbf{E}_G, i, j \in [\tau(e)], i \neq j, \\ & v = \mathbf{vert}_G(e, i), v' = \mathbf{vert}_G(e, j) \}. \end{aligned}$$

A *path from v to v'* in G , (or *linking v to v'*), is a nonempty sequence π in $\mathbf{P}(G)^+$ of the form:

$$\pi = (v, e_1, i_1, j_1, v_1) (v_1, e_2, i_2, j_2, v_2) \cdots (v_{k-1}, e_k, i_k, j_k, v').$$

Its length is k , and its *sequence of vertices* is defined as:

$$\mathbf{vert}(\pi) := (v, v_1, v_2, \dots, v_{k-1}, v').$$

If v_1, \dots, v_{k-1} are internal vertices, then π is an *internal path*. Note that v and v' may be either external or internal.

1.7. Definition (Internally connected components). Let G be a concrete graph. A subgraph H of G is a *concrete internally connected component* of G , a *cicc* for short, if it is:

- (1) either an edge, all vertices of which are external,
- (2) or is $G \upharpoonright (\{v\} \cup E(v))$ where $v \in \mathbf{IV}_G$ and $E(v)$ is defined below.

For every vertex v in \mathbf{IV}_G , we denote by $E(v)$ the set of all edges of G to which belongs a vertex v' in \mathbf{IV}_G , that is either equal to v , or linked to v by an internal path. It is clear that any two distinct vertices of a cicc are linked by an internal path.

A graph is *internally connected* if it is not degenerated and has a unique cicc.

We now give another characterization of the notion of a cicc. By an *item* of a graph (resp. an *internal item*), we mean an edge or a vertex (resp. an edge or an internal vertex). Two internal items i and i' of a graph G are *internally linked*, if there exist two internal vertices v and v' linked by an internal path, and such that, either $i = v$, or v belongs to i , and either $i' = v'$, or v' belongs to i' . This is denoted by $i \approx i'$. The relation \approx is an equivalence relation on the set of internal items of G . It is clear that a cicc of G is a subgraph of G of the form $G \upharpoonright C$, where C is a subset of $\mathbf{V}_G \cup \mathbf{E}_G$, that is is the equivalence class modulo \approx of an internal item of G .

Every path of G^0 is internal. A subgraph H of G such that H^0 is a cicc of G^0 is called a *concrete connected component* of G (a *ccc* for short). This notion is equivalent to the classical notion of a connected component, in the case of graphs with edges of type at most 2. See Example 1.11 below. A graph is *connected* if it has a unique ccc.

For every concrete graph G , we denote by $\mathbf{CICC}(G)$ and by $\mathbf{CCC}(G)$ its sets of concrete internally connected components, and of concrete connected components respectively.

An *internally connected component* (an *icc*) of G (resp. a *connected component* (a *cc*)) is the isomorphism class of a cicc (resp. of a ccc) of G . We denote by $\mathbf{ICC}(G)$ and $\mathbf{CC}(G)$ the sets of icc's, and of cc's of G respectively.

The notions of icc and of cc can be defined for abstract graphs in an obvious way.

In order to express formally a graph as a combination of its cicc's, we introduce new graph operations.

1.8. Definition (Parallel composition). For every $n \geq 0$ we define a graph operation $//_n$ of profile $n \times n \rightarrow n$. It is associated with the tuple (H, e, e') , where H is the graph with set of vertices $[n]$, with two edges e and e' , and such that:

$$\mathbf{vert}_H(e) = \mathbf{vert}_H(e') = \mathbf{src}_H = (1, \dots, n).$$

Intuitively, $G //_n G'$ is the result of gluing G and G' by their sources, in their order. Note that $G //_0 G' = G \oplus G'$ if G and G' are 0-graphs. We shall omit the subscript n , i.e., we shall overload the symbol $//$, and we shall do the same for \oplus .

We now extend it to indexed sets of graphs. Let $(G_i)_{i \in I}$ be an indexed set of concrete n -graphs. The graph $G = //\{G_i \mid i \in I\}$ is defined as $K[G_i/e_i; i \in I]$, where K is as H above with set of edges $\{e_i \mid i \in I\}$, all of them with $(1, \dots, n)$ as sequence of vertices. Intuitively, G is the result of gluing the graphs G_i by their sources, in their order.

If I is a countable set, and $G_i = G$ for all $i \in I$, then the graph $//\{G_i \mid i \in I\}$ is denoted by $//^\infty G$.

If A is a set of n -graphs, we denote by $//A$ the graph $//\{G_i \mid i \in I\}$, where $\{G_i \mid i \in I\}$ is an enumeration of A . (We mean by this that $A = \{G_i \mid i \in I\}$, and that every graph of A is equal to G_i for one and only one $i \in I$). Any two enumerations of A define isomorphic graphs.

These definitions apply to indexed families of abstract graphs as well.

1.9. Definition (*Source-preserving compositions of graphs*). We say that a graph is *source-separated* if any two of its sources are distinct.

A graph operation defined by a tuple (H, e_1, \dots, e_k) is *source-preserving* if H is source-separated, if it has no internal vertex and no other edges than e_1, \dots, e_k , and if each of these edges has pairwise distinct vertices ordered in the way these are in the sequence of sources of H . If f is a source-preserving graph operation, and $G = f(G_1, \dots, G_k)$ we also say that G is a *source preserving composition* of G_1, \dots, G_k .

Let G_1, \dots, G_k be source-separated concrete graphs, and let G be a concrete graph, that is a source-preserving composition of them as above. Because of the ordering of the vertices of the edges of H , each of the graphs G_i is a subgraph of G .

The sources of any of these graphs G_i remain sources of G (after possible fusion with other sources of G_1, \dots, G_k), and each internal vertex of G corresponds to one and only one internal vertex of one of the graphs G_1, \dots, G_k . Every edge of G also corresponds to one and only one edge of one of the graphs G_1, \dots, G_k . Furthermore two internal items of G are internally linked (in G) iff they belong to the same graph G_i , and are internally linked in this graph. Hence, roughly speaking, $\mathbf{CICC}(G) = \mathbf{CICC}(G_1) \cup \dots \cup \mathbf{CICC}(G_k)$. (One has actually a canonical isomorphism rather than an equality.)

Note that the parallel composition of two graphs is a source preserving graph operation.

1.10. Lemma. (1) *Let G be a concrete source-separated graph. There exists a partition $\{A_1, \dots, A_k\}$ of $\mathbf{CICC}(G)$ such that G is isomorphic to a source-preserving composition of the graphs $//A_1, //A_2, \dots, //A_k$.*

(2) If, furthermore, $\text{ICC}(G)$ is finite, then there exist $G_1, \dots, G_n, G_{n+1}, \dots, G_{n+p}$ in $\text{CICC}(G)$ such that G is isomorphic to a source preserving composition of the graphs $G_1, \dots, G_n, \text{\textit{\textcircled{}}}^\infty G_{n+1}, \dots, \text{\textit{\textcircled{}}}^\infty G_{n+p}$.

Proof. (1) Let us form a partition of $\text{CICC}(G)$ by putting in the same component the cicc's that have the same set of sources. This defines a partition with finitely many components A_1, \dots, A_k . Note that since the elements of $\text{CICC}(G)$ are subgraphs of G , any two elements of some A_i have the same sequence of sources. (All the cicc's of type 0 (they are also ccc's) are in a same component of the partition, say A_k .)

It is then clear that G is a source-preserving composition of $\text{\textit{\textcircled{}}}A_1, \dots, \text{\textit{\textcircled{}}}A_k$. We omit the details.

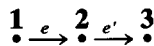
(2) If $\text{ICC}(G)$ is finite, then each component A_i can be partitioned into $\{A_{i,1}, \dots, A_{i,n_i}\}$, such that the graphs of $A_{i,j}$ are all isomorphic to one another.

Let then $\{G_1, \dots, G_n\}$ be an enumeration of $\cup \{A_{i,j} \mid i \in [k], j \in [n_i], A_{i,j} \text{ is finite}\}$, and $\{G_{n+1}, \dots, G_{n+p}\}$ be an enumeration of the set $\{B_{i,j} \mid i \in k, j \in [n_i], A_{i,j} \text{ is infinite}\}$, where each $B_{i,j}$ is an element of $A_{i,j}$. The result follows then from (1), and from the facts that $\text{\textit{\textcircled{}}}A_i$ is isomorphic to $(\text{\textit{\textcircled{}}}A_{i,1})\text{\textit{\textcircled{}}}(\text{\textit{\textcircled{}}}A_{i,2})\text{\textit{\textcircled{}}} \dots \text{\textit{\textcircled{}}}(\text{\textit{\textcircled{}}}A_{i,n})$, and that $\text{\textit{\textcircled{}}}A_{i,j}$ is isomorphic to $\text{\textit{\textcircled{}}}^\infty B_{i,j}$, where $A_{i,j}$ is infinite. \square

Let us finally observe that a nonempty graph is internally connected iff it has no isolated source, and is not of the form $G \text{\textit{\textcircled{}}} G'$, for any two nondegenerated graphs G and G' .

1.11. Example. A graph G is shown in Fig. 1. (The vertices marked **1**, **2**, **3** are its first, second and third sources respectively.) It has countably many ‘‘parallel edges’’ labelled by a . It is connected but not internally connected. The set $\text{CICC}(G)$ consists of countably many copies of the graph a (i.e., $\mathbf{1} \xrightarrow{a} \mathbf{2}$) and of the graph H also shown in Fig. 1.

The graph G can be written $f(\text{\textit{\textcircled{}}}^\infty a, H)$, where f is the graph operation associated with the tuple (K, e, e') , where K is the graph:



2. Systems of equations and equational graphs

2.1. Definitions. Let A be a finite ranked alphabet called the set of terminal labels. A system of graph equations over A , or simply a system, is a sequence of equations of the form $S = \langle u_1 = H_1, \dots, u_n = H_n \rangle$, where $U = \{u_1, \dots, u_n\}$ is a ranked alphabet called the set of unknowns of S , and $H_i \in \text{FCG}(A \cup U)_{\tau(u_i)}$ for all $i = 1, \dots, n$. An edge of $H_i, i \in [n]$ is terminal if its label is in A , and nonterminal otherwise, i.e., if its label is in U .

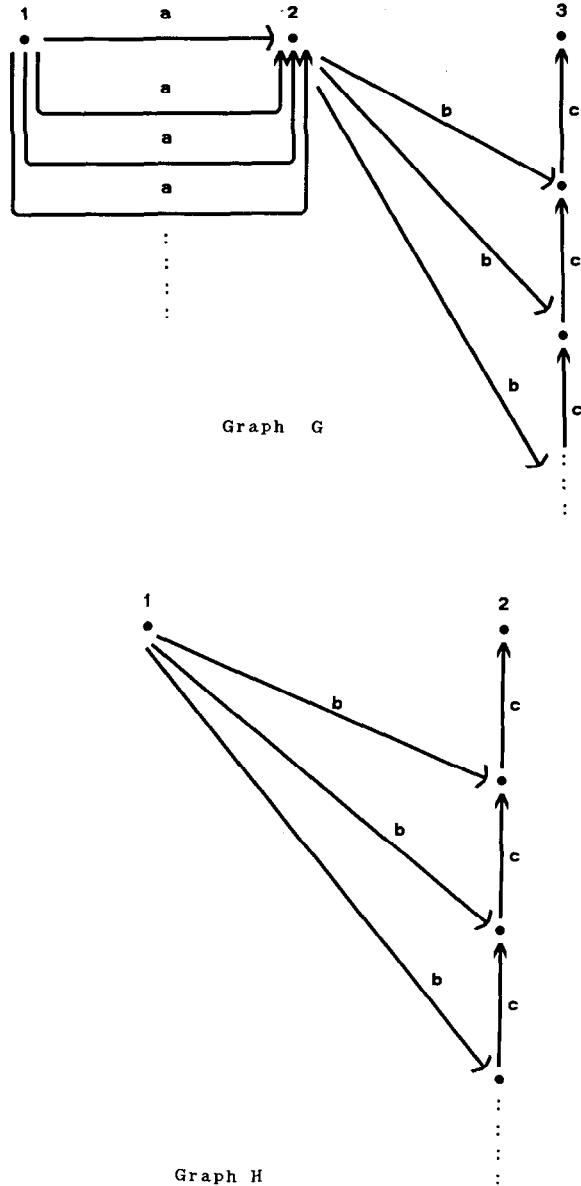


Fig. 1.

A solution of S is an n -tuple (G_1, \dots, G_n) of graphs in $\mathbf{G}(A)$, such that $\tau(G_i) = \tau(u_i)$, and $G_i = H_i[G_1/u_1, \dots, G_n/u_n]$ for all $i = 1, \dots, n$.

A few examples are given below. They show that a system S has in general several solutions. The set of solutions of S contains an *initial* solution. The term initial is used w.r.t. an appropriate category of solutions. The introduction of category theory makes possible to apply the notion of *least fixed point of a functor*

[1]. This notion and its application to the construction of solutions of systems of graph equations are briefly surveyed in the appendix. A complete treatment can be found in Bauderon [2, 3]. An alternative construction of the initial solution is given in Courcelle [11], in terms of *infinite graph expressions*. The solution constructed in this way, independently of category-theoretical notions, is called the *canonical solution* of S . We shall use here a variant of the construction of [11] that is more appropriate to the purposes of this paper.

A graph is *equational* if it is *defined* by a system, i.e., if it is a component of the canonical (equivalently initial) solution of this system. A finite graph is trivially equational.

2.2. Example. A system S is shown in Fig. 2. The terminal alphabet is $A = \{a, b, c\}$, its unknowns are u_1, u_2 , and u_3 . All symbols are of type 2. By S_i , $i = 1, 2, 3$, we denote the system reduced to the equation with left-hand side u_i .

The graph G_1 of Fig. 3 is a solution of S_1 . For every solution G of S_1 , for every graph G' of type 0, the graph $G \oplus G'$ is also a solution of S_1 . But G_1 is *the* (canonical) solution of S_1 . The graphs G_2, G'_2 and $G''_2 := \theta_{1,2}(G_2)$ are three

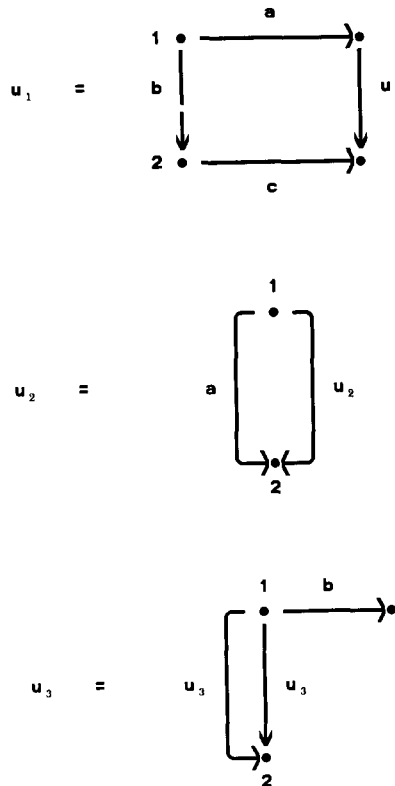


Fig. 2.

solutions of S_2 . (The graph G'_2 is not shown in Fig. 3.) The graphs G_2 and G_3 are the solutions of S_2 and S_3 respectively. Note that G_3 is not connected, although the right-hand side of S_3 is connected. Observe that $G_2 = //^\infty a$ and that $G_3 = //^\infty H$ (where H is the graph: $1 \bullet \xrightarrow{b} \bullet \bullet 2$).

2.3. Definition (Syntactic trees). Let $S = \langle u_i = H_i; i \in [n] \rangle$ be a system over A , with $n_i = \tau(u_i) = \tau(H_i)$ for all $i = 1, \dots, n$.

We shall describe its (canonical) solution (G_1, \dots, G_n) by means of an n -tuple (T_1, \dots, T_n) of infinite trees, where T_i will be called the *syntactic tree* of G_i .

For each i in $[n]$, we let $(e_{i,1}, \dots, e_{i,m_i})$ be a fixed enumeration of the set of

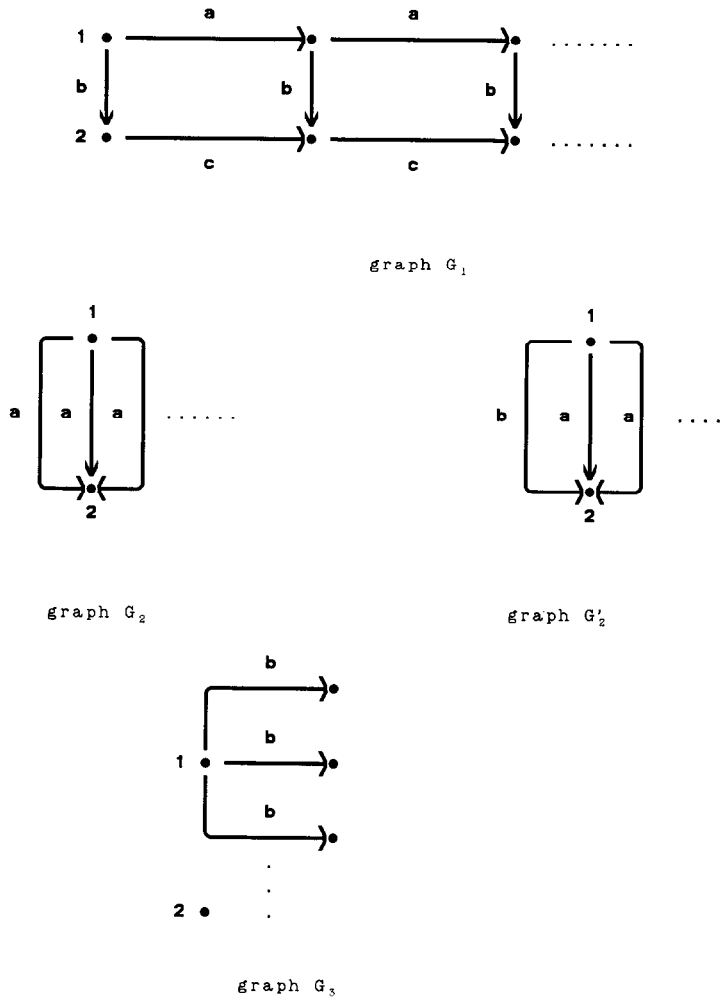


Fig. 3.

nonterminal edges of H_i . We let $(w_{i,1}, \dots, w_{i,m_i})$ be the corresponding sequence of labels (hence $w_{i,j} \in U$). We let $r = \max\{m_i \mid i \in [n]\}$.

An (n, r) -tree is a 4-tuple $T = \langle N, \mathbf{rt}, \mathbf{lab}, (\mathbf{suc}_j)_{j \in [r]} \rangle$ where:

- N is a finite or countable set, the set of *nodes*,
- \mathbf{rt} is a node, the *root* of the tree,
- $\mathbf{lab}: N \rightarrow [n]$ is a total mapping defining the *label* of a node; one lets $N_i = \mathbf{lab}^{-1}(i)$, for $i \in [n]$,
- $\mathbf{suc}_j \subseteq N \times N$ is a binary relation: $(\mu, \mu') \in \mathbf{suc}_j$ means that the node μ' is a j -*successor* of the node μ ; the integer j plays the role of a label attached to the edge from μ to μ' ; hence one assumes that $\mathbf{suc}_j \cap \mathbf{suc}_{j'} = \emptyset$ for $j' \neq j$; one lets $\mathbf{suc} = \bigcup \{\mathbf{suc}_j \mid j \in [r]\}$, and one assumes that $\langle N, \mathbf{rt}, \mathbf{suc} \rangle$ is an unordered tree in the classical sense with $\{\mu' \mid (\mu, \mu') \in \mathbf{suc}\}$ as set of successors (i.e. of sons) of node μ .

We say that $\langle N, \mathbf{rt}, \mathbf{suc} \rangle$ is a tree to mean that for every node μ there is one and only one path from \mathbf{rt} to μ . A path is here a sequence of nodes $(\mu_1, \mu_2, \dots, \mu_k)$ such that $(\mu_i, \mu_{i+1}) \in \mathbf{suc}$ for all $i = 1, \dots, k-1$.

We say that T is a *syntactic tree of S* , if T is an (n, r) -tree, and if the following condition holds:

- (ST): for every node $\mu \in N_i$, for every $j \in [r]$, if $w_{i,j} = u_k$ then μ has one and only one j -successor μ' , and μ' is labeled by k , and if $j > m_i$ then μ has no j -successor.

It is clear that, for every $i \in [n]$, there is one and only one syntactic tree T , the root of which is labeled by i . This tree is denoted by T_i , and is called the *syntactic tree of G_i* .

We denote by T_i^j the initial part of T_i consisting of its first j levels, with T_i^1 reduced to the root, labeled by i .

A tree is a graph, and its nodes could be called vertices. But we keep the term “node” for trees. In complicated proofs involving simultaneously a graph and a tree describing it, having two distinct words makes statements more clear.

A syntactic tree T is *regular* in the following sense. For any two nodes μ and μ' with the same label, the two subtrees T_μ and $T_{\mu'}$, issued from μ and μ' , are isomorphic. (An isomorphism must preserve the labels of nodes and of edges.)

2.4. Definition (*The value of a syntactic tree*). We now show that G_i can be considered as the result of gluing infinitely many copies of finitely many graphs, in a way defined by T_i . We assume that H_1, \dots, H_n are pairwise disjoint concrete graphs.

For every $i \in [n]$, we let $\mathbf{del}(H_i)$ be the graph

$$H_i \upharpoonright (\mathbf{V}_{H_i} \cup \mathbf{E}_{H_i} - \{e_{i,1}, \dots, e_{i,m_i}\}).$$

Hence $\mathbf{del}(H_i)$ is a subgraph of H_i (obtained from H_i by the deletion of some edges, whence the notation).

Let T be a syntactic tree $\langle N, \mathbf{rt}, \mathbf{lab}, (\mathbf{suc}_j)_{j \in [m_1]} \rangle$, and let $i_0 = \mathbf{lab}(\mathbf{rt})$.

Hence $T = T_{i_0}$. Let $k_0 = \tau(u_{i_0}) = \tau(G_{i_0})$. We now construct a graph G that we shall denote by $\mathbf{eval}(T)$.

For each $\mu \in N$, and $H \in \mathbf{CG}(A)$, we denote by $\mu \cdot H$ the concrete graph isomorphic to H , such that $\mathbf{V}_{\mu \cdot H} = \{\mu\} \times \mathbf{V}_H$, $\mathbf{E}_{\mu \cdot H} = \{\mu\} \times \mathbf{E}_H$, (μ, v) corresponds to v for $v \in \mathbf{V}_H$, and (μ, e) corresponds to e for $e \in \mathbf{E}_H$, in the isomorphism $:\mu \cdot H \rightarrow H$.

The graphs $\mu \cdot \mathbf{del}(H_i)^0$, for all $\mu \in N_i$, $i \in [n]$, are pairwise disjoint concrete 0-graphs. Let K be their union, equipped with the sequence of sources of the graph $\mathbf{rt} \cdot \mathbf{del}(H_{i_0})$. We call it the *intermediate graph associated with T* .

We let \simeq be the equivalence relation on \mathbf{V}_K , generated by the set of pairs of vertices of the form:

$$((\mu, \mathbf{vert}_{H_i}(e_{i,j}, k)), (\mu', \mathbf{src}_{H_i}(k)))$$

where $(\mu, \mu') \in \mathbf{suc}_j$, $j \in [r]$, $i = \mathbf{lab}(\mu)$, $i' = \mathbf{lab}(\mu')$, $k \in [\tau(e_{i,j})]$.

The graph associated with T is defined as K/\simeq , and is denoted by $\mathbf{eval}(T)$, or $\mathbf{eval}_S(T)$ if, for some reason, we wish to specify the relevant system. This graph is infinite in general.

This construction also applies to the partial syntactic trees T_i^j , and yields finite graphs, denoted by $\mathbf{eval}(T_i^j)$.

2.5. Proposition. *The n -tuple $(\mathbf{eval}_S(T_1), \dots, \mathbf{eval}_S(T_n))$ is the canonical solution of S .*

The proof is in the Appendix.

2.6. Example. The syntactic trees of the system S of Example 2.2 are shown in Fig. 4. Fig. 5 shows simultaneously the syntactic tree T of G_1 , the graph $\mathbf{del}(H_1)$, the intermediate graph K and its quotient $G_1 = K/\simeq = \mathbf{eval}(T)$. The dotted lines on the drawing of K are not edges. They show pairs of vertices generating \simeq . \square

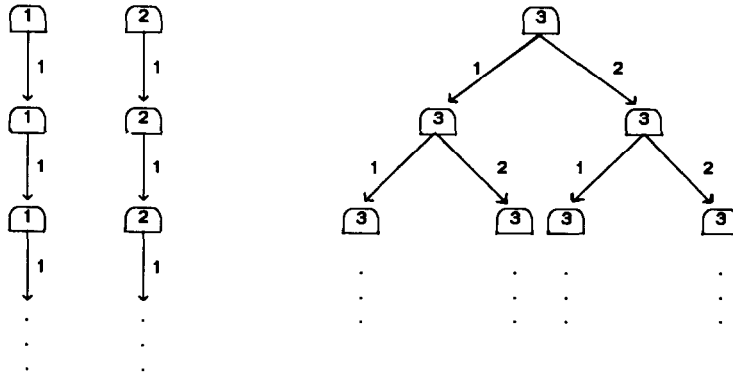


Fig. 4.

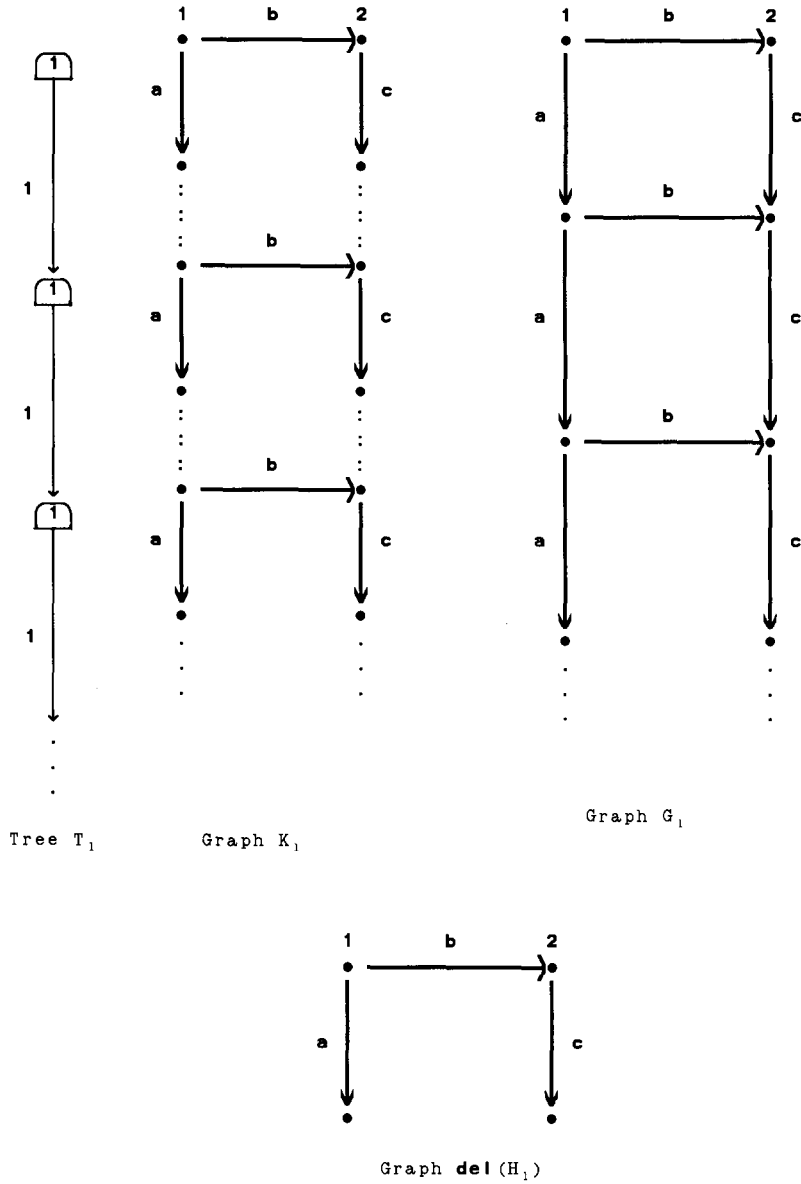


Fig. 5.

2.7. Definition. A system $S = \langle u_i = H_i; i \in [n] \rangle$ is *separated*, if the graphs H_i are source-separated and U -separated, for all i . (We say that a graph is U -separated if the sequence of vertices of an edge having its label in U has no repetition. It is source-separated if its sources are pairwise distinct; see Definition (1.9).)

2.8. Lemma. *The graphs defined by a separated system are source-separated.*

Proof. Let S be separated, let $G_i = \mathbf{eval}(T_i)$ where T_i is the syntactic tree of G_i . By Definition 2.4, $G_i = K_i/\approx$ where K_i is the intermediate graph associated with T_i . It is easy to check that no two sources of K_i are equivalent by \approx . We omit the details. \square

2.9. Proposition. *Let S be a separated system, let G_1, \dots, G_n be its solution. Let us assume that none of the graphs G_i , $i = 1, \dots, n$, is degenerated. Any solution of S consisting of source-separated and internally connected graphs is isomorphic to the canonical one.*

The proof is given in the Appendix.

Remarks. (1) In this proposition, the condition that the graphs of the solution of S have internal items eliminates systems of the forms $\langle u = v, v = u \rangle$ or $\langle w = \sigma_\alpha(w) \rangle$. Note that Proposition 2.5 is valid for them.

(2) The system S_1 of Example 2.2 is separated. The graph G_1 is its unique source-separated and internally connected solution. The system S_2 has no internally connected solution. Two connected source-separated solutions of S_2 , namely G_2 and G'_2 , are shown in Fig. 3. (The graph G'_2 has infinitely many a 's.) \square

For every graph G , we denote by $\mathbf{sep}(G)$ the unique source-separated graph H that is like G except that the ordering of its sources is the ordering of their first occurrences in \mathbf{src}_G , i.e., such that:

- $\mathbf{src}_G = (v_1, v_2, \dots, v_n)$ for some v_1, \dots, v_n ,
- $\mathbf{src}_H = (v_{i_1}, v_{i_2}, \dots, v_{i_m})$,
- i_{k+1} is the least integer j such that $v_j \notin \{v_{i_1}, \dots, v_{i_k}\}$, $k \geq 0$. (Hence $i_1 = 1$.)

There is a unique mapping α , and a (nonunique) mapping β such that $G = \sigma_\alpha(\mathbf{sep}(G))$ and $\mathbf{sep}(G) = \sigma_\beta(G)$.

The next proposition shows that every equational graph is of the form $\sigma_\alpha(G)$, where G is defined by a separated system.

2.10. Proposition. *For every system S , one can construct a separated system defining the graphs $\mathbf{sep}(G_1), \dots, \mathbf{sep}(G_n)$ where (G_1, \dots, G_n) is the solution of S .*

The proof is given in the Appendix.

2.11. Example. Let S be the system shown in Fig. 6, with u of type 2. Let G be its solution. The system S' , also shown in Fig. 6, is separated. Its unknown v is of type 1 and its solution is the pair of graphs $(G, \sigma_1(\theta_{1,2}(G)))$, where u defines G , and v defines $\sigma_1(\theta_{1,2}(G))$.

2.12. Definition (Extended systems). Let A and $U = \{u_1, \dots, u_n\}$ be as in Definition 2.1. Let U^∞ be the new ranked alphabet $\{u_i^\infty \mid i \in [n]\}$, disjoint from $A \cup U$, with rank mapping τ such that $\tau(u_i^\infty) = \tau(u_i)$.

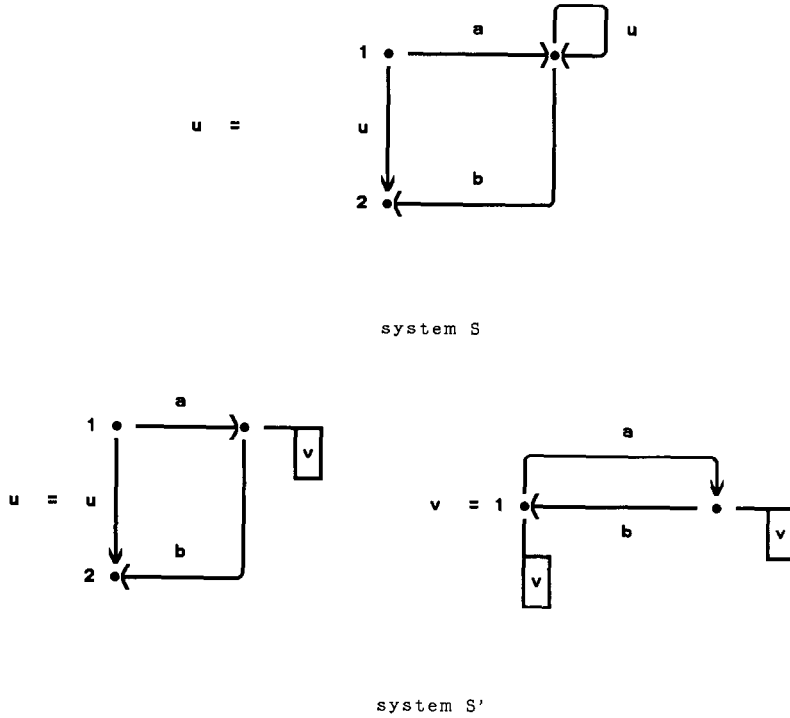


Fig. 6.

An *extended system* with set of unknowns U , is a sequence of equations S of the form $\langle u_1 = H_1, \dots, u_n = H_n \rangle$ where $H_i \in \mathbf{FCG}(A \cup U \cup U^\infty)_{\tau(u_i)}$ for all $i = 1, \dots, n$.

A *solution* of S is an n -tuple $(G_1, \dots, G_n) \in \mathbf{G}(A)^n$ such that:

$$\tau(G_i) = \tau(u_i),$$

and:

$$G_i = H_i[G_1/u_1, \dots, G_n/u_n, (\llcorner^\infty G_1)/u_1^\infty, \dots, (\llcorner^\infty G_n)/u_n^\infty]$$

for all $i = 1, \dots, n$.

Let S^∞ be the (ordinary) system consisting of the equations of S , augmented with the equations:

$$u_i^\infty = u_i \llcorner u_i^\infty$$

for $i = 1, \dots, n$. Its set of unknowns is $U \cup U^\infty$. Remark that the right-hand sides of these equations are not graphs, but graph expressions.

It is clear that the canonical solution of S^∞ is a $2n$ -tuple of graphs of the form $(G_1, \dots, G_n, G'_1, \dots, G'_n)$, where $G'_i = \llcorner^\infty G_i$ for all $i = 1, \dots, n$, and that (G_1, \dots, G_n) is a solution of S . This n -tuple (G_1, \dots, G_n) is called *the solution of S* , and one says that the graphs G_1, \dots, G_n are *defined* by S . Hence, extended systems define equational graphs. (Since an equation of the form $u = u \llcorner G$ has

many solutions, a system of the form S^∞ has solutions that are not of the form $(G_1, G_2, \dots, G_n, //^\infty G_1, \dots, //^\infty G_n)$ for any G_1, \dots, G_n .)

An extended system S as above is *separated*, if the graphs H_i are source-separated and $(U \cup U^\infty)$ -separated. It follows that S is separated iff S^∞ is separated. Hence, by Lemma 2.8, the graphs defined by a separated extended system are source-separated.

We now define syntactic trees for the graphs defined by extended systems. In Definition 2.3, it suffices to do the following modifications:

- $(e_{i,1}, \dots, e_{i,m_i})$ is an enumeration of the set of nonterminal edges of H_i , i.e., of the set of those having a label in $U \cup U^\infty$,
- $(w_{i,1}, \dots, w_{i,m_i})$ is the corresponding sequence of labels, (hence $w_{i,j} \in U \cup U^\infty$),
- in Condition (ST), one adds the following clause: if $\mu \in N_i$, $j \in [m_i]$, and $w_{i,j} = u_k^\infty$, then μ has ω j -successors, and all of them have the label k .

The definition of $\text{eval}_S(T)$ is as in Definition 2.4.

2.13. Proposition. *Let S be an extended system with n unknowns, and let (T_1, \dots, T_n) be the associated sequence of syntactic trees. The n -tuple $(\text{eval}(T_1), \dots, \text{eval}(T_n))$ is the solution of S .*

Proof. Let $(T'_1, \dots, T'_n, T'_{n+1}, \dots, T'_{2n})$ be the sequence of syntactic trees of the (ordinary) system S^∞ , with set of unknowns $\{u_1, \dots, u_n, u_1^\infty, \dots, u_n^\infty\}$.

It is not hard to prove that $\text{eval}_{S^\infty}(T'_i) = \text{eval}_S(T_i)$, and that $\text{eval}_{S^\infty}(T'_{n+i}) = //^\infty(\text{eval}_S(T_i))$ for all $i = 1, \dots, n$, by going back to the definitions of S^∞ . \square

2.14. Proposition. *Proposition 2.9 holds for separated extended systems.*

Proof. See the Appendix. \square

3. Monadic second-order logic and definable graphs

We review or introduce the following notions: relational structures, monadic second-order logic, definability and relative definability of graphs.

3.1. Definitions. Let R be a finite ranked set of symbols such that each element r in R has a rank $\rho(r)$ in \mathbb{N}_+ . A symbol r in R is considered as a $\rho(r)$ -ary relation symbol.

An R -(*relational*) *structure* is a tuple $S = \langle \mathbf{D}_S, (r_S)_{r \in R} \rangle$ where \mathbf{D}_S is a possibly empty set, called the domain of S , and r_S is a subset of $\mathbf{D}_S^{\rho(r)}$ for each r in R . We denote by $\mathcal{S}(R)$ the class of R -structures.

We denote by $\mathcal{L}(R, W)$ the set of formulas of *monadic second-order logic* written with the symbols of R , and with free variables in W , where W is a set of

variables $X, Y, X_1, X_2, Z, Z', \dots$. These variables will denote subsets of \mathbf{D}_S , where S belongs to $\mathcal{S}(R)$.

The atomic formulas are:

$$\begin{aligned} X \subseteq Y, \\ r(X_1, \dots, X_n), \quad \text{where } n = \rho(r), \\ \mathbf{fin}(X), \end{aligned}$$

If X, Y, X_1, \dots, X_n denote subsets $\bar{X}, \bar{Y}, \bar{X}_1, \dots, \bar{X}_n$ of \mathbf{D}_S , $S \in \mathcal{S}(R)$, then these formulas are true iff, respectively:

$$\begin{aligned} \bar{X} \subseteq \bar{Y}, \\ r_S(x_1, \dots, x_n) \text{ holds, for some } x_i \text{ in } \bar{X}_i, i = 1, \dots, n, \\ \bar{X} \text{ is finite.} \end{aligned}$$

The formulas of $\mathcal{L}(R, W)$ are formed with the Boolean connectives \neg and \vee , and existential quantifications.

Let S be an R -structure, let $\varphi \in \mathcal{L}(R, W)$, and γ be a W -assignment in S , i.e., $\gamma(X)$ is a subset of \mathbf{D}_S for every variable X in W (we write this $\gamma: W \rightarrow S$ to be short). We write $(S, \gamma) \models \varphi$ iff φ holds in S for γ . We write $S \models \varphi$ in the case where φ has no free variable.

A set of R -structures L is *definable* if it is the set of R -structures where some (closed) formula φ in $\mathcal{L}(R)$ holds.

In order to make formulas more readable, and also to define conveniently some sublanguages of $\mathcal{L}(R, W)$, we extend the syntax with the connectives \wedge, \Rightarrow and $\forall X$. We shall also use the following new atomic formulas:

$$\begin{aligned} X = Y \quad \text{equivalent to} \quad X \subseteq Y \wedge Y \subseteq X, \\ X = \emptyset \quad \text{equivalent to} \quad \forall Y [Y \subseteq X \Rightarrow X \subseteq Y], \\ \mathbf{sgl}(X) \quad \text{equivalent to} \quad \forall Y [Y \subseteq X \Rightarrow Y = \emptyset \vee Y = X] \end{aligned}$$

(to mean that X is singleton).

We shall use the quantifications on finite sets:

$$\begin{aligned} \exists_f X \varphi \quad \text{equivalent to} \quad \exists X [\mathbf{fin}(X) \wedge \varphi], \\ \forall_f X \varphi \quad \text{equivalent to} \quad \forall X [\mathbf{fin}(X) \Rightarrow \varphi]. \end{aligned}$$

We shall also use lowercase variables x, y, x_1, \dots, x_n to denote singletons, i.e., elements of \mathbf{D}_S , $S \in \mathcal{S}(R)$.

This means that:

$$\begin{aligned} \exists x \varphi \quad \text{is equivalent to} \quad \exists x [\mathbf{sgl}(x) \wedge \varphi], \\ \forall x \varphi \quad \text{is equivalent to} \quad \forall x [\mathbf{sgl}(x) \Rightarrow \varphi], \\ x \in Y \quad \text{is equivalent to} \quad x \subseteq Y. \end{aligned}$$

For every assignment $\gamma: W \rightarrow S$, we shall assume that $\gamma(x)$ is singleton for every lowercase variable x in W . We shall write $\gamma(x) = d$ instead of $\gamma(x) = \{d\}$.

3.2. Definition (Relative definability of structures). Let R and R' be two ranked sets of relation symbols. Let W be a finite set of uppercase variables, called here the set of *parameters*. (It is not a loss of generality to assume that the parameters are set variables; this is just convenient for some proofs.)

An (R, R') -*definition scheme* is a tuple of formulas of the form $\Delta = (\delta, \psi_1, \dots, \psi_k, (\theta_{r,j})_{r \in R, j \in [k]^{\rho(r)}}$ where:

$$\delta \in \mathcal{L}(R', W) \quad (\text{we call } \delta \text{ the domain formula of } \Delta),$$

$$\psi_i \in \mathcal{L}(R', \{x_1\} \cup W), \quad i = 1, \dots, k,$$

$$\theta_{r,j} \in \mathcal{L}(R', \{x_1, \dots, x_{\rho(r)}\} \cup W), \quad r \in R, j \in [k]^{\rho(r)}.$$

Let $S' \in \mathcal{S}(R')$, let γ be a W -assignment in S' . A structure S with domain $\subseteq \mathbf{D}_{S'} \times [k]$ is *defined by* Δ in (S', γ) , (this is denoted by $S = \mathbf{def}_\Delta(S', \gamma)$) if:

$$(S', \gamma) \models \delta, \quad \mathbf{D}_S := \{(d, i) \mid d \in \mathbf{D}_{S'}, i \in [k], (S', \gamma, d) \models \psi_i\}$$

(this set may be empty, and S is still well-defined),

$$r_s = \{((d_1, i_1), \dots, (d_s, i_s)) \mid (S', \gamma, d_1, \dots, d_s) \models \theta_{r,j}\}$$

$$\text{where } j = (i_1, \dots, i_s) \text{ and } s = \rho(r).$$

(By $(S', \gamma, d_1, \dots, d_s) \models \theta_{r,j}$, we mean $(S', \gamma') \models \theta_{r,j}$ where γ' is the assignment extending γ , such that $\gamma'(x_i) = d_i$ for all $i = 1, \dots, s$; analogous notations will be used in the sequel).

In such a case, we say also that S is defined in S' by Δ in terms of γ .

In the special case where $k = 1$, we can replace $\mathbf{D}_S \times \{1\}$ by \mathbf{D}_S , so that $\mathbf{D}_S \subseteq \mathbf{D}_{S'}$, and the tuple Δ can be written more simply $(\varphi, \psi, (\theta_r)_{r \in R})$.

We denote by $\mathbf{def}_\Delta(S')$ the set of structures of the form $\mathbf{def}_\Delta(S', \gamma)$ for some assignment γ . If $W = \emptyset$, then $\mathbf{def}_\Delta(S')$ is either empty or singleton. We write $S = \mathbf{def}_\Delta(S')$ iff it is the singleton reduced to S . It may also happen that W is not empty, but that all the structures in $\mathbf{def}_\Delta(S')$ are isomorphic. Since we are actually interested by structures up to isomorphism, we write in this case $S = \mathbf{def}_\Delta(S')$ where S is any of these isomorphic structures.

We say that S is *definable in* S' if $S = \mathbf{def}_\Delta(S')$ for some definition scheme Δ . This notion is trivial for finite structures (because any two nonempty finite structures are definable in each other), but it is not for infinite ones.

The following proposition says that if $S = \mathbf{def}_\Delta(S', \gamma)$ then the monadic second-order properties of S can be expressed as monadic second-order properties of (S', γ) .

Let $\Delta = (\delta, \psi_1, \dots, \psi_k, (\theta_{r,j})_{r \in R, j \in [k]^{\rho(r)}}$ be written with a set of parameters W . Let V be a set of uppercase variables disjoint from W .

For every variable X in V , for every $i = 1, \dots, k$, we let X_i be a new variable. We let $\vec{V} = \{X_i \mid X \in V, i = 1, \dots, k\}$. For every $\vec{\gamma}: \vec{V} \rightarrow \mathcal{P}(D)$, we let $\gamma: V \rightarrow$

$\mathcal{P}(D \times [k])$ be defined by

$$\gamma(X) = \bar{\gamma}(X_1) \times \{1\} \cup \cdots \cup \bar{\gamma}(X_k) \times \{k\}.$$

With these notations we have:

3.3. Proposition. *For every formula β in $\mathcal{L}(R, V)$, one can construct a formula $\bar{\beta}$ in $\mathcal{L}(R', \bar{V} \cup W)$ such that, for every S' in $\mathcal{S}(R')$, for every $\mu: W \rightarrow S'$, for every $\bar{\gamma}: \bar{V} \rightarrow S'$, we have:*

$$\begin{aligned} \text{def}_\Delta(S', \mu) \text{ is defined} & \quad (\text{let us denote it by } S), \\ \gamma \text{ is a } V\text{-assignment in } S, & \quad \text{and } (S, \gamma) \models \beta \\ \text{iff } (S', \bar{\gamma} \cup \mu) \models \bar{\beta}. & \end{aligned}$$

Proof. We take $\bar{\beta}$ equal to

$$\bar{\beta} \wedge \delta \wedge \bigwedge_i \{ \forall x (x \in X_i \Rightarrow \psi_i(x)) \}$$

where $\bar{\beta}$ is constructed by induction on the structure of β as follows:

- If β is $X \subseteq X'$, then $\bar{\beta}$ is $X_1 \subseteq X'_1 \wedge \cdots \wedge X_k \subseteq X'_k$.
- If β is $r(X^1, \dots, X^n)$, then $\bar{\beta}$ is $\exists y_1, \dots, y_n [\bigwedge_{j \in [k]^n} \{ \theta_{r,j}(y_1, \dots, y_n) \wedge y_1 \in X_{j(1)}^1 \wedge \cdots \wedge y_n \in X_{j(n)}^n \}]$ (where we denote by $j(i)$ the i th element of the sequence j).
- If β is $\text{fin}(X)$, then $\bar{\beta}$ is $\text{fin}(X_1) \wedge \cdots \wedge \text{fin}(X_k)$.
- If β is $\neg\beta_1$, then $\bar{\beta}$ is $\neg\bar{\beta}_1$.
- If β is $\beta_1 \vee \beta_2$, then $\bar{\beta}$ is $\bar{\beta}_1 \vee \bar{\beta}_2$.
- If β is $\exists X \beta_1$, then $\bar{\beta}$ is $\exists X_1, \dots, X_k \bar{\beta}_1$. (We assume that all variables of β are in V .)

The verification that $\bar{\beta}$ satisfies the desired properties is easy by induction on the structure of β . \square

The following proposition is an easy consequence of the previous one.

3.4. Corollary. *Let $S \in \mathcal{S}(R)$, $S' \in \mathcal{S}(R')$, and $S'' \in \mathcal{S}(R'')$. If S is definable in S' , and S' is definable in S'' , then S is definable in S'' .*

The definition scheme defining S in S'' is obtained by the construction of Proposition 3.3 from the definition schemes defining S in S' , and S' in S'' .

3.5. Definition (Relational structures representing graphs). Let A be a finite set of edge labels as in Definition 1.1; let $n \in \mathbb{N}$. Let $\mathbf{R}(A, n)$ be the following set of relation symbols:

$$\begin{aligned} \text{edg}_a & \text{ of rank } \tau(a) + 1, \quad \text{for each } a \text{ in } A, \\ \text{ps}_i & \text{, of rank } 1, \quad \text{for each } i = 1, \dots, n. \end{aligned}$$

For every n -graph G over A , let $|G|$ be the $\mathbf{R}(A, n)$ -structure such that:

$$\begin{aligned} \mathbf{D}_{|G|} &:= \mathbf{V}_G \cup \mathbf{E}_G && \text{(we assume that } \mathbf{V}_G \cap \mathbf{E}_G = \emptyset\text{),} \\ \mathbf{edg}_{a|G|}(x, y_1, \dots, y_k) &= \mathbf{true} && \text{iff } x \in \mathbf{E}_G, \mathbf{lab}_G(x) = a, \text{ and} \\ &&& \mathbf{vert}_G(x) = (y_1, \dots, y_k), \\ \mathbf{ps}_{i|G|}(x) &= \mathbf{true} && \text{iff } x = \mathbf{src}_G(i). \end{aligned}$$

A first-order formula, saying that some atomic formula $\mathbf{edg}_a(x, y_1, \dots, y_k)$ holds for some y_1, \dots, y_k , characterizes the edges among $\mathbf{D}_{|G|}$.

It is clear that the structure $|G|$ represents the graph G , i.e., that for any two graphs G and G' , $|G|$ is isomorphic to $|G'|$ iff G is isomorphic to G' .

A subset L of $\mathbf{G}(A)_n$ is *definable* iff there exists a formula φ in $\mathcal{L}(\mathbf{R}(A, n))$ such that $G \in L$ iff $|G| \models \varphi$, i.e., iff the set of structures representing it is definable. A graph G is *definable* if the set consisting of G is definable, i.e., if this graph is characterized up to isomorphism by the validity in $|G|$ of a closed monadic second-order formula. The notion of a graph being *definable in* another one follows immediately from the corresponding notion concerning structures.

In Courcelle [10–12], a slightly different version of monadic second-order logic is used: the structures representing graphs have two domains, the set of edges, and the set of vertices, and the formulas are written with variables of two possible sorts: the variables of sort “vertex” denote vertices or sets of vertices, and those of sort “edge” denote edges or sets of edges. It is not hard to prove that the two logical languages yield the same definable sets of graphs. The proof is similar to that of Proposition 3.3.

We now recall a basic theorem from Courcelle [11].

3.6. Theorem. *The monadic second-order theory of an equational graph is decidable.*

More precisely, given a system of graph equations and a closed monadic second-order formula of appropriate type, one can decide whether the formula holds in the graph that is the first component of the canonical solution of the system.

Hence, in particular, one can decide whether an equational graph G is connected, or internally connected, or source-separated. One can decide whether G has k , or infinitely many concrete internally connected components, the set of sources of which is a given set of sources of G . This is just to cite a few facts that will be useful in the sequel.

There are two main techniques for proving the decidability of the monadic (second-order) theory of a structure (or of a set of structures). The first one uses automata. It has been introduced by Büchi [5], and culminates with the results of Rabin [18–20]. Some difficult aspects of this approach are better understood in the framework of infinite games, as shown by Gurevich and Harrington [17]. The

second technique uses “generalized sums” that construct structures more or less as we construct graphs by means of graph operations. It has been used extensively by Shelah [22] for several types of ordered sets. Both techniques are surveyed in Gurevich [16].

Our proof of Theorem 3.6 (given in [11]) uses the first of these techniques, by means of a reduction to Rabin’s theorem (i.e., the decidability of the monadic theory of the infinite complete binary tree). Our theorem is also an extension of Rabin’s since the infinite complete binary tree is an equational graph. (But we do not give another proof of Rabin’s theorem.)

We now define a few sublanguages of the language \mathcal{L} of monadic second-order logic.

3.7. Definition (*Weak monadic second-order logic*). We shall denote by $\mathcal{WL}(R, W)$ the subset of $\mathcal{L}(R, W)$ consisting of formulas written with quantifications of the restricted forms $\exists x$, $\forall x$, $\exists_f X$, and $\forall_f X$, i.e., written with quantifications restricted to singletons and to finite sets. In Rabin [19], these formulas are called *weak monadic second-order formulas*. We shall also denote by $\exists\mathcal{WL}(R, W)$ the set of formulas of the form $\exists X_1, \dots, X_n \varphi$, where φ is in $\mathcal{WL}(R, W \cup \{X_1, \dots, X_n\})$. Hence one has a hierarchy of languages:

$$\mathcal{WL} \subseteq \exists\mathcal{WL} \subseteq \mathcal{L},$$

and each of them is strictly more expressive than the preceding one by the results of Rabin [19]. We say that a graph or a set of graphs is \mathcal{WL} - (or $\exists\mathcal{WL}$ -) *definable* if there is a formula in \mathcal{WL} (or in $\exists\mathcal{WL}$ respectively) that defines it in the sense of Definition 3.5.

It is known from Rabin [18] that a subset of $\mathcal{F}(A)$ (we denote in this way the set of infinite complete binary trees, the nodes of which have labels in A) is definable iff it is definable by a finite-state tree automaton (we need not recall its definition here; see Section 6 below, where we shall use these automata). Such a set L is $\exists\mathcal{WL}$ -definable iff it is defined by a “special” automaton in the sense of Rabin [19], i.e., an automaton with Büchi’s accepting condition. It is \mathcal{WL} -definable iff L and $\mathcal{F}(\{A\}) - L$ are both $\exists\mathcal{WL}$ -definable [19]. The emptiness of the set of trees defined by a “special” automaton can be decided in polynomial time (in the size of the automaton) whereas an exponential time is required in general. It follows that the satisfiability in $\mathcal{F}(\{A\})$ of a formula in $\exists\mathcal{WL}$ can be decided more efficiently than that of a general monadic second-order formula. We shall try to characterize graphs and sets of graphs by formulas in \mathcal{WL} or in $\exists\mathcal{WL}$ rather than in \mathcal{L} .

As a useful example, let us consider the class of (n, r) -trees, defined in Definition 2.3. An (n, r) -tree $T = \langle V, \mathbf{rt}, \mathbf{lab}, (\mathbf{suc}_j)_{j \in [r]} \rangle$ is represented by the $\mathbf{K}(n, r)$ -structure, $\langle V, \mathbf{prt}, (\mathbf{plab}_i)_{i \in [n]}, (\mathbf{suc}_j)_{j \in [r]} \rangle$ (also denoted by T), where

$\mathbf{K}(n, r) := \{\mathbf{prt}, \mathbf{plab}_i, \mathbf{suc}_j \mid i = 1, \dots, n, j = 1, \dots, r\}$ and:

$\mathbf{prt}(x)$ holds iff $x = \mathbf{rt}$, and

$\mathbf{plab}_i(x)$ holds iff $\mathbf{lab}(x) = i$.

The set of $\mathbf{K}(n, r)$ -structures S representing (n, r) -trees is \mathcal{WL} -definable. (The key observation is that the existence of a path in S linking x to y is expressible by a weak formula with free variables x and y . The notion of a path is as in Definition 2.3.) Condition (ST) of Definition 2.3 is clearly first-order. It follows that the syntactic trees of a system are \mathcal{WL} -definable. The syntactic trees of an extended system are also \mathcal{WL} -definable because the property:

“there exist infinitely many x such that $\varphi(x)$ ”

can be expressed with quantifications on singletons and finite sets as follows:

$$\exists x \varphi(x) \wedge \forall_f X [\exists x (\neg x \in X) \wedge \varphi(x)].$$

3.8. Proposition. *Let G be a graph of the form $f(G_1, \dots, G_k, //^\infty(G_{k+1}), \dots, //^\infty(G_l))$ where f is a source-preserving graph operation, and the graphs G_1, \dots, G_l are internally connected, and source-separated.*

If the graphs G_1, \dots, G_l are definable, or $\exists \mathcal{WL}$ -definable, or \mathcal{WL} -definable, then the graph G is definable, $\exists \mathcal{WL}$ -definable or \mathcal{WL} -definable respectively.

Proof. We begin with a few preliminary notations and constructions.

Let H be a graph in $\mathbf{G}(A)_m$. For every internal item x of H , we denote by $\mathbf{C}_H(x)$ the set of items of H that are linked to x by an internal path. It follows that $H \upharpoonright \mathbf{C}_H(x)$ is the cicc of H containing x . We also let

$$\mathbf{S}_H(x) = \{i \in [m] \mid \mathbf{src}_H(i) \in \mathbf{C}_H(x)\}.$$

For every subset S of $[m]$, one can construct weak formulas $\lambda(x, y)$ and $\lambda_S(x)$ such that, for every H in $\mathbf{G}(A)_m$, for all internal items x, y of H :

$$H \vDash \lambda(x, y) \text{ iff } y \in \mathbf{C}_H(x), \text{ and}$$

$$H \vDash \lambda_S(x) \text{ iff } S = \mathbf{S}_H(x).$$

Let $S = \{i_1, \dots, i_n\} \subseteq [m]$, with $i_1 < i_2 < \dots < i_n$. Every formula φ in $\mathcal{L}(\mathbf{R}(A, n), \{Y_1, \dots, Y_k\})$ can be converted into a formula $\varphi_S(x, Y_1, \dots, Y_k)$ such that for every graph H in $\mathbf{G}(A)_m$, for every internal item x such that $\mathbf{S}_H(x) = S$, for every $Y_1, \dots, Y_k \subseteq \mathbf{D}_H$:

$$H \vDash \varphi_S(x, Y_1, \dots, Y_k) \text{ iff } H \upharpoonright \mathbf{C}_H(x) \vDash \varphi(Y'_1, \dots, Y'_k)$$

where $Y'_i := Y_i \cap \mathbf{C}_H(x)$ for each $i = 1, \dots, k$. The formula φ_S is obtained from φ by the following transformations.

(i) One replaces the relation symbol \mathbf{ps} , by \mathbf{ps}_i everywhere in φ , for all $j = 1, \dots, n$.

(ii) One replaces atomic formulas like $X \subseteq Y_i$, $Y_i \subseteq X$, $Y_i \subseteq Y_j$ by weak formulas expressing respectively that $X \subseteq Y'_i$, $Y'_i \subseteq X$, $Y'_i \subseteq Y'_j$. (In the first case, the corresponding formula is: $\forall y (y \in X \Rightarrow y \in Y_i \wedge \lambda(x, y))$.)

(iii) One restricts quantifications to subsets of $\mathbf{C}_H(x)$: each subformula of φ of the form $\exists X \theta$ is replaced by $\exists X [\forall y (y \in X \Rightarrow \lambda(x, y)) \wedge \theta]$. A similar transformation is done for quantifications over singletons and finite sets.

It follows that φ_S is a weak formula if φ is a weak formula. If φ belongs to $\exists \mathcal{WL}(\mathbf{R}(A, n))$, then one can manage to construct φ_S in $\exists \mathcal{WL}(\mathbf{R}(A, m))$ as shown in the following typical case:

If φ is $\exists X \exists Y \theta$ where θ is a weak formula, then we let φ_S be the formula:

$$\exists X \exists Y [\forall y (y \in X \Rightarrow \lambda(x, y)) \wedge \forall y (y \in Y \Rightarrow \lambda(x, y)) \wedge \theta_S]$$

where θ_S is the formula in $\mathcal{WL}(\mathbf{R}(A, m), \{x, X, Y\})$ obtained by the initial construction.

Let us now consider a source-separated graph G in $\mathbf{G}(A)_m$ written as $f(G_1, \dots, G_k, (\!/\!\!/^\infty G_{k+1}), \dots, (\!/\!\!/^\infty G_l))$ where f is a source-preserving graph operation, and where G_1, \dots, G_l are internally connected source-separated graphs respectively defined by formulas $\varphi_1, \dots, \varphi_l$. We assume that f is defined by a tuple (K, e_1, \dots, e_l) as in Definition 1.5. Our purpose is to build a formula defining G . We let $S_0 := \{j \in [m] \mid \mathbf{src}_K(j) \text{ is isolated}\}$, and for each $i = 1, \dots, l$, we let $S_i := \{j \in [m] \mid \mathbf{src}_K(j) \in \mathbf{vert}_K(e_i)\}$.

Let H be an arbitrary graph in $\mathbf{G}(A)_m$. It is isomorphic to G iff the following conditions hold:

- (1) For all $i = 1, \dots, m$, the i th source of H is isolated iff $i \in S_0$.
- (2) There exists a set X of pairwise independent internal items satisfying the following conditions (we say that two items are *independent* if they are not internally linked):
 - (2.1) every internal item of H is internally linked to some item x in X ,
 - (2.2) there exist x_1, \dots, x_k in X such that $\mathbf{S}_H(x_i) = S_i$ and $H \upharpoonright \mathbf{C}_H(x_i) = G_i$ for $i = 1, \dots, k$,
 - (2.3) for every y in $X - \{x_1, \dots, x_k\}$ (where x_1, \dots, x_k are chosen satisfying (2.2)), $\mathbf{S}_H(y) = S_i$ and $H \upharpoonright \mathbf{C}_H(y) = G_i$ for some $i = k + 1, \dots, l$,
 - (2.4) for every $i = k + 1, \dots, l$, there are infinitely many elements y of X such that $\mathbf{S}_H(y) = S_i$ and $H \upharpoonright \mathbf{C}_H(y) = G_i$.

These conditions are immediately expressible by a formula in $\mathcal{L}(\mathbf{R}(A, m))$, with the help of the auxiliary formulas $\lambda_{S_i}(x)$ and $\varphi_{iS_i}(x)$. Hence G is definable if G_1, \dots, G_l are so.

Let us now assume that $\varphi_1, \dots, \varphi_l$ are weak formulas. We shall express these conditions by a weak formula. We reformulate condition (2) as follows:

(2') There exist pairwise independent internal items x_1, \dots, x_k as in (2.2), such that:

for every finite set X of pairwise independent items containing $\{x_1, \dots, x_k\}$, for every internal item y , there exist pairwise independent internal items

z_{k+1}, \dots, z_l such that they are independent with the ones in X , $\mathbf{S}_H(z_i) = S_i$ and $H \upharpoonright \mathbf{C}_H(z_i) = G_i$ for all $i = k + 1, \dots, l$, and y is internally linked to an item in $X \cup \{z_{k+1}, \dots, z_l\}$.

From this new formulation, it is clear that condition (2) is expressible by a weak formula. Hence G is \mathcal{WL} -definable.

Let us finally consider the case where each formula φ_i is in $\exists \mathcal{WL}(\mathbf{R}(A, m_i))$, and m_i is the type of G_i . We shall construct φ in $\exists \mathcal{WL}(\mathbf{R}(A, m))$ that defines G .

We can rewrite the formulas $\varphi_1, \dots, \varphi_l$ in such a way that they are of the form $\exists Y_1, \dots, Y_s \psi_i$, $i = 1, \dots, l$, where $\psi_i \in \mathcal{WL}(\mathbf{R}(A, m_i), \{Y_1, \dots, Y_s\})$, and for every graph H in $\mathbf{G}(A)_{m_i}$, for every s -tuple (Y_1, \dots, Y_s) of sets of items of H , if $H \models \psi_i(Y_1, \dots, Y_s)$, then Y_1, \dots, Y_s do not contain any source of H (and, of course, H is equal to G_i).

Condition (2) can be rewritten as follows:

(2'') There exist sets $X, x_1, \dots, x_k, X_{k+1}, \dots, X_l, Y_1, \dots, Y_s$ satisfying the following conditions:

- X is a set of pairwise independent internal items satisfying (2.1),
- x_1, \dots, x_k are singletons, X_{k+1}, \dots, X_l are infinite and X is the disjoint union of $x_1, \dots, x_k, X_{k+1}, \dots$, and X_l ,
- for every y in X , if $y = x_i$, and $i \in [k]$, or if $y \in X_i$, and $i \in \{k + 1, \dots, l\}$, then:

$$\mathbf{S}_H(y) = S_i \quad \text{and} \quad H \upharpoonright \mathbf{C}_H(y) \models \psi_{iS_i}(y, Y_1 \cap \mathbf{C}_H(y), \dots, Y_s \cap \mathbf{C}_H(y)).$$

Condition (2'') can be expressed by a weak formula with free variables $X, x_1, \dots, x_k, X_{k+1}, \dots, X_l, Y_1, \dots, Y_s$. It is quite clear that Condition (2'') implies Condition (2).

Let us conversely assume that (2) holds. Let X, x_1, \dots, x_k satisfy (2.1)–(2.4). By (2.4), one can partition $X - \{x_1, \dots, x_k\}$ in infinite sets X_{k+1}, \dots, X_l such that $\mathbf{S}_H(y) = S_i$ and $H \upharpoonright \mathbf{C}_H(y) = G_i$ for each $i = k + 1, \dots, l$, and each $y \in X_i$.

For each y in X , there exists an s -tuple Y_1^y, \dots, Y_s^y of sets of internal items $\subseteq \mathbf{C}_H(y)$ such that $H \upharpoonright \mathbf{C}_H(y) \models \psi_i(Y_1^y, \dots, Y_s^y)$, and either $y = x_i$, with $1 \leq i \leq k$, or $y \in X_i$, and $k + 1 \leq i \leq l$.

We then let $Y_i = \bigcup \{Y_i^y \mid y \in X\}$.

Since the sets Y_i are sets of internal items of H , $Y_j^y = Y_j \cap \mathbf{C}_H(y)$ for each j . It follows that:

$$H \upharpoonright \mathbf{C}_H(y) \models \psi_{iS_i}(y, Y_1 \cap \mathbf{C}_H(y), \dots, Y_s \cap \mathbf{C}_H(y)).$$

Hence Condition (2'') holds. This concludes the proof. \square

3.9. Remark. From the result of Proposition 3.4, stating that the relative definability of graphs is a transitive relation, one might think that if a graph is definable in a definable graph, then it is definable. This is not true as shown by the following example.

It is not hard to prove that the infinite graph H of type 1 shown in Fig. 7 is

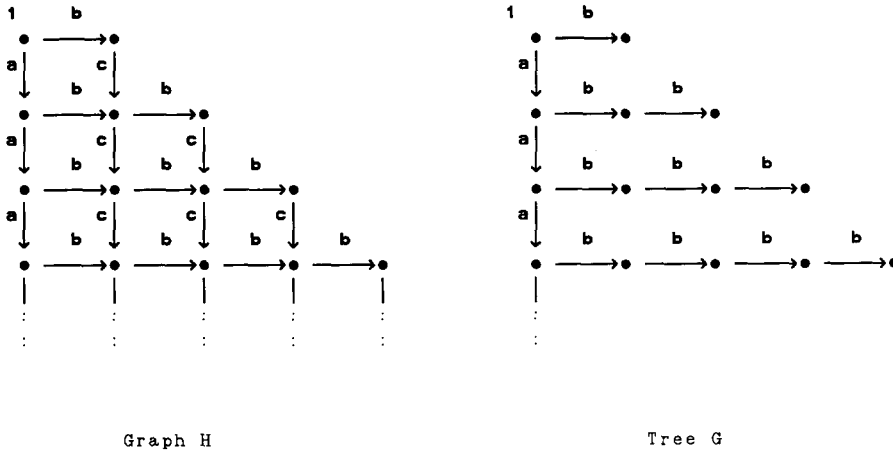


Fig. 7.

definable. The graph $G := H[2/c]$ (obtained from H by deleting all the edges labelled by c , and keeping their vertices) is definable in H . This graph is a $(1, 2)$ -tree. (Its vertices all have the same label, i.e., have no label, and its edges are labelled by a or b , instead of 1 or 2.) This tree is not regular, because it has infinitely many distinct subtrees. Hence it is not definable, because every definable tree is regular. (This follows from the theorem of Rabin [20] saying that a nonempty definable set of trees contains a regular tree.)

4. A special case of the main theorem

In this section, we establish that the equational graphs defined by extended systems of a special form are $\exists \mathcal{WL}$ -definable.

4.1. Definition (Special systems). A *special system* is an extended system $S = \langle u_1 = H_1, \dots, u_n = H_n \rangle$ such that the following conditions are satisfied (we let $U = \{u_1, \dots, u_n\}$, and we recall that H_1, \dots, H_n belong to $\mathbf{FCG}(A \cup U \cup U^\infty)$):

(S1) S is separated (i.e., every graph H_i is source-separated and $(U \cup U^\infty)$ -separated).

(S2) Every graph H_i is internally connected, and, if it has no internal vertex, then its unique edge is terminal (i.e., its label is in A).

(S3) The graphs defined by S are connected and nondegenerated.

4.2. Remarks. Conditions (S1) and (S2) are easily verifiable. Condition (S3) can be decided by the result of Courcelle [11] recalled in Theorem 3.6.

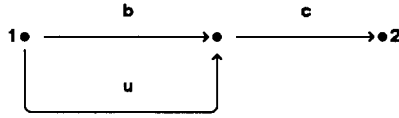
It follows from Conditions (S1) and (S2) that, for every n -tuple (K_1, \dots, K_n) of connected source-separated graphs of respective types $\tau(u_1), \dots, \tau(u_n)$, the

graph $H_i[K_1/u_1, \dots, K_n/u_n, (//^\infty K_1)/u_1, \dots, (//^\infty K_n)/u_n]$ is internally connected. Hence, by (S2) and (S3), the graphs defined by a special system are internally connected. It follows then from Proposition 2.15 that the canonical solution of a special system is the unique one consisting of connected source-separated graphs.

Conditions (S2) and (S3) imply that, if $\tau(u_i) = 0$, then neither u_i nor u_i^∞ occurs in any of the graphs H_1, \dots, H_n .

The system of Example 2.2 is not special.

The graph H of Fig. 1, Example 1.11, is defined by the special system reduced to the equation $u = K$ where K is the following graph:



4.3. Proposition. *The graphs forming the solution of a special system are $\exists \mathcal{WL}$ -definable.*

Here is the idea of the proof. If (G_1, \dots, G_n) is the solution of a special system S , then one can “represent” in each graph G_i its syntactic tree T_i : the nodes of T_i are “represented” by certain well-chosen vertices of G_i , and the **suc** relation of T_i is represented by the existence of certain paths between these vertices. The formula, intended to “say” that a given graph G is isomorphic to G_i , will have to do two things:

(i) to “guess” that G “contains” a representation of the tree T_i in the above sense, and

(ii) to “verify” that G is isomorphic to $\text{eval}_S(T_i)$.

Let us put this more formally. Given S and i , we shall construct a definition scheme Δ with parameters X_1, \dots, X_m , such that the following holds:

(1) $T_i = \text{def}_\Delta(G_i, X_1, \dots, X_m)$ for some sets X_1, \dots, X_m of items of G_i ,

(2) for every graph G , for all sets X_1, \dots, X_m of items of G , if $\text{def}_\Delta(G, X_1, \dots, X_m)$ is defined, then:

(2.1) $\text{def}_\Delta(G, X_1, \dots, X_m)$ is an (n, r) -tree T ,

(2.2) $T = T_i$,

(2.3) $\text{eval}_S(T_i) = G$ (i.e., is isomorphic to G).

It will follow that G_i is defined by the formula $\exists X_1, \dots, X_m \delta$, where δ is the domain formula of Δ . Before starting the proof, we introduce a few definitions and notations, and we establish a few lemmas.

4.4. Definitions and Notations. In what follows, we let S be a special system $\langle u_1 = H_1, \dots, u_n = H_n \rangle$. We assume that H_1, \dots, H_n are pairwise disjoint

concrete graphs. As in Definition 2.12 we let $e_{i,1}, \dots, e_{i,m_i}$ be an enumeration of the set of *nonterminal edges* of H_i , i.e., that have a label in $U \cup U^\infty$, and we let $\mathbf{del}(H_i)$ be obtained from H_i by deleting its nonterminal edges (but $\mathbf{V}_{\mathbf{del}(H_i)} = \mathbf{V}_{H_i}$). We let $r = \mathbf{Max}\{m_i \mid i = 1, \dots, n\}$.

We fix some i_0 in $[n]$, and we let $G = G_{i_0}$. We shall prove that G is definable. Our proof will use $T = T_{i_0}$, the syntactic tree of G_{i_0} .

The notations $N, N_i, \mathbf{lab}, \mathbf{suc}_j, \mathbf{rt}$ will refer to T , as in Definitions 2.3 and 2.12. Let us recall that N is the set of nodes of T , that N_i is the set of its nodes labeled by i , that $\mathbf{lab}: N \rightarrow [n]$ is the labelling function, that $\mathbf{suc}_j = \{(\mu, \mu') \in N \times N \mid \mu' \text{ is a } j\text{-successor of } \mu\}$, and that \mathbf{rt} is the root of T . Hence the label of \mathbf{rt} is i_0 .

Two nodes are *analogous* if they have the same label.

If η is a sequence of integers (j_1, j_2, \dots, j_s) , then $\mathbf{suc}(\eta)$ denotes the set of pairs (μ, μ') such that there exists a sequence $\mu_0, \mu_1, \dots, \mu_s$ with $\mu_0 = \mu$, $\mu' = \mu_s$, and $(\mu_{i-1}, \mu_i) \in \mathbf{suc}_{j_i}$ for all $i = 1, \dots, s$. If $\eta = ()$, then $\mathbf{suc}(\eta) := \{(\mu, \mu) \mid \mu \in N\}$.

We write $\mu' < \mu$ if $(\mu, \mu') \in \mathbf{suc}(\eta)$ for some nonempty sequence η . We say then that μ' is a *descendent* of μ . We write $\mu' \leq \mu$ if $\mu' \leq \mu$ or $\mu' = \mu$.

If $\mu' \leq \mu$, there is a unique sequence η such that $(\mu, \mu') \in \mathbf{suc}(\eta)$. But there may exist several, and even infinitely many, nodes μ' such that $(\mu, \mu') \in \mathbf{suc}(\eta)$ for some μ and η . This is due to the possible presence, in the right-hand sides of the equations of S , of the symbols u_i^∞ , which cause the existence in T of infinitely many j -successors for some nodes.

We now introduce some notations concerning the way G is constructed from T . We denote by K the graph $\bigcup \{\mu \cdot \mathbf{del}(H_i)^0 \mid \mu \in N, i = \mathbf{lab}(\mu)\}$ equipped with the sequence of sources of $\mathbf{rt} \cdot H_i$, as in Definition 2.4. Then $G = K/\simeq$, where \simeq is the equivalence relation defined in Definition 2.4. We let f be the canonical surjective homomorphism: $K \rightarrow K/\simeq = G$.

For every μ , we let K/μ be the graph such that $(K/\mu)^0 = \bigcup \{\mu' \cdot \mathbf{del}(H_i)^0 \mid \mu' \leq \mu, i = \mathbf{lab}(\mu')\}$. This graph is a subgraph of K^0 . We also equip K/μ with the sequence of sources of $\mu \cdot H_{\mathbf{lab}(\mu)}$, as sequence of sources. We let $G/\mu = f(K/\mu)$. Hence $(G/\mu)^0$ is a subgraph of G^0 .

It is clear that, if μ and μ' have the same label i in T , then the subtrees T_μ and $T_{\mu'}$ of T , respectively issued from μ and μ' , are isomorphic. It follows that G/μ and G/μ' are isomorphic: they are both isomorphic to G_i , the i th component of the solution of S .

For every edge e of G , there is a unique pair (μ, e') with $\mu \in N_i$ for some i , and $e' \in \mathbf{E}_{\mathbf{del}(H_i)}$, such that $f((\mu, e')) = e$. We denote it by $(\mu(e), \mathbf{rep}(e))$. We say that $\mathbf{rep}(e)$ *represents* e . Since $f: K \rightarrow K/\simeq = G$ is not injective on \mathbf{V}_K , a vertex w may be equal to $f((\mu, v))$ for several pairs (μ, v) , equivalent to one another w.r.t. \simeq .

Claim. *There is a unique $<$ -maximal node μ of T such that $f((\mu, v)) = w$ for some v .*

This node is denoted by $\mu(w)$, and the associated vertex v by $\mathbf{rep}(w)$.

Proof. If $f((\mu, v)) = f((\mu', v)) = w$, and neither $\mu \leq \mu'$ nor $\mu' \leq \mu$, then there exist v_1, v'_1 such that $w = f((\mu_1, v_1)) = f((\mu'_1, v'_1))$ where μ_1 is the father node of μ , and μ'_1 is the father node of μ' in T . By repeating the argument if necessary, one obtains the existence of a unique $<$ -maximal node μ such that $w = f((\mu, v))$ for some v . Hence μ is the desired $\mu(w)$. (Since the root is at a finite distance, this argument yields the result after finitely many iterations. This proof works for an arbitrary extended system, even if it is not special.)

Since S is separated, and since the graphs G_1, \dots, G_n it defines are source-separated, the homomorphism f is injective on each graph $\mu \cdot \mathbf{del}(H_i)^0$, where $\mu \in N_i$, $i \in [n]$. The unicity of v such that $w = f((\mu(w), v))$ follows. \square

In addition, let us observe that if w is internal in G , then $\mathbf{rep}(w)$ is internal in H_i , where $i = \mathbf{lab}(\mathbf{rep}(w))$. Hence if w is internal in G , then $(\mu(w), \mathbf{rep}(w))$ is the unique pair (μ, v) such that $w = f((\mu, v))$ and v is an internal vertex of H_i , $i = \mathbf{lab}(\mu)$. If w is a source of G , then $\mu(w)$ is the root of T , and $\mathbf{rep}(w)$ is a source of H_{i_0} .

4.5. Lemma. *Let $e \in \mathbf{E}_G$, $i \in [\tau(e)]$ and $w = \mathbf{vert}_G(e, i)$. Then $\mu(e) \leq \mu(w)$. Let furthermore w be internal, let η be such that $(\mu(w), \mu(e)) \in \mathbf{suc}(\eta)$, let μ, μ' be two nodes of T such that $(\mu, \mu') \in \mathbf{suc}(\eta)$, and μ is analogous to $\mu(w)$. If $e' = f((\mu', \mathbf{rep}(e)))$ and $w' = \mathbf{vert}_G(e', i)$, then $\mu(w') = \mu$. Hence $w' = f((\mu, \mathbf{rep}(w)))$.*

Proof. The first assertion is clear.

There is a unique (possibly empty) sequence $\eta = (j_1, \dots, j_m)$ such that $(\mu(w), \mu(e)) \in \mathbf{suc}(\eta)$.

There is also a sequence $\bar{\mu}_0, \bar{\mu}_1, \dots, \bar{\mu}_m$ of nodes of T , of respective labels l_0, l_1, \dots, l_m , such that $\bar{\mu}_0 = \mu(w)$, $\bar{\mu}_m = \mu(e)$, and $(\bar{\mu}_{k-1}, \bar{\mu}_k) \in \mathbf{suc}_{j_k}$ for all $k = 1, \dots, m$.

There is a sequence of integers i_0, \dots, i_{m-1} such that the following conditions hold:

(A1) In H_{l_m} , the i th vertex of $\mathbf{rep}(e)$ is equal to the i_{m-1} th source.

(A2) In H_{l_k} , the i_k th vertex of $e_{l_k, j_{k+1}}$ is equal to the i_{k-1} th source, for $1 \leq k < m$,

(A3) In H_{l_0} , the i_0 th vertex of e_{l_0, j_1} is the internal vertex $\mathbf{rep}(w)$.

(Let us recall from Definitions 2.3, 2.12, and 4.4 that $(e_{i_1, 1}, \dots, e_{i_m, m})$ denotes an enumeration of the set of edges of H_i that have a label in $U \cup U^\infty$.)

Let us now assume that $(\mu, \mu') \in \mathbf{suc}(\eta)$, and that μ is analogous to $\mu(w)$. Let $\mu_0, \mu_1, \dots, \mu_m$ be the sequence such that $\mu' = \mu_m < \mu_{m-1} < \dots < \mu_2 < \mu_1 < \mu_0 = \mu$. Hence $\mathbf{lab}(\mu_i) = l_i$ for all i and μ' is analogous to $\mu(e)$.

Let $e' = f((\mu', \mathbf{rep}(e)))$, and $w' = \mathbf{vert}_G(e', i)$.

By (A1), w' is the i_{m-1} th source of $G/\mu' = G/\mu_m$.

By (A2), w' is the i_{k-1} th source of G/μ_k for $k = m - 1, m - 2, \dots, 1$ (this can be proved by reverse induction from $m - 1$ to 1).

By (A3), w' is $f((\mu_0, \mathbf{rep}(w)))$, and $\mathbf{rep}(w)$ is internal in H_{l_0} .

This proves that $\mu(w') = \mu_0 = \mu$. \square

4.6. Definitions (Paths and their traces). By Definition 1.6, a path γ in G can be written:

$$\gamma = (w_0, e_1, i_1, j_1, w_1)(w_1, e_2, i_2, j_2, w_2) \cdots (w_{m-1}, e_m, i_m, j_m, w_m). \quad (1)$$

We define its *trace* as the sequence;

$$\begin{aligned} \mathbf{tr}(\gamma) = & (\mathbf{rep}(w_0), \mathbf{rep}(e_1), i_1, j_1, \mathbf{rep}(w_1)) \\ & (\mathbf{rep}(w_1), \mathbf{rep}(e_2), i_2, j_2, \mathbf{rep}(w_2)) \cdots \\ & \cdots (\mathbf{rep}(w_{m-1}), \mathbf{rep}(e_m), i_m, j_m, \mathbf{rep}(w_m)). \end{aligned}$$

This sequence is a word in B^+ , where B is the finite set: $V \times E \times [r_A] \times [r_A] \times V$, and $V = \mathbf{V}_{H_1} \cup \dots \cup \mathbf{V}_{H_n}$, $E = \mathbf{E}_{\mathbf{del}(H_1)} \cup \dots \cup \mathbf{E}_{\mathbf{del}(H_n)}$, $r_A = \mathbf{Max}\{\tau(a) \mid a \in A\}$.

We also denote by $\mathbf{edg}(\gamma, i)$ the i th edge of γ , i.e., the edge e_i of γ written as above.

We say that γ is *strongly internal* in G , if it is internal in G/μ where $\mu = \mu(w_0)$. In this case, we let $\eta_0, \dots, \eta_m, \bar{\eta}_1, \dots, \bar{\eta}_m$ be the sequences in \mathbb{N}^* such that $(\mu, \mu(w_i)) \in \mathbf{suc}(\eta_i)$, and $(\mu, \mu(e_i)) \in \mathbf{suc}(\bar{\eta}_i)$. Note that $\eta_0 = ()$, and that η_i is a prefix of $\bar{\eta}_i$ and of $\bar{\eta}_{i+1}$ for all $i = 0, \dots, m$ (this is a consequence of Lemma 4.5).

Two strongly internal paths are *analogous* if they have the same trace, whence the same length m , and the same associated sequences $(\eta_i)_{0 \leq i \leq m}$ and $(\bar{\eta}_i)_{1 \leq i \leq m}$. Hence, if two analogous paths link respectively w to w' , and w_1 to w'_1 , then $\mathbf{rep}(w_1) = \mathbf{rep}(w)$, $\mathbf{rep}(w'_1) = \mathbf{rep}(w')$, $(\mu(w), \mu(w')) \in \mathbf{suc}(\eta)$ and $(\mu(w_1), \mu(w'_1)) \in \mathbf{suc}(\eta)$ for some $\eta \in \mathbb{N}^*$. Conversely:

4.7. Lemma. *Let γ be a strongly internal path linking w to w' . For every node μ' analogous to $\mu(w)$, for every μ'' such that $(\mu', \mu'') \in \mathbf{suc}(\eta)$, where η is such that $(\mu(w), \mu(w')) \in \mathbf{suc}(\eta)$, there is a path γ' analogous to γ , linking $f((\mu', \mathbf{rep}(w)))$ to $f((\mu'', \mathbf{rep}(w')))$.*

Proof. The construction of γ' from γ is straightforward. \square

Note that since the relations \mathbf{suc}_j are not functional in general (as we have already noted), there may exist several paths γ' satisfying this lemma. We assume that a canonical way to select one of them has been defined. The corresponding path is denoted by $\mu' \cdot \gamma$ (with $\mu' \cdot \gamma = \gamma$ if $\mu' = \mu(w)$, and w is the origin of γ).

In the following lemma, we let γ be a strongly internal path of the above form (1), and we let $E_{\gamma, i} = \{\mathbf{edg}(\mu' \cdot \gamma, i) \mid \mu' \text{ is analogous to } \mu\}$ for $1 \leq i \leq m$. When a single path γ is considered, the notation E_i is used instead of $E_{\gamma, i}$. Here is a

reconstruction lemma, saying that a path of the form $\mu' \cdot \gamma$ is defined in a unique way by its origin and the sets $E_{\gamma,i}$.

4.8. Lemma. *Let π be an internal path of G with trace $\mathbf{tr}(\gamma)$, and such that $\mathbf{edg}(\pi, i) \in E_{\gamma,i}$ for all $i = 1, \dots, m$. Then $\pi = \mu' \cdot \gamma$ for some node μ' of T that is analogous to μ .*

Proof. Since $\mathbf{tr}(\pi) = \mathbf{tr}(\gamma)$, the path π is of the form:

$$\pi = (w_0'', e_1'', i_1, j_1, w_1'') \cdots (w_{m-1}'', e_m'', i_m, j_m, w_m'').$$

Let $\mu' = \mu(w_0'')$. We have $w_0'' = f((\mu', \mathbf{rep}(w_0)))$, and μ' is analogous to μ , since, from the equality $\mathbf{tr}(\pi) = \mathbf{tr}(\gamma)$, we have $\mathbf{rep}(w_0'') = \mathbf{rep}(w_0)$.

Let then $\gamma' := \mu' \cdot \gamma$ be written as:

$$\gamma' = (w_0', e_1', i_1, j_1, w_1') \cdots (w_{m-1}', e_m', i_m, j_m, w_m') \quad \text{with } w_0' = w_0''.$$

We shall prove, by induction on k , that, for all $k = 0, \dots, m-1$:

$$w_k'' = w_k' \quad \text{and} \quad e_{k'}'' = e_{k'}' \quad \text{for all } k', 1 \leq k' < k.$$

Basis ($k=0$). The equality $w_0'' = w_0'$ follows from the definition of γ' .

Inductive step. We assume that $w_k'' = w_k'$ and $e_{k'}'' = e_{k'}'$ for all $k' < k$.

The sequences $(\eta_i)_{0 \leq i \leq m}$ and $(\bar{\eta}_i)_{1 \leq i \leq m}$ are as in Definition 4.6, and they are the same for γ , and for γ' .

By the definition of γ' , we have:

$$e_{k+1}' = f((\mu(e_{k+1}'), \mathbf{rep}(e_{k+1}))) \quad \text{with } (\mu', \mu(e_{k+1}')) \in \mathbf{suc}(\bar{\eta}_{k+1}) \quad (1)$$

We also have:

$$w_k' = \mathbf{vert}_G(e_{k+1}', i_{k+1}), \quad (2)$$

and

$$(\mu', \mu(w_k')) \in \mathbf{suc}(\eta_k). \quad (3)$$

By Lemma 4.5 we have $\mu(e_{k+1}') \leq \mu(w_k') \leq \mu'$ and:

$$(\mu(w_k'), \mu(e_{k+1}')) \in \mathbf{suc}(\eta), \quad (4)$$

where η is such that $\bar{\eta}_{k+1} = \eta_k \cdot \eta$.

Let us now consider π . Since $e_{k+1}'' \in E_{\gamma, k+1}$, we have $e_{k+1}'' = \mathbf{edg}(v \cdot \gamma, k+1)$, and:

$$e_{k+1}'' = f((\bar{v}, \mathbf{rep}(e_{k+1}))) \quad \text{and} \quad (v, \bar{v}) \in \mathbf{suc}(\bar{\eta}_{k+1}), \quad (5)$$

for some v analogous to μ (and to μ'), and some \bar{v} , analogous to $\mu(e_{k+1}')$.

We also have:

$$w_k'' = \mathbf{vert}_G(e_{k+1}'', i_{k+1}). \quad (6)$$

By (5), and since η is a suffix of $\bar{\eta}_{k+1}$, there exists v' such that $\bar{v} \leq v' \leq v$, and:

$$(v', \bar{v}) \in \mathbf{suc}(\eta). \quad (7)$$

It follows that:

$$(v, v') \in \mathbf{succ}(\eta_k). \quad (8)$$

Hence v' is similar to $\mu(w'_k)$.

Let us consider conditions (2) and (4), (6) and (7), and (5). Since w'_k is internal (because γ' is similar to γ), we can apply Lemma 4.5, and we obtain:

$$v' = \mu(w''_k).$$

Since we assume that $w'_k = w''_k$, we have;

$$v' = \mu(w'_k).$$

Hence, by (3) and (8), we have:

$$v = \mu'.$$

Hence:

$$\begin{aligned} e''_{k+1} &= \mathbf{edg}(v \cdot \gamma, k+1) && \text{(by the definition of } \pi) \\ &= \mathbf{edg}(\mu' \cdot \gamma, k+1) && \text{(since } v = \mu') \\ &= e'_{k+1}. \end{aligned}$$

It follows immediately that $w''_{k+1} = w'_{k+1}$. \square

Let us recall that H_1, \dots, H_n are pairwise disjoint concrete graphs. Let $\{v_1, v_2, \dots, v_N\}$ be an enumeration of the set $\mathbf{IV}_{H_1} \cup \mathbf{IV}_{H_2} \cup \dots \cup \mathbf{IV}_{H_n}$. Let also $\{e_1, \dots, e_{N'}\}$ be an enumeration of the set of edges $\mathbf{E}_{\mathbf{del}(H_1)} \cup \dots \cup \mathbf{E}_{\mathbf{del}(H_n)}$. Let X_i be the set of internal vertices w of G , such that $\mathbf{rep}(w) = v_i$. Let Y_j be the set of edges e of G , such that $\mathbf{rep}(e) = e_j$. Let also $E_{\gamma,1}, \dots, E_{\gamma,m_\gamma}$ be associated with a path γ as in Lemma 4.8.

With these notations, we can state the following corollary of Lemma 4.8, where $X_1, \dots, X_N, Y_1, \dots, Y_{N'}, E_{\gamma,1}, \dots, E_{\gamma,m_\gamma}$ are simultaneously the above defined sets of vertices and of edges, and set variables denoting them. This ambiguity simplifies the statement and is harmless.

4.9. Corollary. *Let $i \in [n]$, let c be an internal vertex of H_i , let v be another vertex of H_i . Let γ be an internal path in G/μ linking $f((\mu, c))$ to $f((\mu, v))$ for some μ in N_i . One can construct a first-order formula $\varphi(w, w', E_{\gamma,1}, \dots, E_{\gamma,m_\gamma}, X_1, \dots, X_N, Y_1, \dots, Y_{N'})$ such that, if w and w' are vertices of G then:*

$$\begin{aligned} &G \models \varphi(w, w', E_{\gamma,1}, \dots, E_{\gamma,m_\gamma}, X_1, \dots, X_N, Y_1, \dots, Y_{N'}) \\ \text{iff} & \\ &w = f((\mu', c)) \quad \text{and} \quad w' = f((\mu', v)), \quad \text{for some } \mu' \text{ in } N_i. \end{aligned}$$

Proof. Let γ be as in Definition 4.6. One can construct a first-order formula φ that says the following:

there exist w_1, w_2, \dots, w_{m-1} in \mathbf{V}_G , there exist f_1, \dots, f_m in \mathbf{E}_G such that:
 $(w, f_1, i_1, j_1, w_1)(w_1, f_2, i_2, j_2, w_2) \cdots (w_{m-1}, f_m, i_m, j_m, w')$
 is an internal path of G with trace $\mathbf{tr}(\gamma)$, such that
 $f_i \in E_{\gamma, i}$ for all $i = 1, \dots, m$.

If $w = f((\mu', c))$, and $w' = f((\mu', v))$ then $\mu' \cdot \gamma$ is such a path, hence $\varphi(w, w', E_{\gamma, 1}, \dots, X_1, \dots, Y_1, \dots)$ holds.

Conversely, if $\varphi(w, w', E_{\gamma, 1}, \dots, X_1, \dots, Y_1, \dots)$ holds, there is an internal path π linking w to w' , such that $\mathbf{tr}(\pi) = \mathbf{tr}(\gamma)$ and $\mathbf{edg}(\pi, i) \in E_{\gamma, i}$ for all $i = 1, \dots, m$.

By Lemma 4.8, $\pi = \mu' \cdot \gamma$ for some $\mu' \in N$. Hence, by Lemma 4.7, π links $f((\mu', c))$ to $f((\mu', v))$. Hence w and w' are as wanted. \square

We can now prove Proposition 4.3.

Proof of Proposition 4.3. We first assume that each of the graphs H_i has at least one internal vertex. This restriction will be lifted in the third part of the proof.

We also assume that each $i \in [n]$ labels at least one node of T . The integers i labelling no node correspond to unknowns of S that are useless for the definition of $G = G_{i_0}$. They can be easily detected and eliminated.

First part: Representation of T in G

The (n, r) -tree T will be defined as a relational structure $\langle V, \mathbf{prt}, (\mathbf{plab}_i)_{i \in [n]}, (\mathbf{suc}_j)_{j \in [r]} \rangle$ (see Definition 3.8). Our aim is to construct a definition scheme Δ defining T in G in terms of the sets $X_1, \dots, Y_1, \dots, E_{\gamma, 1}, \dots$ introduced above and a few others.

For every $i \in [n]$, let us select an internal vertex c_i of H_i . Let $g: N \rightarrow \mathbf{V}_G$ be the mapping such that $g(v) = f((v, c_i))$ where $i = \mathbf{lab}(v)$. Let $V_i = g(N_i)$. Note that $\mu(g(v)) = v$, and that $\mathbf{rep}(g(v)) = c_i$ if $i = \mathbf{lab}(v)$. The mapping g defines a bijection $N_i \rightarrow V_i$ for all i . We have $g(\mathbf{rt}) \in V_{i_0}$ (the label of T is i_0). We let $V := V_1 \cup \dots \cup V_n$.

We shall establish that the binary relation on V :

$$\mathbf{SUC}_j := \{(g(v), g(v')) \mid (v, v') \in \mathbf{suc}_j\}$$

is definable in G in terms of the sets of vertices $V_1, \dots, V_n, X_1, \dots, X_N$, of the sets of edges Y_1, \dots, Y_N (the sets X_i, Y_j are as in Corollary 4.9), and of sets of edges $E_{\gamma, 1}, \dots, E_{\gamma, m_\gamma}$ associated as in Lemma 4.8 with finitely many paths γ of a certain finite set Γ .

We first define Γ . For every $i \in [n]$, for every vertex v of H_i , such that $v \neq c_i$, we let $\gamma_{i,v}$ be an internal path in G/μ that links $f((\mu, c_i))$ to $f((\mu, v))$, where μ is a

node of N_i . (We have assumed that $N_i \neq \emptyset$ for all i at the beginning of the proof.) Let Γ be the set of all these paths.

For each path γ in Γ , let m_γ be its length, and let $E_{\gamma,j} = \{\mathbf{edg}(\mu' \cdot \gamma, j) \mid \mu' \in N_i\}$, for $1 \leq j \leq m_\gamma$, as in Lemma 4.8.

We shall denote $\{X_1, \dots, X_N\}$ by \mathcal{X} , $\{Y_1, \dots, Y_{N'}\}$ by \mathcal{Y} , $\{V_1, \dots, V_n\}$ by \mathcal{V} , and a fixed enumeration of $\{E_{\gamma,1}, \dots \mid \gamma \in \Gamma\}$ by \mathcal{E} . We shall denote by $\varphi_{i,v}(w, w', \mathcal{V}, \mathcal{X}, \mathcal{Y}, \mathcal{E})$ the first-order formula associated with $\gamma_{i,v}$ by Corollary 4.9. If $v = c_i$, we let $\varphi_{i,v}$ be the formula $w = w'$.

Claim 1. \mathbf{SUC}_j is definable in terms of \mathcal{V} , \mathcal{X} , \mathcal{Y} and \mathcal{E}

Let us recall that $r = \mathbf{Max}\{m_i \mid i \in [n]\}$. Let L be the set of triples (i, j, i') in $[n] \times [r] \times [n]$, such that the j -successors of a node μ of N_i are in $N_{i'}$. Let $(i, j, i') \in L$. Let also

$$v_k = \mathbf{vert}_{H_i}(e_{i,j}, k), \quad \text{and} \quad v'_k = \mathbf{src}_{H_{i'}}(k), \quad \text{for all } 1 \leq k \leq m = \tau(e_{i,j}).$$

For every $\mu \in N_i$, $\mu' \in N_{i'}$ such that $(\mu, \mu') \in \mathbf{suc}_j$, we have

$$f((\mu, v_k)) = f((\mu', v'_k)), \quad \text{for all } k = 1, \dots, m.$$

Hence $\mu \cdot \gamma_{i,v_k}$ links $g(\mu) = f((\mu, c_i))$ to $f((\mu, v_k))$, and $\mu' \cdot \gamma_{i',v'_k}$ links $g(\mu') = f((\mu', c_{i'}))$ to $f((\mu', v'_k)) = f((\mu, v_k))$.

It follows that $(g(\mu), g(\mu'))$ satisfies the first-order formula $\psi_{i,j,i'}(w, w', \mathcal{V}, \mathcal{X}, \mathcal{Y}, \mathcal{E})$ defined as:

$$\exists w_1, \dots, w_m \left[\bigwedge_{1 \leq k \leq m} \varphi_{i,v_k}(w, w_k, \mathcal{V}, \mathcal{X}, \dots) \wedge \varphi_{i',v'_k}(w', w_k, \mathcal{V}, \mathcal{X}, \dots) \right].$$

Let us conversely assume that (w, w') satisfies this formula for some w_1, \dots, w_m . Corollary 4.9 says that:

$$\begin{aligned} w &= f((\mu, c_i)) = g(\mu), \\ w'' &= f((\mu, v_k)) = f((\mu', v'_k)), \quad \text{for all } 1 \leq k \leq m, \quad \text{and} \\ w' &= f((\mu', c_{i'})) = g(\mu') \end{aligned}$$

for some $\mu \in N_i$, and some $\mu' \in N_{i'}$. Since S is special, this implies that $(\mu, \mu') \in \mathbf{suc}_j$.

Hence, \mathbf{SUC}_j is defined by the (finite) disjunction of the formulas $\psi_{i,j,i'}$, for all i, i' such that $(i, j, i') \in L$. \square

Let $X_1, \dots, V_1, \dots, Y_1, \dots, E_{\gamma,1}, \dots$ be as above. Let in addition $X_{\mathbf{rt}} = \{g(\mathbf{rt})\}$. (We recall that \mathbf{rt} is the root of T .) We now define a definition scheme Δ , using these sets as parameters, intended to define T in G . We let:

$$\Delta = \langle \delta, \psi, \theta_{\mathbf{prt}}, (\theta_{\mathbf{plab}})_{i \in [n]}, (\theta_{\mathbf{suc}})_{j \in [r]} \rangle$$

where

$\psi(x)$ says that x belongs to $V_1 \cup \dots \cup V_n$,

$\theta_{\text{prt}}(x)$ says that x belongs to X_{rt} ,

$\theta_{\text{plab}_i}(x)$ says that x belongs to V_i ,

$\theta_{\text{suc}_j}(x, y)$ is the formula of Claim 1, that defines SUC_j .

Finally, we let δ express that the $\mathbf{K}(n, r)$ -structure defined by these formulas is the syntactic tree T . (The existence of such a formula follows from Proposition 3.3.) This achieves goals (1), (2.1), and (2.2) presented after the statement of Proposition 4.3.

Second part: Verification that a graph G' in which the tree T is represented is isomorphic to $\text{eval}(T)$

All notations are as in the first part.

We shall construct a formula $\delta'(\mathcal{V}, \mathcal{X}, \mathcal{Y}, \mathcal{E}, X_{\text{rt}})$ such that:

(V1) $(G, \mathcal{V}, \mathcal{X}, \mathcal{Y}, \mathcal{E}, X_{\text{rt}}) \models \delta'$,

(V2) for every graph G' , if $\langle \mathcal{V}', \mathcal{X}', \mathcal{Y}', \mathcal{E}', X'_{\text{rt}} \rangle$ represents T in G' ,
and if $(G', \mathcal{V}', \mathcal{X}', \mathcal{Y}', \mathcal{E}', X'_{\text{rt}}) \models \delta'$, then G' is isomorphic to G .

The construction of δ' satisfying (V1) and (V2) completes the proof, because G is then defined by the formula:

$$\exists \mathcal{V}, \mathcal{X}, \mathcal{Y}, \mathcal{E}, X_{\text{rt}} [\delta \wedge \delta'].$$

where δ is as in the first part.

Let us recall that \approx denotes an equivalence relation on the intermediate graph K , defined as the union of the graphs $\mu \cdot \text{del}(H_i)^0$ for all $i \in [n]$, $\mu \in N_i$, and that the vertices of K are pairs of the form (μ, v) for $\mu \in N_i$, $v \in \mathbf{V}_{H_i}$, $i \in [n]$.

Claim 2. For every $i, i' \in [n]$, $v \in \mathbf{V}_{H_i}$, $v' \in \mathbf{V}_{H_{i'}}$, the binary relation $\approx_{v, v'}$ on N such that $\mu \approx_{v, v'} \mu'$ iff $(\mu, v) \approx (\mu', v')$ is \mathcal{WL} -definable in the relational structure representing T .

Proof. If a binary relation is \mathcal{WL} -definable, then so is its transitive closure, hence, so is the equivalence relation it generates. By Definitions 2.3 and 2.13, \approx is the equivalence relation generated by a binary relation on \mathbf{V}_K , that is first-order definable in terms of prt , $(\text{plab}_i)_{i \in [n]}$, $(\text{suc}_j)_{j \in [r]}$. The result follows, but we omit the technical details. \square

Claim 3. For every i, i', v, v' as in Claim 2, one can construct a weak formula $\eta_{v, v'}(w, w', \mathcal{V}, \mathcal{X}, \mathcal{Y}, \mathcal{E}, X_{\text{rt}})$ such that, for every $w, w' \in \mathbf{V}_G$

$$(G, w, w', \mathcal{V}, \mathcal{X}, \mathcal{Y}, \mathcal{E}, X_{\text{rt}}) \models \eta_{v, v'}$$

iff

$$w \in V_i, \quad w' \in V_{i'}, \quad \text{and} \quad g^{-1}(w) \approx_{v, v'} g^{-1}(w').$$

Proof. If in Proposition 3.3, the formulas forming Δ , and the formula β are weak formulas, then the formula β can be constructed to be weak. This claim is then an immediate consequence of Claim 2 since the relational structure representing T is defined in G' by a definition scheme made of weak formulas. \square

It follows that $f((\mu, v)) = f((\mu', v'))$ iff the pair $(g(\mu), g(\mu'))$ satisfies the formula $\eta_{v,v'}$, where $\mu \in N_i$, $\mu' \in N_{i'}$, $v \in \mathbf{V}_{H_i}$, $v' \in \mathbf{V}_{H_{i'}}$.

Let us now consider a graph G' , and arbitrary sets of vertices and edges \mathcal{V}' , \mathcal{X}' , \mathcal{Y}' , \mathcal{E}' , X'_n satisfying δ , i.e., representing T in G' .

Let $i \in [n]$, $v \in \mathbf{V}_{H_i}$. We say that a vertex w of G' is (i, v) -defined from s , where s is also a vertex of G' if:

$$G' \models \varphi_{i,v}(s, w, \mathcal{V}', \mathcal{X}', \mathcal{Y}', \mathcal{E}', X'_n).$$

Let us consider the following conditions on G' , \mathcal{V}' , \mathcal{X}' etc.

(D1) For every $i \in [n]$, for every $v \in \mathbf{V}_{H_i}$, for every $s \in V_i$, there exists one and only one vertex of G' , that is (i, v) -defined from s .

(D2) Every vertex w of G' is (i, v) -defined from s , for some i in $[n]$, some v in \mathbf{V}_{H_i} , some s in V_i .

(D3) For every $i, i' \in [n]$, every $v \in \mathbf{V}_{H_i}$, every $v' \in \mathbf{V}_{H_{i'}}$, every $s \in V_i$, every $s' \in V_{i'}$, if a vertex w of G' is (i, v) -defined from s , then it is also (i', v') -defined from s' iff $\eta_{v,v'}(s, s')$ holds in G' .

(D4) For every $i \in [n]$, $e \in \mathbf{E}_{\text{del}(H_i)}$ and $v_1, \dots, v_k \in \mathbf{V}_{H_i}$ such that $\mathbf{vert}_{H_i}(e) = (v_1, \dots, v_k)$, then, for every $s \in V_i$, there is in G' a unique edge e_s such that $\mathbf{lab}_{G'}(e_s) = \mathbf{lab}_{H_i}(e)$, $\mathbf{vert}_{G'}(e_s) = (w_1, \dots, w_k)$ where w_j is (i, v_j) -defined from s for every $j = 1, \dots, k$, and e_s belongs to the component Y'_i of \mathcal{Y}' corresponding to e (see the notations of Corollary 4.9; we have $1 \leq i \leq N'$).

(D5) Every edge of G' is e_s for one and only one pair (e, s) as in (D4).

(D6) For all $k \in [\tau(G)]$, $\mathbf{src}_{G'}(k)$ is (i_0, v) -defined from the unique vertex in X'_n , where v is the k th source of H_{i_0} .

It is clear that each of these conditions is expressible by a weak formula. Hence their conjunction can be expressed by a weak formula $\delta'(\mathcal{V}, \mathcal{X}, \mathcal{Y}, \mathcal{E}, X'_n)$ so that $\langle G', \mathcal{V}', \mathcal{X}', \mathcal{Y}', \mathcal{E}', X'_n \rangle$ satisfies (D1)–(D6) iff:

$$G' \models \delta'(\mathcal{V}', \mathcal{X}', \mathcal{Y}', \mathcal{E}', X'_n).$$

Let us verify that $\langle G, \mathcal{V}, \mathcal{X}, \mathcal{Y}, \mathcal{E}, \{\mathbf{g}(\mathbf{rt})\} \rangle$ satisfies (D1)–(D6), whence δ' . The validity of (D1), (D2), and (D4)–(D6) follows from Corollary 4.9. That of (D3) follows from Claim 3.

Hence, the formula δ' satisfies condition (V1). We shall now verify condition (V2). Let $\langle G', \mathcal{V}', \mathcal{X}', \mathcal{Y}', \mathcal{E}', X'_n \rangle$ satisfy δ and δ' . We shall construct an isomorphism between G' and G .

Since δ holds, $T' := \mathbf{def}_\Delta(G', \mathcal{V}', \mathcal{X}', \dots)$ is isomorphic to T by some function $g': T \rightarrow T'$. Hence g' defines a bijection $N_i \rightarrow V'_i$. Let K be the intermediate graph, and $f': K \rightarrow G'$ be such that:

(1) $f'((\mu, v))$ is the unique vertex of G' that is (i, v) -defined from $g'(\mu)$, where $i = \mathbf{lab}(\mu)$; it is well-defined by (D1).

(2) $f'((\mu, e))$ is the edge e , defined by (D4) with $s = g'(\mu)$.

Conditions (D4) and (D6) entail that f' is a homomorphism. Conditions (D2) and (D5) entail that it is surjective. Condition (D5) entails that f' is injective on \mathbf{E}_K . Finally, it follows from condition (D3), and Claims 2 and 3, that the kernel of f' is \simeq . Hence f' factorizes as $h \circ f$ where h is an isomorphism: $K/\simeq = G \rightarrow G'$ and f is the canonical surjection: $K \rightarrow G$.

Third part: Lifting a restriction

At the beginning, we have assumed that each graph H_i has at least one internal vertex, and this allowed us to choose an internal vertex c_i in H_i for each $i = 1, \dots, n$.

If H_i has no internal vertex, then it is reduced to an edge labeled in A , all vertices of which are sources.

We take this edge as item c_i . The construction done in the first two parts must be adapted accordingly. We omit the lengthy technical details that do not need any new idea. \square

5. Proof of the main theorem

Here is the main result of this paper.

5.1. Theorem. *Every equational graph is $\exists\mathcal{WL}$ -definable.*

Proof. Let G be an equational graph. It follows from Lemma 2.10 that the source-separated graph $\mathbf{sep}(G)$ is equational. From a formula in $\exists\mathcal{WL}$ defining $\mathbf{sep}(G)$, it is easy to construct a formula of the same type defining G . Hence, we shall do the proof for a source-separated graph G .

Our purpose is to construct a formula defining G from a system of equations defining it, not only to prove the existence of a formula.

We know by Lemma 1.10 that G can be expressed as a source-preserving composition:

$$G = f(G_1, \dots, G_p, //^\infty G_{p+1}, \dots, //^\infty G_m)$$

of its internally connected components. We shall construct f and a special system defining (G_1, \dots, G_m) . The result will follow then from Theorem 4.4, saying that these graphs are $\exists\mathcal{WL}$ -definable, and from Proposition 3.8, saying that source-preserving compositions of $\exists\mathcal{WL}$ -definable internally connected graphs are $\exists\mathcal{WL}$ -definable. Since these results are effective, a formula defining G can be effectively obtained.

The proof consists of three parts.

First part: Definition of a recursive set \mathbf{DEF}' , and of a bijection of \mathbf{DEF}' onto $\mathbf{CICC}(G)$

We let G be a concrete equational graph, defined as a component of the solution of a separated system $S = \langle u_1 = H_1, \dots, u_n = H_n \rangle$. We let $U = \{u_1, \dots, u_n\}$, and (G_1, \dots, G_n) be its solution.

We let T be the syntactic tree of G , we let K be the intermediate graph used to define $\mathbf{eval}(T)$, and we let f be the surjective canonical graph homomorphism: $K \rightarrow G := K/\approx$. We shall use the notations of the proof of Proposition 4.3.

We denote by \mathbf{S}_G the set of sources of G .

We denote by $\mathbf{CICC}_0(G)$ the set of cicc's of G that have no internal vertex (they are reduced to single edges, all vertices of which are sources), and by $\mathbf{CICC}_1(G)$ the set of the others.

Let C belong to $\mathbf{CICC}_0(G)$. There is a unique pair (μ, e) such that μ is a node of T , i is its label, e is an edge of H_i , and $C = G \upharpoonright \{f((\mu, e))\}$. We let \mathbf{DEF}_0 be the set of such pairs, and $\mathbf{C}(\mu, e)$ be the cicc associated as above with (μ, e) . It can be characterized as follows:

$$(\mu, e) \in \mathbf{DEF}_0 \quad \text{iff} \quad \mu \in N_i, i = \mathbf{lab}(\mu), e \in \mathbf{E}_{H_i}, \text{ and for every vertex } s \text{ of } e, f((\mu, s)) \in \mathbf{S}_G \text{ (this is possible only if } s \text{ is a source of } H_i).$$

Let us make this characterization effective. For every $i = 1, \dots, n$, for every source vertex v of H_i , for every source vertex w of G , one can construct a regular language $L(i, v, w)$, such that, for every node μ of T with label i :

$$f((\mu, v)) = w \quad \text{iff} \quad (\mathbf{rt}, \mu) \in \mathbf{suc}(\eta) \quad \text{for some } \eta \text{ in } L(i, v, w).$$

If A and B are sets of vertices, we let $L(i, v, B)$ denote the union of the sets $L(i, v, w)$ for w in B , and we let $L(i, A, B)$ denote the intersection of the sets $L(i, v, B)$ for v in A . Hence:

$$(\mu, e) \in \mathbf{DEF}_0 \quad \text{iff} \quad \text{every vertex of } e \text{ is a source of } H_i, \text{ and } (\mathbf{rt}, \mu) \in \mathbf{suc}(\eta) \text{ for some } \eta \text{ in } L(i, V, \mathbf{S}_G), \text{ where } i \text{ is the label of } \mu, \text{ and } V \text{ is the set of vertices of } e.$$

Hence $\mathbf{CICC}_0(G) = \{\mathbf{C}(x) \mid x \in \mathbf{DEF}_0\}$, and it is clear that $\mathbf{C}(x) = \mathbf{C}(x')$ iff $x = x'$.

We shall obtain a similar, but more complicated indexing of $\mathbf{CICC}_1(G)$. Let $C \in \mathbf{CICC}_1(G)$. There exists a unique \leftarrow -maximal node μ of T such that $f((\mu, v))$ is an internal vertex of C (hence of G), for some vertex v . We denote this node by $\mu(C)$, and we denote by $\mathbf{REP}(C)$ the set of vertices v such that $f((\mu, v)) \in \mathbf{IV}_C$. Note that $\mathbf{REP}(C) \subseteq \mathbf{IV}_{H_i}$ where $i = \mathbf{lab}(\mu(C))$. For every $v \in \mathbf{REP}(C)$, the subgraph C is the unique cicc of G containing $f((\mu(C), v))$. We let $\mathbf{DEF}_1 := \{(\mu(C), v) \mid v \in \mathbf{REP}(C), C \in \mathbf{CICC}_1(G)\}$. We say that C is *defined* by (μ, v) if $\mu = \mu(C)$ and $v \in \mathbf{REP}(C)$, and we write in this case $C = \mathbf{C}(\mu, v)$.

In order to characterize \mathbf{DEF}_1 , we give another definition. If H is a graph in

$\mathbf{FCG}(A \cup U)$, if v, v' are two vertices of H , a *strong internal path in H , linking v to v'* , is an internal path of the form:

$$(v, e_1, i_1, j_1, v_1)(v_1, e_2, i_2, j_2, v_2) \cdots (v_{k+1}, e_k, i_k, j_k, v')$$

where, for every $m \in [k]$, either e_m is a terminal edge, or e_m is labeled by u_p for some $p \in [n]$, and $\mathbf{src}_{G_p}(i_m)$ is linked to $\mathbf{src}_{G_p}(j_m)$ by a path in G_p . In this situation, we say that v and v' are *strongly linked* in H . We denote by $\mathbf{SL}(i, v)$ the set of source vertices of H_i , $i \in [n]$ that are strongly linked to some vertex v of H_i .

5.2. Lemma. *Let $\mu \in N_i$, $v \in \mathbf{IV}_{H_i}$ for some $i \in [n]$. Then $(\mu, v) \in \mathbf{DEF}_1$ iff for every v' in $\mathbf{SL}(i, v)$, the vertex $f((\mu, v'))$ is a source of G .*

Proof. Let $v' \in \mathbf{SL}(i, v)$. If $w' = f((\mu, v'))$ belongs to \mathbf{IV}_G , then the cicc C of G containing $w = f((\mu, v))$ is not defined by (μ, v) , because it contains also w' . Hence $\mu(C) \geq \mu(w') > \mu(w)$, and (μ, v) does not belong to \mathbf{DEF}_1 .

The proof of the other direction is similar. \square

Note that, if $(\mu, v) \in \mathbf{DEF}_1$, the set of sources of $\mathbf{C}(\mu, v)$ is the set $\{f((\mu, v')) \mid v' \in \mathbf{SL}(i, v)\}$ where $i = \mathbf{lab}(\mu)$. Hence, the set \mathbf{DEF}_1 can be effectively characterized as follows:

$$(\mu, v) \in \mathbf{DEF}_1 \quad \text{iff} \quad \text{if } i = \mathbf{lab}(\mu), \text{ then } v \text{ is an internal vertex of } H_i, \text{ and} \\ (\mathbf{rt}, \mu) \in \mathbf{suc}(\eta) \text{ for some } \eta \text{ in } L(i, \mathbf{SL}(i, v), S_G).$$

Hence, we have $\mathbf{CICC}_1(G) = \{\mathbf{C}(\mu, v) \mid (\mu, v) \in \mathbf{DEF}_1\}$, but this is not the desired indexing, since some cicc's may occur several times. The following lemma will allow us to restrict the set \mathbf{DEF}_1 appropriately.

5.3. Lemma. *If (μ, v) and $(\mu', v') \in \mathbf{DEF}_1$, $i = \mathbf{lab}(\mu)$, then $\mathbf{C}(\mu, v) = \mathbf{C}(\mu', v')$ iff $\mu = \mu'$, and v and v' are strongly linked in H_i .*

Proof. “If”. By the definitions, a strong internal path in H_i , where $i = \mathbf{lab}(\mu)$, linking v to v' yields an internal path in G/μ (since this graph is isomorphic to $H_i[G_1, \dots, G_n]$), that links $f((\mu, v))$ to $f((\mu, v'))$. Hence $\mathbf{C}(\mu, v) = \mathbf{C}(\mu, v')$.

The other direction is similar. \square

Being strongly linked is an equivalence relation on the set of internal vertices. Let us select a vertex in each class. The selected vertices will be called the *canonical internal vertices*.

We define \mathbf{DEF}'_1 as the set of pairs (μ, v) in \mathbf{DEF}_1 such that v is canonical, and we let $\mathbf{DEF}' := \mathbf{DEF}_0 \cup \mathbf{DEF}'_1$. Hence, every cicc of G is $\mathbf{C}(\mu, x)$ for one and only one pair (μ, x) in \mathbf{DEF}' .

The following lemma is an easy consequence of the above definitions and lemmas.

5.4. Lemma. *For every item (μ, x) of K , one can effectively decide whether:*

(1) $f((\mu, x))$ is a source of G , (and one can determine its rank if this is the case), or:

(2) $f((\mu, x))$ is an internal item of G , and in this case, one can determine the pair w in \mathbf{DEF}' such that $\mathbf{C}(w)$ is the cicc of G containing it.

Second part: Constructing a source-preserving decomposition of G over its cicc's

Let us say that two cicc's of G are equivalent, written $C \equiv C'$, if they are isomorphic and have the same set of sources (if this is the case, they have the same sequence of sources).

We define the *type* of a pair (μ, e) in \mathbf{DEF}_0 as (i, R) , where $i = \mathbf{lab}(\mu)$ and R is the set of pairs (v, w) such that v is a vertex of e (v is also a source of H_i), $w = f((\mu, v))$ and belongs to \mathbf{S}_G .

We define the *type* of a pair (μ, v) in \mathbf{DEF}_1 , as the pair (i, R) where $i = \mathbf{lab}(\mu)$, and R is the set of pairs (v', w) such that $v' \in \mathbf{SL}(i, v)$, $f((\mu, v')) = w$, and $w \in \mathbf{S}_G$.

For every item x , and every (possible) type (i, R) , one can construct a regular language $L'(x, i, R)$, such that, for every node μ of T , (μ, x) is of type (i, R) iff $i = \mathbf{lab}(\mu)$ and $(\mathbf{rt}, \mu) \in \mathbf{suc}(\eta)$ for some η in $L'(x, i, R)$. To be precise, one takes for $L'(x, i, R)$ the intersection of the languages $L(i, v'_j, w_j)$, $j = 1, \dots, k$, where $\mathbf{SL}(i, v) = \{v'_1, \dots, v'_k\}$ and $R = \{(v'_j, w_j), j = 1, \dots, k\}$ in the case where x is an internal vertex of H_i . If x is an edge e , then v'_1, \dots, v'_k are its vertices. If R is not of the appropriate form, then $L'(x, i, R)$ is empty.

If $w = (\mu, x)$ and $w' = (\mu', x')$ belong to \mathbf{DEF} , we let $w \equiv w'$ iff $x = x'$, and w and w' are of the same type.

5.5. Lemma. *If w and w' belong to \mathbf{DEF}' , and $w \equiv w'$, then $\mathbf{C}(w) \equiv \mathbf{C}(w')$.*

Proof. The case of w and w' in \mathbf{DEF}_0 is clear.

Let $w = (\mu, v)$ and $w' = (\mu', v)$ belong to \mathbf{DEF}'_1 , and be of the same type. We first make a few observations concerning $\mathbf{C}(\mu, v)$.

(1) Every item x of $\mathbf{C}(\mu, v)$ is $f((\mu_1, y))$ for some $\mu_1 \leq \mu$, and some item y .

(2) If x , as in (1), is $f((\mu_1, y))$, and if μ_1 is not less than μ , then $\mu = \mu_1$, and x is a source, both of G and of $\mathbf{C}(\mu, v)$.

(3) If $f((\mu_1, x)) \in \mathbf{C}(\mu, v)$, $(\mu, \mu_1) \in \mathbf{suc}(\eta)$, and u'_1 is such that $(\mu, \mu'_1) \in \mathbf{suc}(\eta)$, then $f((\mu'_1, x)) \in \mathbf{C}(\mu, v)$. (Easy proof by induction on the length of η .)

Remark that the subtrees of T issued from μ and μ' , that we denote respectively by T_μ and $T_{\mu'}$, are isomorphic (since μ and μ' have the same label, and T is a syntactic tree). Let $h: T_\mu \rightarrow T_{\mu'}$ be an isomorphism (it preserves the labels of nodes and edges in an obvious way). Let \tilde{h} be the binary relation $\subseteq (\mathbf{V}_G \times \mathbf{V}_G) \cup (\mathbf{E}_G \times \mathbf{E}_G)$ defined by:

$$\tilde{h} = \{(f((\mu_1, x)), f((h(\mu_1), x))) \mid \mu_1 \leq \mu, \\ x \text{ is an item of the appropriate graph}\}.$$

It is not hard to prove that it is bijective. It is an isomorphism $\bar{h}: G/\mu \rightarrow G/\mu'$. If γ is an internal path in G linking a vertex w to $f((\mu, v))$, then γ is in $\mathbf{C}(\mu, v)$. Hence $w = f((\mu_1, v'))$. Its image under \bar{h} is a path γ' , linking $w' := f((h(\mu_1), v'))$ to $f((\mu', v))$. The path γ' is internal, because (μ, v) and (μ', v) are of the same type. Hence it belongs to $\mathbf{C}(\mu', v)$.

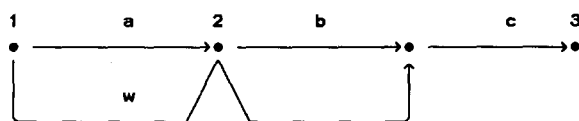
Hence \bar{h} is a bijection: $\mathbf{C}(\mu, v) \rightarrow \mathbf{C}(\mu', v)$. Since μ and μ' are of the same type, $\mathbf{C}(\mu, v)$ and $\mathbf{C}(\mu', v)$ have the same set of sources, namely, the set of vertices \bar{w} , such that $(\bar{v}, \bar{w}) \in R$ for some \bar{v} , where (i, R) is the common type of (μ, v) and (μ', v) . \square

We can construct a finite subset \mathbf{DEF}'' of \mathbf{DEF}' that has one and only one element in each class of \equiv . It is finite because there are finitely many types, whence finitely many equivalence classes. Its construction can be done effectively by means of the languages $L'(x, i, R)$. Furthermore, one can also determine, for every w in \mathbf{DEF}'' , the cardinality $\mathbf{c}(w)$ of its equivalence class in \mathbf{DEF}' w.r.t. \equiv .

Let us define $\mathbf{A}(w) = \{\mathbf{C}(w') \mid w' \in \mathbf{DEF}', w' \equiv w\}$. The family $(\mathbf{A}(w))$, for w in \mathbf{DEF}'' , is a partition of $\mathbf{CICC}(G)$, every set of which consists of isomorphic graphs, having the same sequence of sources. The cardinality of $\mathbf{A}(w)$ is $\mathbf{c}(w)$. From these data, and by Lemma 1.10, one can achieve the goal of this second part:

5.6. Lemma. *One can construct a source-preserving decomposition of G over $\{\mathbf{C}(w) \mid w \in \mathbf{DEF}''\}$.*

The graph G of Fig. 1, Example 1.11, is defined by the system reduced to the equation $w = K$, where K is the following graph:



This system is not special. The source-preserving decomposition of the graph G over its internally connected components is $f(//^\infty a, H)$, where f and H are defined in Example 1.11. The graph H is defined by a special system defined in Remarks 4.2.

For (μ, x) in \mathbf{DEF} , we denote by $\bar{\mathbf{C}}(\mu, x)$ the abstract graph, (i.e., the icc of G), isomorphic to $\mathbf{C}(\mu, x)$.

At this point of our proof, we do not know how to decide whether $\mathbf{C}(\mu, v)$ is isomorphic to $\mathbf{C}(\mu', v')$ for $(\mu, v), (\mu', v') \in \mathbf{DEF}'_1$. This is actually decidable, but only by Corollary 5.9 of the theorem we are now proving. Hence we do not have an enumeration of $\mathbf{ICC}(G)$. A same icc may appear several times in the indexed set $(\bar{\mathbf{C}}(\mu, x))_{(\mu, x) \in \mathbf{DEF}'}$.

Third part: Construction of a special system defining ICC(G)

5.7. Proposition. *One can construct a special system defining the graphs $\bar{\mathbf{C}}(\mu, x)$ for all $(\mu, x) \in \mathbf{DEF}''$.*

Proof. All what we did up to now, was relative to a fixed component $G = G_i$ of the solution of a given system S . In this proof, we shall consider all components simultaneously, hence we shall make this dependence explicit. The above introduced objects will be denoted by $\mathbf{DEF}(i)$, $\mathbf{DEF}'(i)$, $\mathbf{C}(i, \mu, x)$, etc.

For each i , each (μ, x) in $\mathbf{DEF}''(i)$, let us define a new symbol $[i, \mu, x]$ of type $\tau(\mathbf{C}(i, \mu, x))$. Let U'' be this new set of symbols.

From Lemma 5.6, one can construct, for each i in $[n]$, a finite graph K_i over $U'' \cup U''^\infty$, such that:

$$G_i = K_i[\mathbf{C}(i, \mu, x)/[i, \mu, x], (//^\infty \mathbf{C}(i, \mu, x))/[i, \mu, x]^\infty]$$

(The substitution is done simultaneously for all $[i, \mu, x]$ in U'' and all $[i, \mu, x]^\infty$ in U''^∞ . This is omitted in order to obtain a simpler formula. The same will be done below in similar cases.)

It is easy to observe that $\mathbf{C}(i, \mu, x)$ is isomorphic to $\mathbf{C}(j, \varepsilon, x)$, if (μ, x) belongs to $\mathbf{DEF}(i)$ and j is the label of the node μ of T_i . Hence, it is isomorphic to $\mathbf{C}(j, \varepsilon, x')$ for a unique x' such that (ε, x') belongs to $\mathbf{DEF}''(j)$. We let $\mathbf{rep}([i, \mu, x])$ be the new symbol $[j, x']$, and $\mathbf{rep}([i, \mu, x]^\infty)$ be similarly $[j, x']^\infty$. We let U' be the set of new symbols $[j, x]$, for (ε, x) in $\mathbf{DEF}(j)$, with $\tau([j, x]) := \tau(\mathbf{C}(j, \varepsilon, x))$. We let:

$$K'_i := K_i[\mathbf{rep}([i', \mu, x])/[i', \mu, x], \mathbf{rep}([i', \mu, x]^\infty)/[i', \mu, x]^\infty].$$

Hence, K'_i is a finite graph over $U' \cup U'^\infty$, and we have:

$$G_i = K'_i[\mathbf{C}(j, \varepsilon, x)/[j, x], (//^\infty \mathbf{C}(j, \varepsilon, x))/[j, x]^\infty]. \quad (1)$$

Our purpose is to construct a special system

$$S' = \langle [j, x] = H_{j,x}; [j, x] \in U' \rangle$$

defining $(\bar{\mathbf{C}}(j, \varepsilon, x))_{[j,x] \in U'}$.

If $(\varepsilon, x) \in \mathbf{DEF}_0(j)$, then we let simply $H_{j,x} := \bar{\mathbf{C}}(j, \varepsilon, x)$. (This graph is reduced to a single edge, all vertices of which are sources.)

Let us now consider (ε, x) in $\mathbf{DEF}''(j)$, where x is an internal vertex of H_j . Let then $L_j := H_j[K'_1/u_1, \dots, K'_n/u_n]$. This graph belongs to $\mathbf{FCG}(A \cup U' \cup U'^\infty)_{\tau(u_j)}$.

Let x' be the internal vertex of L_j to which the internal vertex x of H_j is mapped, in the substitution defining L_j (see Definition 1.4). We let $H_{j,x}$ be the cicc of L_j containing x' . This completes the definition of S' .

Claim 1. $\mathbf{C}(j, \varepsilon, x)$ is isomorphic to the graph:

$$H_{j,x}[\mathbf{C}(i', \varepsilon, x')/[i', x'], (//^\infty \mathbf{C}(i', \varepsilon, x'))/[i', x']^\infty].$$

Proof. We have

$$\begin{aligned}
G_j &= H_j[G_1, \dots, G_n] \\
&= H_j[K'_1[\mathbf{C}(i', \varepsilon, x')/[i', x'], \dots], \dots] \quad (\text{by (1)}) \\
&= L_j[\mathbf{C}(i', \varepsilon, x')/[i', x'], (//^\infty \mathbf{C}(i', \varepsilon, x'))/[i', x']^\infty] \quad (2)
\end{aligned}$$

by the definition of L_j and the associativity of graph substitutions.

Let us observe that, for all graphs G, L, C_1, \dots, C_m :

if $G = L[C_1/u_1, \dots, C_m/u_m]$,
if C_1, \dots, C_m are source-separated and connected,
if v is an internal vertex of L , mapped onto v' in the
substitution defining G , then
 v' is internal in G , and the cicc of G containing v' is isomorphic
to $L'[C_1/u_1, \dots, C_m/u_m]$, where L' is the cicc of L containing v .

In order to complete the proof, one applies this remark with $G = G_j$, $L = L_j$, $v = x'$, and the sequence C_1, \dots, C_m equal to an enumeration of the set of graphs $\mathbf{C}(i', \varepsilon, x')$, for $[i', x']$ in U' . It follows that $H_{j,x}[\dots]$ is the cicc of G_j containing the vertex corresponding to x' in the substitution of equation (2). This vertex is $f((\varepsilon, x))$, and the cicc containing it is $\mathbf{C}(j, \varepsilon, x)$. \square

Hence $(\bar{\mathbf{C}}(j, \varepsilon, x))_{[j,x] \in U'}$, is a solution of the extended system S' . The following claim completes the proof.

Claim 2. S' is a special system, and $(\bar{\mathbf{C}}(j, \varepsilon, x))_{[j,x] \in U'}$ is its solution.

Proof. Since the graphs G_j are source-separated, and by (1), the graphs K'_j are also source-separated. Since S is separated, the graphs L_i are also source-separated. The cicc's of the graphs G_1, \dots, G_n are source-separated too. Hence S' is separated. This establishes condition (S1) of Definition 4.1.

Since the extended system S' is separated, since $(\bar{\mathbf{C}}(j, \varepsilon, x))_{[j,x] \in U'}$ is a solution of S' consisting of source-separated and internally connected graphs, it is actually the solution of S , by Proposition 2.14.

Hence condition (S3) of Definition 4.1 holds. Condition (S2) also holds since the right-hand sides of S' are internally connected (by construction). If they have no internal vertex, then they are of the form $\bar{\mathbf{C}}(j, \varepsilon, e)$, for (ε, e) in $\mathbf{DEF}_0(j)$, hence, they are reduced to single terminal edges. Hence, S' is a special system, as was to be proved. \square

This completes the proof of Theorem 5.1. \square

The following corollary collects a few results, established in the course of the above proof, or easily derivable from it.

5.8. Corollary. (1) *An equational graph has finitely many internally connected components, and finitely many connected components.*

(2) *The internally connected components of a source-separated equational graph are equational. They are defined by a special system.*

Here is another corollary of Theorem 5.1.

5.9. Corollary. *The equality (i.e., the isomorphism) of two equational graphs is decidable.*

Proof. Let G be an equational graph. Let φ be the formula defining it. This formula can be constructed since the proof of Theorem 5.1 is effective. Let G' be another equational graph. We know from Courcelle [11, Corollary (9.3)], that it can be decided whether $G' \models \varphi$, i.e., whether G' is isomorphic to G . \square

The algorithm derived from this proof is actually intractable: the formula derived from Proposition 4.3, that characterizes an equational graph defined by a special system S , is of length $\mathbf{O}(\text{size}(S)^4)$. (The *size* of a system is the sum of the sizes of the graphs forming its equations; the *size* of a graph H of type n is $n + \text{Card}(\mathbf{V}_H) + \sum \{\tau(e) \mid e \in \mathbf{E}_H\}$.) But the transformation of an arbitrary system into a source-separated one is exponential (see the proof of Proposition 2.10 given in the Appendix). Hence the formula constructed in the proof of Theorem 5.1 to define an equational graph G is of exponential length in the size of the system S . It can be transformed (by the results of [11]) into a formula θ in $\exists\mathcal{WL}$ that defines the set of infinite graph expressions (that are infinite trees), the value of which is G . This formula can be converted into a “special” automaton \mathcal{A} (Rabin [19]). The size of this automaton can be expressed with a level of exponentiation proportional to the maximum number of nested quantifications in θ . The second graph G' (that we test for isomorphism with G) can be represented by an infinite graph expression that is a regular tree. One can decide whether this tree is accepted by the automaton \mathcal{A} , in polynomial time in the sizes of \mathcal{A} and of the system of equations defining G' . However, the automaton \mathcal{A} is of super-exponential size in that of the system defining G . (This intractability is mainly due to the construction of \mathcal{A} .) An efficient algorithm can perhaps be found by other techniques, like the ones presented in Courcelle [6].

5.10. Conjecture: *Every equational graph is \mathcal{WL} -definable.*

This conjecture holds for regular trees in the sense of Courcelle [7] (because, in a regular tree, the set of paths from the root to the nodes labelled by any given symbol form a regular language, and regular languages can be defined in \mathcal{WL} by the fundamental result of Büchi [5]). It holds also for the syntactic trees of equational graphs, as we have seen in Section 3.

It suffices to establish it for the equational graphs defined by special systems. The proof of Theorem 5.1 works then with \mathcal{WL} instead of $\exists\mathcal{WL}$, since, by Proposition 3.8, the \mathcal{WL} -definability of graphs is preserved by source-preserving compositions.

6. The relative definability of equational graphs

The connected component H of the first source of a graph G is definable in this graph. If furthermore G is equational, then H is also equational. This fact is a by-product of the proof of Theorem 5.1, (see Corollary 5.8). It suggests the following generalization:

6.1. Theorem. *Every graph definable in an equational graph is equational.*

The proof of this proposition will use a few lemmas.

We let T denote the unique infinite complete binary tree, all nodes of which are labeled by 0 (equivalently, have no label). As in Section 3, we denote by $\mathcal{T}(A)$ the set of infinite complete binary trees, the nodes of which are labeled in A .

6.2. Lemma. *Let \simeq be an equivalence relation on T defined by a monadic second-order formula $\varphi(x, y)$. There exists an integer k such that \simeq is generated by its restriction the pairs of vertices at distance at most k .*

Proof. A tree t in $\mathcal{T}(\{0, 1\})$ is completely defined by the total mapping $\|t\|: \{l, r\}^* \rightarrow \{0, 1\}$ that defines the label of any node. (The nodes are words on the two letters l and r ; the left successor of u is ul , its right successor is ur .)

We let the *distance* of two words w and w' in $\{l, r\}^*$ be their distance in T , considered as a graph. In other words, $d(w, w') = |v| + |v'|$ where $w = uv$, $w' = uv'$ and u is the longest common prefix of w and w' .

For every tree t in $\mathcal{T}(\{0, 1\})$ we let:

$$\delta(t) = \mathbf{Min}(\{|x| \mid x \in \{l, r\}^+, \|t\|(x) = 1\}).$$

If x, y are two nodes of t with label 1, we let $\mu(t, x, y)$ be the least integer k such that there is a sequence of nodes $x_0 = x, x_1, \dots, x_m = y$, all of them with label 1, and such that $d(x_i, x_{i+1}) \leq k$ for all $i = 0, \dots, m - 1$. (We have $\mu(t, x, y) = 0$ if $x = y$.) We also let:

$$\mu(t) = \mathbf{Max}\{\mu(t, x, y) \mid x, y \in \mathbf{V}_T, \|t\|(x) = \|t\|(y) = 1\}$$

(We may have $\delta(t) = \infty$ and $\mu(t) = \infty$.)

We let L be the set of trees t in $\mathcal{T}(\{0, 1\})$ such that $\|t\|^{-1}(1)$ is an equivalence class of the relation \simeq . In order to prove the lemma, we need only prove that $\mathbf{Max}\{\mu(t) \mid t \in L\}$ is finite.

Claim 1. *If t is regular, then $\mu(t)$ is finite.*

Proof. Let t be a regular tree in $\mathcal{T}(\{0, 1\})$ and let $\{t_0, \dots, t_m\}$ be the set of its subtrees. Let $k = \mathbf{Max}\{\delta(t_i) \mid i = 0, \dots, m, \delta(t_i) < \infty\}$ (and $k = 0$ if $\delta(t_i) = \infty$ for all i).

We shall establish that $\mu(t) \leq 2k + 1$.

Let x, y be two nodes of t with label 1. Let $y_0 = x, y_1, y_2, \dots, y_n = y$ be the sequence of nodes of the unique loop-free path in t from x to y . The nodes y_1, \dots, y_{n-1} are ancestors of either x or y or both. Hence, the subtrees $t/y_1, \dots, t/y_{n-1}$ have at least one vertex labeled by 1. Let x_i be one such vertex of t/y_i of minimal depth (i.e., of minimal length as a word in $\{l, r\}^*$); this depth is at most k). Let us now consider the sequence of vertices of t :

$$y_0 = x, y_1x_1, y_2x_2, \dots, y_{n-1}x_{n-1}, y_n = y.$$

We have:

$$d(x, y_1x_1) \leq 1 + |x_1| \leq k + 1,$$

$$d(y_ix_i, y_{i+1}x_{i+1}) \leq 1 + |x_i| + |x_{i+1}| \leq 2k + 1,$$

$$d(y_{n-1}x_{n-1}, y) \leq 1 + |x_{n-1}| \leq k + 1.$$

Hence $\mu(t, x, y) \leq 2k + 1$. This completes the proof of the claim. \square

We now continue the proof of the lemma. Since the equivalence relation \approx is definable, the set L is definable. Hence, by the main theorem of Rabin [18], it is also definable by a finite-state automaton $\mathcal{A} = \langle \{0, 1\}, Q, M, Q_0, Q_F \rangle$ with set of states Q , with set of initial states $Q_0 (\subseteq Q)$, transition table M , and set of sets of accepting states Q_F . For every q in Q , we denote by $\mathbf{L}(q)$ the set of trees accepted by \mathcal{A} with initial (root) state q . It follows that $L = \bigcup \{\mathbf{L}(q) \mid q \in Q_0\}$. We let $Q' \subseteq Q$ be the set of states q such that $\mathbf{L}(q)$ is reduced to a single regular tree, denoted by $\mathbf{t}(q)$.

We let $\delta = \mathbf{Max}\{\delta(\mathbf{t}(q)) \mid q \in Q', \delta(\mathbf{t}(q)) < \infty\}$ and $\mu = \mathbf{Max}\{\mu(\mathbf{t}(q)) \mid q \in Q'\}$.

Claim 2. For every t in L , $\mu(t) < \mathbf{Max}\{\mu, 2\delta + 2\}$.

Proof. Let $t \in L$, let x and y be two vertices of t labeled by 1.

First case: x is an ancestor of y . Without loss of generality, we can assume that $y = xrx'$, with $x' \in \{l, r\}^*$. Let us consider an accepted run f of \mathcal{A} on t . (A run is defined as a mapping $f: \{l, r\}^* \rightarrow Q$ such that $f(\varepsilon) \in Q_0$ and f satisfies the local conditions specified by the transition relation M . It is *accepted* if on every infinite branch of t , the set of states occurring infinitely many times is in Q_F .)

Let $q = f(xr)$. Then $t/xr \in \mathbf{L}(q)$. Any other tree in $\mathbf{L}(q)$ can be substituted for t/xr in t (as subtree issued from xr), thus giving a tree t' in L . By the definition of L , if two trees t and t' in L are such that $\|t\|(x) = \|t'\|(x) = 1$ for some x in $\{l, r\}^*$, then $t = t'$. It follows that $\mathbf{L}(q)$ is reduced to a single tree. Since every nonempty set of the form $\mathbf{L}(q)$ contains a regular tree by a result of Rabin [20], t/xr is regular and $q \in Q'$.

Let x'' be a node in t/xr with label 1 that is as close as possible to the root. We get:

$$\mu(t, x, y) \leq \mathbf{Max}\{d(x, xrx''), \mu(t, xrx'', y)\} \leq \mathbf{Max}\{1 + \delta, \mu\}.$$

since $|x''| \leq \delta$ and $\mu(t, xrx'', y) \leq \mu(t/xr, x'', y) \leq \mu$.

Second case: $x = ulx'$, $y = ury'$ for some u, x', y' in $\{l, r\}^*$. If u has label 1, then we get

$$\mu(t, x, y) \leq \mathbf{Max}\{\mu(t, x, u), \mu(t, u, y)\} \leq \mathbf{Max}\{1 + \delta, \mu\}$$

by the first case.

Let us now assume that u has label 0. Let f be a run of \mathcal{A} on t . Let $q = f(ul)$ and $q' = f(ur)$. As in the first case we obtain that q and q' belong to Q' . Letting x'' and y'' be two shortest words such that ulx'' and ury'' have both the label 1, it follows that:

$$\begin{aligned} \mu(t, x, y) &\leq \mathbf{Max}\{\mu(\mathbf{t}(q), x', x''), \mu(\mathbf{t}(q'), y', y''), d(ulx'', ury'')\} \\ &\leq \mathbf{Max}\{\mu, 2\delta + 2\}. \end{aligned}$$

since x'' and y'' are of length at most δ . This completes the proof of Claim 2 and of Lemma 6.2. \square

In the following lemma, we shall use the notion of tree-width of a graph, borrowed from Robertson and Seymour [21].

6.3. Definition (Tree-width). Let G be a graph. A *tree-decomposition* of G is a pair (U, f) consisting of an undirected tree U , and a mapping $f: \mathbf{V}_U \rightarrow \mathcal{P}(\mathbf{V}_G)$ such that:

- (1) $\mathbf{V}_G = \bigcup \{f(i) \mid i \in \mathbf{V}_U\}$,
- (2) every edge of G has all its vertices in $f(i)$ for some i ,
- (3) if $i, j, k \in \mathbf{V}_U$, and if j is on the unique loop-free path in U from i to k , then $f(i) \cap f(k) \subseteq f(j)$,
- (4) all sources of G are in $f(i)$ for some i in \mathbf{V}_U .

The *width* of such a decomposition is defined as

$$\mathbf{Max}\{\mathbf{Card}(f(i)) \mid i \in \mathbf{V}_U\} - 1,$$

and the tree-width of G is the minimum width of a tree-decomposition of G . It is denoted by $\mathbf{twd}(G)$. (For a 0-graph, Condition (4) is satisfied in a trivial way.)

6.4. Lemma. *Let G be a graph in $\mathbf{CG}(A)_n$ such that $\mathbf{V}_G \subseteq \{l, r\}^*$, and such that any two sources and any two vertices belonging to a same edge are at distance at most k . The tree-width of G is at most $2^{k+1} - 2$.*

Proof. We let (T, f) be the tree decomposition of G such that T is the unique tree in $\mathcal{I}(\{0\})$, (as in Lemma 6.2), and $f(w) = \mathbf{V}_G \cap \{wu \mid u \in \{l, r\}^*, |u| \leq k\}$, for every node w of T .

We first verify that (T, f) is actually a tree-decomposition. Each vertex of G belongs to some set $f(w)$. This ensures that Condition (1) of Definition 6.3 holds. Consider any two vertices x and y of G such that $d(x, y) \leq k$. We have $x = wx'$, and $y = wy'$, with $|x'|, |y'| \leq k$. It follows that x, y belong both to $f(w)$. This remark establishes Conditions (2) and (4). Condition (3) is easy to verify.

The cardinality of $f(w)$ is at most $1 + 2 + \dots \cdot 2^k = 2^{k+1} - 1$. Hence, $\mathbf{twd}(G) \leq 2^{k+1} - 2$. \square

Proof of Theorem 6.1. We first prove that if a graph G is definable in the infinite binary complete tree T , then it is equational.

Let $G = \mathbf{def}_\Delta(T)$ for some definition scheme Δ with parameters X_1, \dots, X_n , and domain formula φ . This means that:

$$T \models \exists X_1, \dots, X_n \varphi$$

and that, for every n -tuple (X_1, \dots, X_n) of subsets of \mathbf{V}_T satisfying φ , the defined graph $\mathbf{def}_\Delta(T, X_1, \dots, X_n)$ is isomorphic to G .

We shall transform Δ by several steps. We first eliminate parameters. By the result of Rabin [20] establishing that every nonempty definable set of infinite complete binary trees contains a regular tree, it follows that, if there exists a n -tuple (X_1, \dots, X_n) satisfying φ , there exists one, (X_1^0, \dots, X_n^0) , corresponding to a regular tree in the construction of [18]. This n -tuple is definable. It follows that Δ can be transformed into a parameterless definition scheme Δ' such that:

$$\mathbf{def}_{\Delta'}(T) = \mathbf{def}_\Delta(T, X_1^0, \dots, X_n^0) = G.$$

For simplicity, we shall denote Δ' by Δ . This definition scheme is of the form $(\varphi, \psi_1, \dots, \psi_k, \theta, \dots)$, so that:

$$\mathbf{V}_G \cup \mathbf{E}_G \subseteq \{l, r\}^* \times [k].$$

Our next aim is to reduce k to 1. The integers in $[k]$ can be encoded as words in $\{l, r\}^m$ where $k \leq 2^m$. Via this encoding, we get:

$$\mathbf{V}_G \cup \mathbf{E}_G \subseteq \{l, r\}^* \cdot \{l, r\}^m \subseteq \{l, r\}^*.$$

It follows that Δ can be modified into a definition scheme without parameters, also denoted by Δ , and such that $G = \mathbf{def}_\Delta(T)$, and $\mathbf{V}_G \cup \mathbf{E}_G \subseteq \mathbf{V}_T = \{l, r\}^*$.

We make a last technical assumption concerning Δ . We let r_A be the maximum type of a label in A and M be such that $r_A \leq 2^M$. We assume that, if e is a word in $\{l, r\}^*$ representing an edge of G , then no word of the form ew with $w \in \{l, r\}^*$, $1 \leq |w| \leq M$ is in $\mathbf{V}_G \cup \mathbf{E}_G$. Some extra transformation of Δ may be necessary to ensure this last condition.

Our objective is to construct an equational graph G' over $A \cup \{\$, \clubsuit\}$, where $\$$ and \clubsuit are two new edge labels of type 2 such that:

$$G = G'[2/\$, \sigma_{1,1}(\mathbf{1})/\clubsuit]. \quad (1)$$

The effect of this substitution is to delete the edges labeled by $\$$ (while keeping their vertices), and to contract the edges labeled by \clubsuit . (In a contraction, an edge disappears and its two vertices are fused into a single one.)

The construction of G' will achieve the proof because, from a system of equations S' defining G' , one obtains a system defining G by doing the substitutions of $\mathbf{2}$ for $\$$ and $\sigma_{1,1}(\mathbf{1})$ for \clubsuit on the righthand side of all the equations of S' .

The following technical fact is easy to establish.

Claim. *For every nonempty subset V of \mathbf{V}_T , defined as $V = \{x \mid T \models \mu(x)\}$, where μ is a given monadic second-order formula, one can construct two monadic second-order formulas, defining two sets E and E' forming a partition of the set of edges of T , such that in the simultaneous substitution in T of $\mathbf{2}$ for all edges e in E , and of $\sigma_{1,1}(\mathbf{1})$ for all edges e' in E' , every node of T gets identified with one and only one node in V , and no two nodes of V get identified.*

We now start the proof itself. We first assume that no symbol in A is of type 0. We let w_1, \dots, w_{r_A} be a fixed sequence of pairwise distinct words in $\{l, r\}^*$, with $|w_i| = M$. We let \approx be the equivalence relation on \mathbf{V}_T generated by the set of pairs of the form (ew_i, v) such that $e \in \mathbf{E}_G$, $v \in \mathbf{V}_G$, and $\mathbf{vert}_G(e, i) = v$ for $i \in [\tau(e)]$. (Recall that e, w_1, v are words in $\{l, r\}^*$.) Since G is $\mathbf{def}_\Delta(T)$, it follows that the equivalence relation \approx is definable in T . Let k be the integer associated with it by Lemma 6.2.

We let G' be the graph such that $\mathbf{V}_{G'} = \mathbf{V}_T = \{l, r\}^*$, $\mathbf{src}_{G'} = \mathbf{src}_G$ (recall that $\mathbf{V}_G \subseteq \mathbf{V}_T$), and having the following edges:

- an edge with label $\mathbf{lab}_G(e)$ and sequence of vertices (ew_1, \dots, ew_m) where $m = \tau(e)$, for every edge e of G ,
- an edge from v to v' with label \clubsuit , for every two vertices v and v' of T such that $v \neq v'$, $v \approx v'$ and $d(v, v') \leq k$,
- the edges of T labeled by $\$$ or \clubsuit according to whether they belong to E or E' , where (E, E') is the partition associated with \mathbf{V}_G by the above claim.

This graph G' satisfies equation (1). (This follows in particular from Lemma 6.2, ensuring that all vertices of an equivalence class of \approx get identified into a single vertex of G .)

Since T is definable, since \mathbf{V}_G is a definable subset of \mathbf{V}_T , and since the edges of G' are established in a definable way, it follows that G' is definable.

By Lemma 6.3, the graph G' is of finite tree-width. Hence G' is of width at most k' , for some k' , by Proposition 2.6 of [12]. It is proved in Courcelle [11, Proposition 9.5] that if a graph is definable and of finite width, then it is equational. Hence G' is equational and this concludes the proof that G is equational, if it is definable in the tree T , in the case where A has no symbol of type 0. If it has, then one can determine the labels of G of type 0 (that is the cardinality of the set of edges with label a for each a in A of type 0). One obtains that $G[\mathbf{0}/a; a \in A, \tau(a) = \mathbf{0}]$ is equational. It follows easily that G is equational.

Let us now assume that G is definable in an equational graph H . The construction of Theorem 8.1 of Courcelle [11] shows actually that H is definable in the infinite complete k -ary tree for some suitable k . Since this tree is itself definable in T , and since relative definability is a transitive relation (by Proposition 3.4, it follows that G is definable in T . Hence, G is equational by the first part of the proof. \square

As a conclusion, we collect the equivalent characterizations of equation graphs we have obtained in the following theorem.

6.5. Theorem. *The following properties of a graph are equivalent:*

- (1) G is equational,
- (2) G is definable and has a finite width (or tree-width),
- (3) G is $\exists\mathcal{WL}$ -definable and has a finite width (or tree-width),
- (4) G is definable in an equational graph,
- (5) G is definable in the infinite complete binary tree T .

Proof. That G has a finite width iff it has a finite tree-width follows from Courcelle [12, Proposition 2.6].

(3) \Rightarrow (2) is trivial.

(2) \Rightarrow (1) is proved in Courcelle [11, Proposition 9.5].

(1) \Rightarrow (3). The $\exists\mathcal{WL}$ -definability is proved in Theorem 5.1; the finiteness of the width is proved in Courcelle [11, Section (5.8)].

The equivalences (1) \Leftrightarrow (5) \Leftrightarrow (4) are proved in the proof of Theorem 6.1. \square

Appendix

This appendix contains the proofs of a few technical propositions, that have been stated without proofs in Section 2.

We first review a few definitions from Adamek and Koubek [1] and Bauderon [2, 3].

A.1. Definition (*The least fixed-point of a functor*). Let K be a category, let $F: K \rightarrow K$ be a functor. A *fixed point* of F is a pair (X, h) where X is an object of K , and h is an isomorphism: $FX \rightarrow X$.

A *least fixed point* of F is a fixed point (X_0, h_0) of F such that, for every fixed point (X, h) of F , there is a unique morphism $f: X_0 \rightarrow X$ making the following diagram commutative:

$$\begin{array}{ccc} FX_0 & \xrightarrow{h_0} & X_0 \\ Ff \downarrow & & \downarrow f \\ FX & \xrightarrow{h} & X \end{array}$$

If F has a least fixed point, then it is unique, up to isomorphism. We shall also call it the *initial solution* of the equation $X = FX$.

The following lemma is a special case of Proposition 5 of Ademek and Koubek [1].

A.2. Lemma. *Let K be a category having an initial object $\mathbb{1}$. Let $F:K \rightarrow K$ be a functor. Let w be the unique morphism $\mathbb{1} \rightarrow F\mathbb{1}$. If the diagram.*

$$\mathbb{1} \xrightarrow{w} F\mathbb{1} \xrightarrow{Fw} F^2\mathbb{1} \xrightarrow{F^2w} \dots \longrightarrow F^n\mathbb{1} \xrightarrow{F^nw} \dots$$

has a colimit X , and if the canonical morphism $h:X \rightarrow FX$ is an isomorphism, then (X, h^{-1}) is the least fixed point of F .

The canonical morphism h is defined as follows. Since X is the colimit, one has morphisms $w_n:F^n\mathbb{1} \rightarrow X$, such that $w_n = F^m w \cdot w_{n+1}$. Hence, one has a family of morphisms $Fw_n:F^{n+1}\mathbb{1} \rightarrow FX$, such that $Fw_n = F^{n+1}w \cdot Fw_{n+1}$, and a unique morphism $h:X \rightarrow FX$ by the universal property of the colimit.

We now recall from Bauderon [2, 3], how these definitions can be applied to systems of graph equations.

A.3. Definition (*The initial solution of a system of graph equations*). Let $S = \langle u_1 = H_1, \dots, u_n = H_n \rangle$ be a system of graph equations of the general form of Definition 2.1. Here we assume that H_1, \dots, H_n are pairwise disjoint concrete graphs. We let $p_i = \tau(u_i) = \tau(H_i)$ for all i .

We associate with S , a category K_S , and a functor $F_S:K_S \rightarrow K_S$.

The objects of K_S are n -tuples of concrete graphs (K_1, \dots, K_n) in $\mathbf{CG}(A)_{p_1} \times \dots \times \mathbf{CG}(A)_{p_n}$.

A morphism $h:(K_1, \dots, K_n) \rightarrow (K'_1, \dots, K'_n)$ is an n -tuple $h = (h_1, \dots, h_n)$ of graph homomorphisms where $h_i:K_i \rightarrow K'_i$ for all i , and each mapping $h_{i\mathbb{k}}:\mathbf{E}_{K_i} \rightarrow \mathbf{E}_{K'_i}$ is injective. We shall say that the homomorphisms h_1, \dots, h_n are *edge-injective*. This category has an initial object, denoted by $\mathbb{1}$, equal to the n -tuple of graphs $(\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n)$.

We now define a functor $F_S:K_S \rightarrow K_S$ as follows:

- $F_S(K_1, \dots, K_n) = (K'_1, \dots, K'_n)$, where $K'_i = H_i[K_1/u_1, \dots, K_n/u_n]$ for all i ,
- if $h = (h_1, \dots, h_n):(K_1, \dots, K_n) \rightarrow (K'_1, \dots, K'_n)$ then $F_S(h)$ is the morphism (h'_1, \dots, h'_n) where $h'_i = H_i[h_1, \dots, h_n]$ for all i . (See Definition 1.4.)

It is proved in Bauderon [2, 3] that K_S and F_S satisfy the conditions of Lemma A.2. Hence F_S has a least fixed point, that is the initial solution of S .

Letting T_1, \dots, T_n be the syntactic trees associated with S by Definition 2.4, we have the following lemma, that is nothing but Proposition 2.5.

A.4. Lemma. *The least fixed point of F_S is isomorphic to $(\mathbf{eval}(T_1), \dots, \mathbf{eval}(T_n))$.*

Proof. Let T_i^j be the initial part of T_i , obtained by restricting T_i to its first j levels. Let K_i^j be the intermediate graph associated with T_i^j , as defined in Definition 2.4. Clearly $T_i^j \subseteq T_i^{j+1}$ and $K_i^j \subseteq K_i^{j+1}$.

From this inclusion, one obtains a homomorphism

$$h_i^j: \mathbf{eval}(T_i^j) \rightarrow \mathbf{eval}(T_i^{j+1}).$$

Let us now consider the diagram

$$\mathbb{1} \xrightarrow{w} F_S \mathbb{1} \xrightarrow{F_S w} \dots \longrightarrow F_S^m \mathbb{1} \xrightarrow{F_S^m w} \dots$$

For each $m \geq 0$, let $(G_1^m, \dots, G_n^m) = F_S^m \mathbb{1}$ and (G_1, \dots, G_n) be the colimit of this diagram. Let us denote $F_S^m w$ by (w_1^m, \dots, w_n^m) . Here, w_i^m is a homomorphism: $G_i^m \rightarrow G_i^{m+1}$. It is easy to verify that there are isomorphisms $g_i^m: \mathbf{eval}(T_i^m) \rightarrow G_i^m$ such that the diagrams

$$\begin{array}{ccc} \mathbf{eval}(T_i^m) & \xrightarrow{h_i^m} & \mathbf{eval}(T_i^{m+1}) \\ s_i \downarrow & & \downarrow s_i^{m+1} \\ G_i^m & \xrightarrow{w_i^m} & G_i^{m+1} \end{array}$$

all commute.

Hence, to establish the lemma, it suffices to establish that $\mathbf{eval}(T_i)$ is the colimit of the diagram:

$$\mathbf{eval}(T_i^1) \xrightarrow{h_i^1} \mathbf{eval}(T_i^2) \xrightarrow{h_i^2} \dots \longrightarrow \mathbf{eval}(T_i^j) \xrightarrow{h_i^j} \dots$$

But $\mathbf{eval}(T_i^j)$ is a quotient of the intermediate graph K_i^j , and h_i^j is the canonical homomorphism derived from the inclusion $K_i^j \subseteq K_i^{j+1}$. The situation is fully described by the following lemma:

A.5. Lemma. Let K be a concrete graph, let (K^i) , $i \geq 0$, be an increasing sequence of concrete subgraphs of K . Let \approx be an equivalence relation on \mathbf{V}_K , and let \approx_i be an equivalence relation on \mathbf{V}_{K^i} for all i . Let us assume that $\approx = \bigcup \{\approx_i \mid i \geq 0\}$ and that $\approx_i \subseteq \approx_{i+1}$ for all i . Let $G^i = K^i / \approx_i$, let f^i and h^i be the canonical morphisms making commutative all diagrams of the form:

$$\begin{array}{ccc} K^i & \subseteq & K^{i+1} \\ f^i \downarrow & & \downarrow f^{i+1} \\ G^i & \xrightarrow{h^i} & G^{i+1} \end{array}$$

Then, K / \approx is the colimit of the diagram:

$$G^0 \xrightarrow{h^0} G^1 \xrightarrow{h^1} G^2 \longrightarrow \dots \longrightarrow G^i \xrightarrow{h^i} \dots$$

The proof of this lemma is a routine verification that we omit.

The desired result follows from the lemma, if we take $K = K_i$, $K^j = K_i^j$, $G^j = \mathbf{eval}(T_i^j)$ and $h^j = h_i^j$. \square

A.6. Proof of Proposition 2.10. Given a system $S = \langle u_1 = H_1, \dots, u_n = H_n \rangle$, with solution (G_1, \dots, G_n) , we shall construct a separated system defining the graphs $\mathbf{sep}(G_1), \dots, \mathbf{sep}(G_n)$.

We need a few definitions.

For every n -graph G , we let

$$\Delta(G) := \{(i, j) \in [n]^2 \mid \mathbf{src}_G(i) = \mathbf{succ}_G(j)\}.$$

We also let, for every $e \in \mathbf{E}_G$,

$$\Delta(G, e) = \{(i, j) \in [\tau(e)]^2 \mid \mathbf{vert}_G(e, i) = \mathbf{vert}_G(e, j)\}.$$

If G is equational, then $\Delta(G)$ can be computed by the decidability result recalled in Theorem 3.6.

A special case. We first assume that S and its solution satisfy the following conditions:

(C1) $\Delta(H_i, e) = \Delta(H_j)$ for every nonterminal edge e of H_i with label u_j , for every i in $[n]$,

(C2) $\Delta(G_i) = \Delta(H_i)$ for all $i \in [n]$.

One can establish that (C2) is a consequence of (C1) (this is a generalization of Lemma 2.8), but this fact is not needed in this proof. Condition (C1) is decidable because it concerns finitely many finite graphs, and condition (C2) is decidable, since $\Delta(G)$ is computable for an equational graph G .

For each u_i , let us define a new unknown w_i , of type $\tau(\mathbf{sep}(G_i))$. Let U' be this set of unknowns. Let α_i and β_i be such that $G_i = \sigma_{\alpha_i}(\mathbf{sep}(G_i))$ and $\mathbf{sep}(G_i) = \sigma_{\beta_i}(G_i)$. These two mappings can be constructed from $\Delta(G_i)$. Let $S' = \langle w_1 = H'_1, \dots, w_n = H'_n \rangle$ where, for every i :

$$H'_i = \sigma_{\beta_i}(H_i[\sigma_{\alpha_i}(w_1)/u_1, \dots, \sigma_{\alpha_n}(w_n)/u_n]).$$

(Let us recall that w_i denotes the graph reduced to one edge labeled by w_i . See Definition 1.1.)

Claim 1. *The system S' is separated, and its solution is the n -tuple $(\mathbf{sep}(G_1), \dots, \mathbf{sep}(G_n))$.*

Proof. It follows from the construction that each graph H'_i is source-separated and U' -separated. Let (G'_1, \dots, G'_n) be the solution of S' . The syntactic trees associated with S and S' are the same. Let us denote them by T_i , $i = 1, \dots, n$. Since, for each i , $\mathbf{del}(H_i)^0 = \mathbf{del}(H'_i)^0$, the intermediate graphs K_1, \dots, K_n associated with T_1, \dots, T_n are the same for the two systems. So are the associated equivalence relations on the sets $\mathbf{V}_{K_1}, \dots, \mathbf{V}_{K_n}$.

Hence, $G_i'^0 = G_i^0$ for all $i \in [n]$. From the definition of S' , and Definition 2.4, we have $G_i' = \sigma_{\beta_i}(G_i) = \mathbf{sep}(G_i)$. \square

The general case. Let $S = \langle u_1 = H_1, \dots, u_n = H_n \rangle$ be a system with solution (G_1, \dots, G_n) as in the first case.

For every $i \in [n]$, let $\delta_i = \Delta(G_i) \in \mathbf{Eq}([p_i])$ where $p_i = \tau(u_i)$.

For every $\delta \in \mathbf{Eq}([p_i])$ such that $\delta_i \subseteq \delta$, let $[u_i, \delta]$ be a new unknown, of type $\tau(u_i)$. Let U' be the set of these unknowns.

We shall construct a system $S' = \langle [u_i, \delta] = H'_{i,\delta}; [u_i, \delta] \in U' \rangle$ with solution $(\theta_\delta(G_i))_{[u_i, \delta] \in U'}$.

In order to construct $H'_{i,\delta}$, we let \sim be the equivalence relation on \mathbf{V}_{H_i} , generated by the following sets of pairs of vertices:

$$\begin{aligned} &(\mathbf{src}_{H_i}(j), \mathbf{src}_{H_i}(j')) \quad \text{for all } (j, j') \in \delta, \\ &(\mathbf{vert}_{H_i}(e, j), \mathbf{vert}_{H_i}(e, j')) \quad \text{for all } (j, j') \in \Delta(G_k), \\ &\quad \text{all } k \in [n], \text{ all edges } e \text{ of } H_i, \text{ the label of which is } u_k. \end{aligned}$$

For every nonterminal edge e of H_i , let $\delta(e)$ be the equivalence relation on $[\tau(e)]$ such that:

$$(j, j') \in \delta(e) \Leftrightarrow \mathbf{vert}_{H_i}(e, j) \sim \mathbf{vert}_{H_i}(e, j').$$

Hence, $\delta(e) \supseteq \Delta(H_i, e) \cup \Delta(G_k)$ if $\mathbf{lab}(e) = u_k$. We now define

$$H'_{i,\delta} = \theta_\delta[\theta_{\delta(e)}([\mathbf{lab}(e), \delta(e)]) / e; e \in E]$$

where E is the set of nonterminal edges of H_i .

Claim 2. *The system S' satisfies conditions (C1) and (C2) of the first case, and its solution is $(\theta_\delta(G_i))_{[u_i, \delta] \in U'}$.*

Proof. Each graph $\mathbf{del}(H'_{i,\delta})$ is a quotient of $\mathbf{del}(H_i)$. Hence, the intermediate graph $K'_{i,\delta}$ associated with $[u_i, \delta]$ and S' , is a quotient of the intermediate graph K_i associated with u_i and S .

Hence we have a commutative diagram, with canonical surjective homomorphisms:

$$\begin{array}{ccc} K_i & \xrightarrow{k} & K'_{i,\delta} = K_i / \sim \\ f \downarrow & & \downarrow f' \\ G_i = K_i / \approx & \xrightarrow{g} & G'_{i,\delta} = K'_{i,\delta} / \approx' \end{array}$$

We wish to establish that $G'_{i,\delta} = \theta_\delta(G_i)$.

The homomorphism g is surjective since f' and k are surjective.

For every (j, j') in δ , $g(\mathbf{src}_{G_i}(j)) = g(\mathbf{src}_{G_i}(j'))$ since, by the definition of \sim , we have $k(\mathbf{src}_{G_i}(j)) = k(\mathbf{src}_{G_i}(j'))$.

Hence, there exist surjective homomorphisms g' and h making the following diagram commutative:

$$\begin{array}{ccc} G_i & \xrightarrow{g} & G'_{i,\delta} \\ g' \searrow & & \swarrow h \\ & \theta_\delta(G_i) & \end{array}$$

Let v, v' be vertices of K_i such that $f'(k(v)) = f'(k(v'))$. We shall prove that $g'(f(v)) = g'(f(v'))$, and it will follow that h is an isomorphism. Since f' and k are canonical homomorphisms associated with quotients, it suffices to consider pairs of vertices belonging to the sets that generate the corresponding equivalences.

Hence, we only consider the following cases:

First case: (v, v') belongs to the set generating \sim on K_i . The various possibilities are as follows.

(1) $v = (\mathbf{rt}, \mathbf{src}_{H_i}(j))$, $v' = (\mathbf{rt}, \mathbf{src}_{H_i}(j'))$ where \mathbf{rt} denotes, as in Proposition 4.3, the root of T_i , and $(j, j') \in \delta$.

(2) $v = (\mu', \mathbf{vert}_{H_i}(e, j))$, $v' = (\mu', \mathbf{vert}_{H_i}(e, j'))$ where μ' is a node of T_i , e is an edge of H_i , i' is the label of μ' , and $(j, j') \in \delta(e)$.

(3) $v = (\mu, \mathbf{src}_{H_i}(j))$, $v' = (\mu, \mathbf{src}_{H_i}(j'))$ and, for some μ', i', e as in case (2), the node μ of T_i is the successor of μ' corresponding to e , the label of e in H_i is $u_{i'}$ (hence the label of μ in T_i is i''), and $(j, j') \in \delta(e)$.

In case (1), one has $g'(f(v)) = g'(f(v'))$ by the definition of g' . In cases (2) and (3), one has $f(v) = f(v')$ by the definition of $\delta(e)$.

Second case: $(k(v), k(v'))$ belongs to the set generating the equivalence relation \approx' on $K'_{i,\delta}$. (See Definition 2.4.)

These exists v'' in K_i such that (v, v'') is as in the first case, and (v'', v') belongs to the set generating the equivalence relation \approx on K_i .

Hence $f(v'') = f(v')$. By the first case, $g(f(v)) = g(f(v''))$. It follows that $g(f(v)) = g(f(v'))$.

Hence, the solution of S' is $(\theta_\delta(G_i))_{[u_i, \delta] \in U'}$. Since $\delta \supseteq \Delta(G_i)$, we have $\Delta(\theta_\delta(G_i)) = \delta$, for all $[u_i, \delta] \in U'$, and that S' satisfies (C1) and (C2), follows from its construction. \square

The construction of Proposition 2.10 consists in the following steps. Given a system S , with solution (G_1, \dots, G_n) , one uses Claim 2 to construct a system S' , defining the graphs $\theta_\delta(G_i)$ for all $[u_i, \delta]$ in U' . This system satisfies conditions (C1) and (C2). The construction of Claim 1 allows to transform it into a system S'' defining the graphs $\mathbf{sep}(\theta_\delta(G_i))$ for all $[u_i, \delta]$ in U' .

Since for each i , $\mathbf{sep}(G_i) = \mathbf{sep}(\theta_\delta(G_i))$ where $\delta = \Delta(G_i)$, the system S'' defines the graphs $\mathbf{sep}(G_1), \dots, \mathbf{sep}(G_n)$, together with some other graphs. \square

Proof of Proposition 2.9. We shall prove that, if S is a separated system with initial solution (G_1, \dots, G_n) , if (K_1, \dots, K_n) is a solution of S , if each graph G_i has at least one internal item, and if each graph K_i is internally connected and source-separated, then G_i is isomorphic to K_i for all i .

The proof will be done formally in a simplified case. In the general case, it is no more difficult, except for the notations.

Let S be the system reduced to the single equation $\langle u = H \rangle$ where $\tau(u) = n$ and H is in $\mathbf{FCG}(A \cup \{u\})_n$. We let $\{e_1, \dots, e_r\}$ be an enumeration of the set of nonterminal edges of H .

Let $K \in \mathbf{CG}(A)_n$ be a solution of S , that is source-separated and internally connected. We have an isomorphism $k_0: H[K/e_1, \dots, K/e_r] \rightarrow K$. Let (k_1, \dots, k_r) be the r -tuple of homomorphisms, $k_i: K^0 \rightarrow H[K/e_1, \dots, K/e_r]$, associated with $H[K/e_1, \dots, K/e_r]$ by Definition 1.4.

Fig. 8 illustrates the situation.

Since S is separated, the homomorphisms k_1, \dots, k_r are injective.

Let G be the canonical solution of S . It is characterized as the colimit of the diagram:

$$\mathbf{n} = G_0 \xrightarrow{w_0} G_1 \xrightarrow{w_1} G_2 \xrightarrow{w_2} \dots \rightarrow G_i \xrightarrow{w_i} \dots$$

where $G_i = H[G_{i-1}/e_1, \dots, G_{i-1}/e_r]$, $w_i = H[w_{i-1}, \dots, w_{i-1}]$, and w_0 exists (uniquely) by initiality. Hence we have the following commutative diagram (we shall call it Γ):

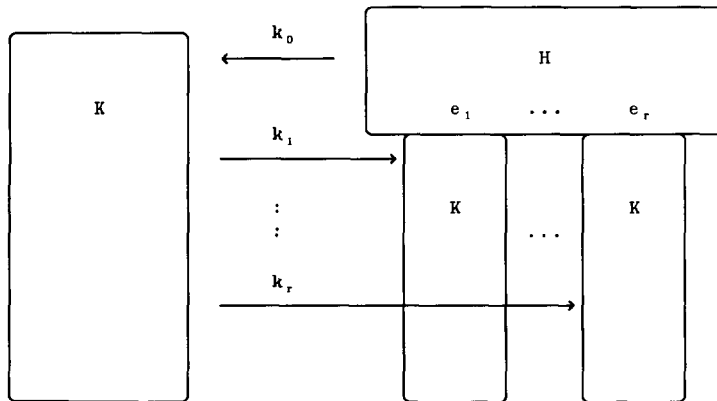
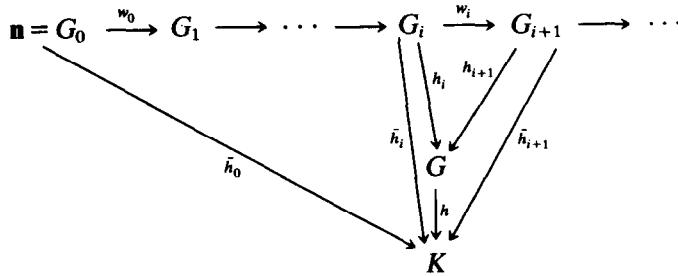


Fig. 8.

All morphisms of this diagram are injective. This follows from the following facts:

- (1) w_0 is injective since S is separated,
- (2) \tilde{h}_0 is injective since K is source-separated,
- (3) if g_1, \dots, g_r are injective morphisms, then $H[g_1, \dots, g_r]$ is injective: this follows from the hypothesis that S is separated.

Hence (1) and (3) give that the homomorphisms w_i are injective for all $i \geq 1$. So are the morphisms h_i . By (1) and (3), we obtain that the morphisms \tilde{h}_i are injective. It follows that h is injective too.

Letting $(g_{i,1}, \dots, g_{i,r})$ be the r -tuple of homomorphisms associated with $H[G_i, \dots, G_i]$, $(g_{i,j}: G_i \rightarrow H[G_i, \dots, G_i] = G_{i+1})$, we have the following commutative diagram (we shall call it $\Gamma_{i,j}$):

$$\begin{array}{ccccc}
 G_i & \xrightarrow{g_{i,j}} & H[G_i, \dots, G_i] = G_{i+1} & & \\
 \tilde{h}_i \downarrow & & \downarrow H[\tilde{h}_i, \dots, \tilde{h}_i] & \searrow \tilde{h}_{i+1} & \\
 K & \xrightarrow{k_j} & H[K, \dots, K] & \xrightarrow{k_0} & K
 \end{array}$$

A.7. Lemma. *Let $v \in h(\mathbf{V}_G)$. Let v_0, v_1, v_2, \dots be an infinite sequence such that for every j , there exists i such that $k_0(k_i(v_{j+1})) = v_j$. There exists m such that v_m is a source of K .*

Proof. As in Theorem 5.1, we let \mathbf{S}_K be the set of sources of K . Let v_0, v_1, v_2, \dots be as in the statement.

Since G is the colimit of the diagram Γ , there is an integer m such that $\tilde{h}_m(v) = v_0$ for some vertex v of G_m . Let m be the smallest such integer.

If $m = 0$, then $v_0 \in \mathbf{S}_K$ and we are done.

Otherwise, let us consider v_1 such that $k_0(k_i(v_1)) = v_0$. We claim that $\tilde{h}_{m-1}(v') = v_1$ for some v' in G_{m-1} .

Let us consider the diagram $\Gamma_{m-1,i}$.

The vertex v belongs to $H[G_{m-1}, \dots, G_{m-1}]$. Let v'' be its image in $H[K, \dots, K]$ under the homomorphism $H[\tilde{h}_{m-1}, \dots, \tilde{h}_{m-1}]$. Since $k_i^{-1}(v'')$ is defined (and equal to v_1), $g_{m-1,i}^{-1}(v)$ is defined and equal to a vertex v' in G_{m-1} .

By the commutativity of $\Gamma_{m-1,i}$, we have

$$v'' = k_i(\tilde{h}_{m-1}(v')) \quad \text{and} \quad k_0(v'') = v_0.$$

We have assumed that $k_0(k_i(v_1)) = v_0$.

Since the morphisms k_0, k_1, \dots, k_r are injective, we have $v_1 = \tilde{h}_{m-1}(v') \in h(\mathbf{V}_G)$. By repeating the argument, one has $v_m = \tilde{h}_0(v')$ for some $v' \in G_0 = \mathbf{n}$. Hence $v_m \in \mathbf{S}_K$, as was to be proved. \square

A few more technical definitions are needed. A path π in K is a $K \setminus G$ -path if the following conditions hold:

- (1) if $\mathbf{vert}(\pi) = (v, v_1, v_2, \dots, v_m, v')$, then v_1, \dots, v_m are in $\mathbf{V}_K - h(\mathbf{V}_G)$, and
- (2) the edges of π are all in $\mathbf{E}_K - h(\mathbf{E}_G)$.

For every $e \in \mathbf{E}_K - h(\mathbf{E}_G)$, we let $\mathbf{D}(e)$ be the set of edges e' of $\mathbf{E}_K - h(\mathbf{E}_G)$ such that there is a $K \setminus G$ -path linking v to v' , where v and v' belong to $\mathbf{V}_K - h(\mathbf{V}_G)$, v belongs to e , v' belongs to e' . We let $\mathbf{C}(e)$ be the 0-graph $K^0 \upharpoonright \mathbf{D}(e)$. We let $\mathbf{c}(e)$ be the number of vertices of $\mathbf{C}(e)$ that are sources of K .

For every subgraph C of K^0 , we denote by $\mathbf{Fr}(C)$ the set $\mathbf{V}_C \cap h(\mathbf{V}_G)$.

Note that h is a bijection $\mathbf{S}_G \rightarrow \mathbf{S}_K$.

A.8. Lemma. *Let $e \in \mathbf{E}_K - h(\mathbf{E}_G)$. There exist $i \in [r]$ and $e' \in \mathbf{E}_K - h(\mathbf{E}_G)$ such that $e = k_0(k_i(e'))$. Then, $\mathbf{C}(e)$ is isomorphic to $\mathbf{C}(e')$ by $k_0 \circ k_i$, and $\mathbf{c}(e) \leq \mathbf{c}(e')$.*

Proof. We first make an observation. If w is an item of K , that does not belong to $h(G)$, then there exists a unique i such that w is an item of $k_0(k_i(K))$, and a unique w' in K such that $w = k_0(k_i(w'))$. This item w' is not in $h(G)$, otherwise, w would be also in $h(G)$.

Since graph homomorphisms preserve the incidences (i.e., technically, the mapping \mathbf{vert}), if e_1 and e_2 are two edges in $\mathbf{E}_K - h(\mathbf{E}_G)$ having a common vertex v in $\mathbf{V}_K - h(\mathbf{V}_G)$, then the three integers i, j, m such that

$$e_1 = k_0(k_i(e'_1)), \quad e_2 = k_0(k_j(e'_2)), \quad v = k_0(k_m(v'))$$

are equal.

It follows that every item of $\mathbf{C}(e)$ belongs to $k_0(k_i(\mathbf{C}(e')))$, where (i, e') is the unique pair such that $e = k_0(k_i(e'))$.

It is not hard to see that, for every item w of $\mathbf{C}(e')$, its image under $k_0 \circ k_i$ is also an item of $\mathbf{C}(e)$.

Hence $\mathbf{C}(e)$ and $\mathbf{C}(e')$ are isomorphic by $k_0 \circ k_i$. Let v be a source of K , belonging to $\mathbf{C}(e)$. We have $v = \mathbf{vert}_K(\bar{e}, j)$ for some integer i , and some edge \bar{e} of $\mathbf{C}(e)$. Let $\bar{e}' = k_i^{-1}(k_0^{-1}(\bar{e}))$. Then $v = k_0(k_i(\mathbf{vert}_K(\bar{e}', j)))$. Hence $k_i(\mathbf{vert}_K(\bar{e}', j))$ is a source of $H[K/e_1, \dots, K/e_r]$ and $\mathbf{vert}_K(\bar{e}', j)$ must be a source of K , otherwise its image under k_i would be internal vertex. Hence

$$v' = k_i^{-1}(k_0^{-1}(v)) \in \mathbf{Fr}(\mathbf{C}(e')) \cap \mathbf{S}_K.$$

It follows that $\mathbf{Fr}(\mathbf{C}(e)) \cap \mathbf{S}_K$ is in bijection by $k_0 \circ k_i$ with a subset of $\mathbf{Fr}(\mathbf{C}(e')) \cap \mathbf{S}_K$. Hence $\mathbf{c}(e) \leq \mathbf{c}(e')$. \square

A.9. Lemma. *If the set $\mathbf{E}_K - h(\mathbf{E}_G)$ is not empty, it contains an element e such that $\mathbf{Fr}(\mathbf{C}(e)) \subseteq \mathbf{S}_K$.*

Proof. Let us assume that $\mathbf{E}_K - h(\mathbf{E}_G)$ is not empty. Let e in this set be such that $\mathbf{c}(e)$ is maximal. We claim that $\mathbf{Fr}(\mathbf{C}(e)) \subseteq \mathbf{S}_K$.

If this is not the case, let $v \in \mathbf{S}_K - \mathbf{Fr}(\mathbf{C}(e))$.

Let i_1, i_2, \dots be the sequence in $[r]$, and $e^{(0)}, e^{(1)}, e^{(2)}, \dots$ be the sequence of edges of $\mathbf{E}_K - h(\mathbf{E}_G)$ such that:

$$e^{(0)} = e,$$

$$e^{(m-1)} = k_0(k_{i_m}(e^{(m)})) \quad \text{for all } m \geq 1.$$

There exist such sequences by Lemma A.8. Let v_0, v_1, \dots , be such that

$$v_0 = v, \quad v_{m-1} = k_0(k_{i_m}(v_m)).$$

Each vertex v_m belongs to $\mathbf{C}(e^{(m)})$, and also to $h(G)$. Hence $v_m \in \mathbf{S}_K$ for some m . Let m_0 be the least such integer.

The sequence $(c(e^{(m)}))_{m \geq 0}$ is nondecreasing by Lemma A.8:

$$c(e^{(0)}) \leq c(e^{(1)}) \leq \dots \leq c(e^{(m)}) \leq \dots$$

but we have $c(e^{(m_0-1)}) < c(e^{(m_0)})$ since v_m is an element of $\mathbf{Fr}(\mathbf{C}(e^{(m_0)})) \cap \mathbf{S}_K$ that is *not* in correspondence by $k_0 \circ k_i$ (where $i = i_{m_0}$) with any element of $\mathbf{Fr}(\mathbf{C}(e^{(m_0-1)})) \cap \mathbf{S}_K$. This contradicts the maximality assumption on $c(e) = c(e^{(0)})$. Hence $\mathbf{Fr}(\mathbf{C}(e)) \subseteq \mathbf{S}_K$. \square

A.10. Proof of Proposition 2.9. Let K be a source-separated solution of S , that is internally connected. Let G be the initial solution of S , we assume that it has internal items. Let $h: G \rightarrow K$ be the morphism that exists by initiality.

We have observed above that h is injective. If $h_E: \mathbf{E}_G \rightarrow \mathbf{E}_K$ is not surjective, then there is, by Lemma A.9, an edge e in $\mathbf{E}_K - h_E(\mathbf{E}_G)$, such that $\mathbf{Fr}(\mathbf{C}(e)) \subseteq \mathbf{S}_K$.

The vertices of $\mathbf{C}(e)$ belonging to $h(G)$ are sources of K . Hence $K \upharpoonright \mathbf{C}(e)$ is a cicc of K , having no edge in common with $h(G)$. Hence, if G has internal items, K has at least two cicc's. Hence it is not internally connected.

Hence h_E is surjective. Since K is connected, it has no isolated vertex, and $h_V: \mathbf{V}_G \rightarrow \mathbf{V}_K$ is surjective. Hence h is an isomorphism, as was to be proved.

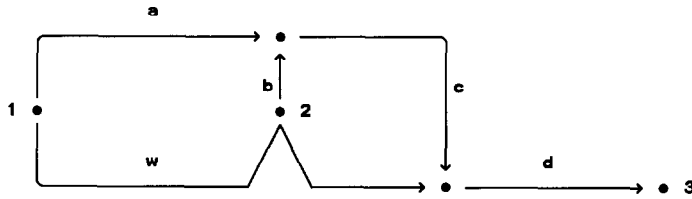
The proof is essentially the same if S has several equations.

Let us only observe that, in the corresponding extension of Lemma A.7 the vertices v_1, v_2, v_3, \dots do not belong necessarily to the same graph as v_0 . Similarly, in the lemma corresponding to Lemma A.8, the two isomorphic graphs $\mathbf{C}(e)$, and $\mathbf{C}(e')$, are not necessarily subgraphs of the same component of the considered solution of S . \square

A.11. Remarks. (The hypotheses of Proposition 2.9 are not superfluous.)

(1) The hypothesis that G contains internal items cannot be omitted. If u is an unknown of type 1, then, the initial solution of the equation $u = u$ is the graph **1**. This equation is separated. But a and b are two nonisomorphic source-separated internally connected solutions, if a and b are of type 1.

(2) The special system consisting of the equation $w = K$ where K is shown in Fig. 9 (the unknown w is of type 3) has an internally connected solution that is



Graph K

Fig. 9.

not source-separated, and not isomorphic to the initial one: namely $\theta_{1,2}(G)$ where G is the initial solution. \square

A.12. Proof of Proposition 2.15. An extended system can be considered as a system with right-hand sides of equations, having possibly infinitely many nonterminal edges. The definitions and results of A.1 to A.5 can be adapted to this case. The proof of Proposition 2.9 can be adapted too (the reader can note that a countable set can be used instead of the set $[r]$). We omit the formal details. \square

Acknowledgements

I thank Y. Gurevich and A. Salwicki for their interest in this work, and their comments about it, that have been the source of substantial improvements.

Note added in proof

G. Sénizergues (Bordeaux-1 University) has established Conjecture 5.10 for the context-free graphs of Muller and Schupp (“The theory of ends, pushdown automata, and second-order logic”, *Theoret. Comput. Sci.* 37 (1985) 51–75) that form a proper subclass of the class of equational graphs.

References

- [1] J. Adamek and V. Koubek, Least fixed point of a functor, *J. Comput. System Sci.* 19 (1979) 163–178.
- [2] M. Bauderon, On systems of equations defining infinite graphs, in: *Lecture Notes in Computer Science 344* (Springer, Berlin, 1989) 54–73.
- [3] M. Bauderon, Infinite hypergraphs, *Theoret. Comput. Sci.*, to appear.
- [4] M. Bauderon and B. Courcelle, Graph expressions and graph rewritings, *Math. Systems Theory* 20 (1987) 83–127.

- [5] J.R. Büchi, Weak second-order logic and finite automata, *Z. Math. Logik Grundlag. Math.* 5 (1960) 66–92.
- [6] B. Courcelle, An axiomatic approach to the Korenjak–Hopcroft algorithms, *Math. Systems Theory* 16 (1983) 191–231.
- [7] B. Courcelle, Fundamental properties of infinite trees, *Theoret. Comput. Sci.* 25 (1983) 95–169.
- [8] B. Courcelle, Equivalences and transformations of regular systems. Applications to recursive program schemes and grammars, *Theoret. Comput. Sci.* 42 (1986) 1–122.
- [9] B. Courcelle, An axiomatic definition of context-free rewriting and its application to NLC graph grammars, *Theoret. Comput. Sci.* 55 (1987) 141–181.
- [10] B. Courcelle, The monadic second-order logic of graphs I: Recognizable sets of finite graphs, *Information and Computation* 85 (1990) 12–75.
- [11] B. Courcelle, The monadic second-order logic of graphs II: Infinite graphs of bounded width, *Math. Systems Theory* 21 (1989) 187–221.
- [12] B. Courcelle, The monadic second-order logic of graphs III: Tree-width, forbidden minors, and complexity issues, Report 88-52, 1988, submitted.
- [13] B. Courcelle, The monadic second-order logic of graphs V: On closing the gap between definability and recognizability, *Theoret. Comput. Sci.*, to appear.
- [14] B. Courcelle, The monadic second-order logic of graphs: definable sets of finite graphs, in: *Lecture Notes in Computer Science* 344 (Springer, Berlin, 1989) 30–53.
- [15] B. Courcelle, Graph rewriting: an algebraic and logic approach, in: J. Van Leeuwen, ed., *Handbook of Theoretical Computer Science*, Vol. B (North-Holland, Amsterdam, 1990), Ch. 5.
- [16] Y. Gurevich, Monadic second-order theories, in: J. Barwise and S. Feferman, eds., *Model-theoretic Logic*, (Springer, Berlin, 1985) 479–506.
- [17] Y. Gurevich and L. Harrington, Trees, automatas and games, *Proc. of Symp. on Theory of Comput.*, San Fransisco, 1982, 60–65.
- [18] M. Rabin, Decidability of second-order theories and automata on infinite trees, *Trans. Amer. Math. Soc.* 141 (1969) 1–35.
- [19] M. Rabin, Weakly definable relations and special automata, in: Y. Bar-Hillel, ed., *Mathematical Logic and Foundations of Set Theory* (North-Holland, Amsterdam, 1972) 1–23.
- [20] M. Rabin, Automata on infinite objects and Church’s problem, *A.M.S. Regional Conf. Series in Math.* 13 (1972).
- [21] N. Robertson and P. Seymour, some new results on the well-quasi ordering of graphs, *Ann. Discrete Math.* 23 (North-Holland, Amsterdam, 1984) 343–354.
- [22] S. Shelah, The monadic theory of order, *Ann. of Math.* 102 (1975) 379–419.