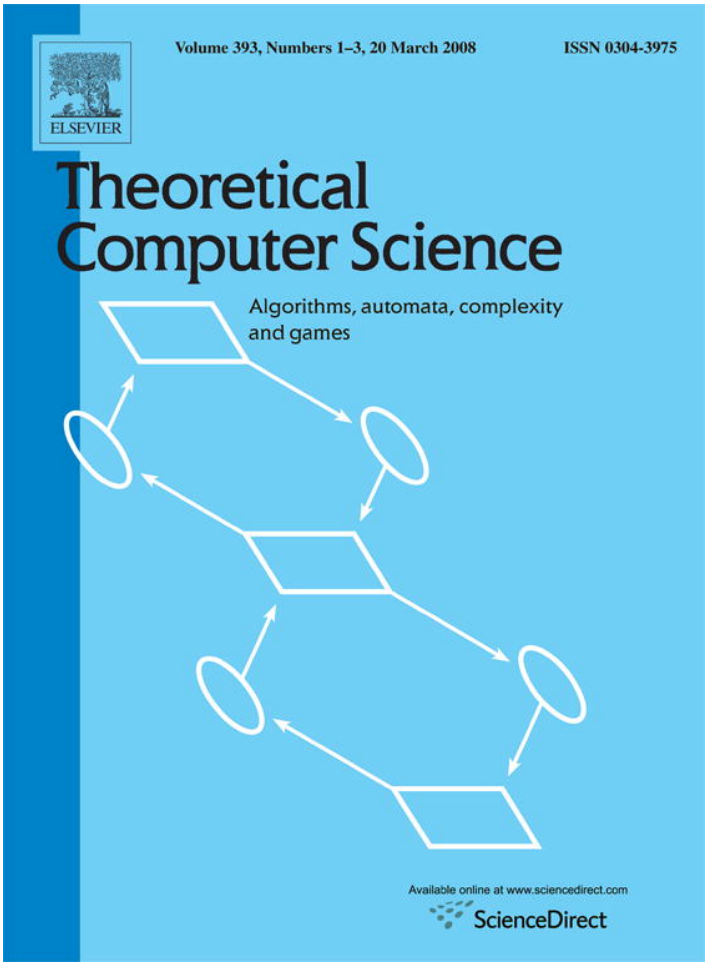
Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article was published in an Elsevier journal. The attached copy is furnished to the author for non-commercial research and education use, including for instruction at the author's institution, sharing with colleagues and providing to institution administration.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright

Author's personal copy

Available online at www.sciencedirect.com
ScienceDirect
Theoretical Computer Science 393 (2008) 147–165
Theoretical Computer Science
www.elsevier.com/locate/tcs

Graph automata

Symeon Bozapalidis, Antonios Kalampakas*

Aristotle University of Thessaloniki, Department of Mathematics, 54124, Thessaloniki, Greece

Received 5 July 2006; received in revised form 26 November 2007; accepted 30 November 2007

Communicated by Z. Esik

Abstract

Magmoids satisfying the 15 fundamental equations of graphs, namely graphoids, are introduced. Automata on directed hypergraphs are defined by virtue of a relational graphoid. The closure properties of the so-obtained class are investigated, and a comparison is being made with the class of syntactically recognizable graph languages.
© 2007 Elsevier B.V. All rights reserved.

Keywords: Automata; Directed hypergraphs; Magmoids

1. Introduction

Various models of graph generation have been introduced in the literature and have been investigated in many directions (for an overview see [9]). However a general theory of graph automata (i.e., automata with a graph as input) is still missing from formal graph language theory. Historically, Arbib and Give'on were the first who extended tree automata to operate on planar directed ordered acyclic graphs (cf. [1]).

In order to model derivations of 0-type grammars, Kamimura and Slutzki studied automata on rooted pdags, i.e., planar, directed, acyclic graphs (cf. [18,19]). The automata considered by Bossut, Dauchet and Warin, work exactly in the same way except that they use pdags with an unbounded number of roots and that they control the sequences of initial and final states by rational languages of state words (cf [5]). They have shown that the class of automaton definable languages of pdags is the smallest class containing finite sets of pdags and closed under union, (iterated) serial composition and (iterated) nondeterministic parallel composition.

On the other hand there is no systematic approach towards recognizing devices for general graph languages. Our aim, in this paper, is to extend the above automata on general graphs and to compare the corresponding class with the class of syntactically recognizable graph languages (cf. [7]).

To be able to construct graph automata we have to detect the category in which the graphs, over a specified alphabet constitute the free object. In the case of tree automata, the corresponding category is that of algebraic theories (cf. [4]). Recognizability of tree languages in locally finite algebraic theories was considered in [13]. At the level of trees, recognizability defined through congruences is equivalent to the recognizability defined via automata.

* Corresponding author. Tel.: +30 2310214088.
E-mail addresses: bozapali@math.auth.gr (S. Bozapalidis), akalamp@math.auth.gr (A. Kalampakas).

0304-3975/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.tcs.2007.11.022

Author's personal copy

More precisely, let Γ be a finite ranked alphabet, $X_n = \{x_1, \ldots, x_n\}$, $(n \geq 0)$ and denote by $T_\Gamma(X_n)$ the set of all Γ-trees over X_n. Then $T_\Gamma(X) = (T^n_\Gamma(X_m))_{m,n\geq 0}$, with tree substitution as operation, is the free algebraic theory over Γ. A tree language $F \subseteq T_\Gamma(X)$ is recognizable if there is a locally finite theory M and a theory morphism

$$h : T_\Gamma(X) \to M$$

so that $F = h^{-1}(h(F))$, or equivalently F is saturated by a locally finite congruence $\sim$ on $T_\Gamma(X)$.

A tree automaton is a system $\mathcal{A} = (\Gamma, Q, h_\mathcal{A}, I_\mathcal{A}, T_\mathcal{A})$, where Q is the finite set of states, $I_\mathcal{A}, T_\mathcal{A} \subseteq Q$ are the sets of initial and final states and $h_\mathcal{A} : \Gamma \to R(Q)$ is the transition function, where $R(Q)$ denotes the theory of relations over Q. The behavior of $\mathcal{A}$ is the tree language $|\mathcal{A}| = \bar{h}^{-1}_\mathcal{A}(\mathcal{T})$, where $\bar{h}_\mathcal{A} : T_\Gamma(X) \to R(Q)$ is the theory morphism uniquely extending $h_\mathcal{A}$ and $\mathcal{T} \subseteq R(Q)$ is constructed by $I_\mathcal{A}, T_\mathcal{A}$ in the obvious way. It is well known that $F \subseteq T_\Gamma(X)$ is recognizable if and only if it is the behavior of such an automaton (cf. [14,4]).

The notion of magmoids, introduced by Arnold and Dauchet (cf. [2,3]), is the algebraic structure which has, at the level of graphs, the role that algebraic theories have for the case of trees. Recall that a magmoid is a double ranked set $M = (M_{m,n})_{m,n\geq 0}$ equipped with two operations

$$\circ : M_{m,n} \times M_{n,k} \to M_{m,k}, \quad m, n, k \geqslant 0,$$
$$\square : M_{m,n} \times M_{m',n'} \to M_{m+m',n+n'}, \quad m, n, m', n' \geqslant 0,$$

which are associative, unitary and mutually coherent in a canonical way.

It is known that the set of (hyper)graphs $GR(\Sigma)$, with hyperedges labelled over a finite doubly ranked alphabet Σ, can be organized into such a structure with $\circ$ being the graph product and $\square$ the graph sum (cf. [6,7]). Moreover Engelfriet and Vereijken have proved that, $GR(\Sigma)$ is finitely generated, that is, any graph can be built from a specific finite set of elementary graphs by using the two operations of product and sum (Theorem 7 of [12]).

However, for any given hypergraph, there are infinitely many expressions representing it. This ambiguity was recently settled by constructing a *finite* set of equations $\mathcal{E}$ with the property that two expressions represent the same hypergraph if and only if one can be transformed into the other through these equations (cf. [6]).

Therefore, $GR(\Sigma)$ is characterized as the quotient of the free magmoid generated by Σ, divided by $\mathcal{E}$, or equivalently, it is the free object generated by Σ within the category of all magmoids that satisfy the equations $\mathcal{E}$, called *graphoids*.

In this respect, various algebraic properties of graphs and graph languages can be investigated inside the framework of magmoids. The scope of the present paper is to study automata on patterns and graphs.

A graph automaton is a system $\mathcal{A} = (\Sigma, Q, \delta_\mathcal{A}, I_\mathcal{A}, T_\mathcal{A})$, where Q is the finite set of states, $I_\mathcal{A}, T_\mathcal{A} \subseteq Q^*$ initial and final rational languages, $\delta_\mathcal{A} : \Sigma \to Rel(Q)$ is the transition function, and $Rel(Q) = (Rel_{m,n}(Q))$, $Rel_{m,n}(Q) = \{R \mid R \subseteq Q^m \times Q^n\}$, $m, n \geq 0$, is the graphoid of relations (over Q) with operations composition and boxing of relations. The behavior of $\mathcal{A}$ is $|\mathcal{A}| = \bar{\delta}^{-1}_\mathcal{A}(\mathcal{T})$, where $\bar{\delta}_\mathcal{A} : GR(\Sigma) \to Rel(Q)$ is the graphoid morphism uniquely extending $\delta_\mathcal{A}$ and $\mathcal{T} \subseteq Rel(Q)$ is constructed from $I_\mathcal{A}, T_\mathcal{A}$ in the obvious way. A graph language is called recognizable if it is obtained as a behavior of a graph automaton. As opposed to the case of context free graph languages (cf. [11]), the membership problem is decidable in polynomial time for recognizable graph languages.

In [7] graph language recognizability through congruences is considered. To any language $L \subseteq GR(\Sigma)$, a congruence $\sim_L$ (called syntactic) is associated, which constitutes the most economical congruence saturating L. Thus L is syntactically recognizable if and only if $\sim_L$ has locally finite index that is if the quotient graphoid $GR(\Sigma)/\sim_L$ (called syntactic) is locally finite.

In this paper we shall see that the above two notions of graph recognizability are not equivalent — as it is the case for trees. Actually, we prove that the class of recognizable graph languages is strictly contained into that of syntactically recognizable languages.

The paper is divided into 6 sections. The notion of a (semi-)magmoid, together with some preliminary matter, is presented in Section 2. Examples of this algebraic structure are considered. We particularly insist on the construction of the *magmoid of hypergraphs* by recalling the definition of hypergraphs introduced in [12] together with the operations product and sum. Moreover, we construct the free magmoid generated by a doubly ranked alphabet and give the definitions of pattern and pattern language.

In Section 3 we introduce graphoids. An important example of a graphoid, which will be useful in the construction of graph automata, is presented. Moreover, we prove that the set of all graphs over the doubly ranked alphabet Σ, is the

Author's personal copy

free graphoid generated by Σ. Automata on graphs are introduced in Section 4. We prove that the membership problem is decidable in polynomial time for the class of recognizable graph languages. Closure properties are investigated in Section 5. Automaton recognizability is shown to be closed under union, intersection, $\Box$-product and inverse graph homomorphism but not under complement and $\circ$-product.

In Section 6 we compare the notions of recognizability and syntactic recognizability. The class of recognizable graph languages is strictly contained into that of syntactically recognizable graph languages and the complement of any recognizable graph language is syntactically recognizable.

2. Preliminaries

Recall that a doubly ranked set (or a doubly ranked alphabet) $(A_{m,n})_{m,n\in\mathbb{N}}$ is a set A together with a function $rank : A \to \mathbb{N} \times \mathbb{N}$, where $\mathbb{N}$ is the set of natural numbers. For $m, n \in \mathbb{N}$, $A_{m,n}$ is the set $\{a \in A \mid rank(a) = (m, n)\}$. In what follows we will drop the subscript $m, n \in \mathbb{N}$ and denote a doubly ranked set simply by $(A_{m,n})$.

A *semi-magmoid* is a doubly ranked set $M = (M_{m,n})$ equipped with two operations

$$\circ : M_{m,n} \times M_{n,k} \to M_{m,k}, \quad m, n, k \geqslant 0$$
$$\Box : M_{m,n} \times M_{m',n'} \to M_{m+m',n+n'}, \quad m, n, m', n' \geqslant 0$$

which are associative in the obvious way and satisfy the distributivity law

$$(f \circ g) \Box (f' \circ g') = (f \Box f') \circ (g \Box g')$$

whenever all the above operations are defined.

A *magmoid* is a semi-magmoid $M = (M_{m,n})$, equipped with a sequence of constants $e_n \in M_{n,n}$ $(n \geqslant 0)$, called units, such that

$$e_m \circ f = f = f \circ e_n, \qquad e_0 \Box f = f = f \Box e_0$$

for all $f \in M_{m,n}$ and all $m, n \geqslant 0$, and the additional condition

$$e_m \Box e_n = e_{m+n}, \quad \text{for all } m, n \geqslant 0$$

holds.

Notice that, due to the last equation, the element e_n $(n \geq 2)$ is uniquely determined by e_1. From now on e_1 will be simply denoted by e.

In other words a magmoid is nothing but an *x-category* (cf. [8,16,17]) or strict monoidal category (cf. Chapter VII of [20]) whose set of objects is the set of natural numbers. Subsemi-magmoids, morphisms, congruences and quotients of semi-magmoids (resp. magmoids) are defined in the obvious way.

In order to fix our notation we need the following definitions. Let $M = (M_{m,n})$ be a magmoid. We say that a doubly ranked family $L = (L_{m,n})$ is a *subset* of M (notation $L \subseteq M$), whenever $L_{m,n} \subseteq M_{m,n}$ for all $m, n \in \mathbb{N}$. The boolean operations on subsets of M are defined in the obvious way.

Given subsets L, L' of a magmoid M (with unit sequence e_n) we define their $\circ$-*product* $L \circ L'$ by setting

$$(L \circ L')_{m,n} = \bigcup_{k\geqslant 0} L_{m,k} \circ L'_{k,n}, \quad m, n \in \mathbb{N}$$

and their $\Box$-*product* $L \Box L'$ by setting

$$(L \Box L')_{m,n} = \bigcup_{\substack{\kappa+\kappa'=m \\ \lambda+\lambda'=n}} L_{\kappa,\lambda} \Box L'_{\kappa',\lambda'}, \quad m, n \in \mathbb{N}.$$

The subsets E and F of M given by $E_{m,n} = \{e_n\}$ if $m = n$ and $\emptyset$ else, while $F_{m,n} = \{e_0\}$ if $m = n = 0$ and $\emptyset$ else, are the units of the operations $\circ$ and $\Box$ respectively.

The $\circ$-*star* is the union of the successive $\circ$-*powers* of $L \subseteq M$:

$$L^\circ = \bigcup_{k\geqslant 0} L^{\circ,k},$$

Author's personal copy

where $L^{\circ,k}$ is inductively given by

$$L^{\circ,0} = E, \qquad L^{\circ,1} = L, ..., L^{\circ,k+1} = L \circ L^{\circ,k}.$$

The $\square$-*star* $L^{\square}$ is defined analogously.

Example 1 (*Magmoids of Functions and Relations*). The sets $Funct_{m,n}(Q)$ of all functions from Q^m to Q^n

$$Funct_{m,n}(Q) = \{f \mid f : Q^m \to Q^n\}, \quad m, n \geqslant 0,$$

can be structured into a magmoid with $\circ$ being the usual function composition, while the operation $\square$ is the function boxing defined as follows: for $f \in Funct_{m,n}(Q)$ and $f' \in Funct_{m',n'}(Q)$

$$(f \square f')(uu') = f(u) \square f'(u'), \quad u \in Q^m,\ u' \in Q^{m'}.$$

In a similar way the sets

$$Rel_{m,n}(Q) = \{R \mid R \subseteq Q^m \times Q^n\}$$

of all relations from Q^m to Q^n can be organized into a magmoid, $\circ$ being the relation composition and $\square$ the relation boxing, defined as above. Its units $e_n \subseteq Q^n \times Q^n$ are given by

$$e_n = \{(u, u) \mid u \in Q^n\}, \quad n > 0, \quad \text{and} \quad e_0 = \{(\varepsilon, \varepsilon)\}, \quad \varepsilon \text{ the empty word of } Q^*.$$

Clearly $Funct(Q) = (Funct_{m,n}(Q))$ is a sub-magmoid of $Rel(Q) = (Rel_{m,n}(Q))$.

Let Σ be a doubly ranked alphabet. *We denote by* $SM(\Sigma) = (SM_{m,n}(\Sigma))$ *the smallest doubly ranked set satisfying the next items*:

- $\Sigma_{m,n} \subseteq SM_{m,n}(\Sigma)$ for all $m, n \geq 0$,
- if $p \in SM_{m,n}(\Sigma)$ and $q \in SM_{n,k}(\Sigma)$ then their horizontal concatenation $p\,q \in SM_{m,k}(\Sigma)$,
- if $p \in SM_{m,n}(\Sigma)$ and $p' \in SM_{m',n'}(\Sigma)$ then their vertical concatenation $\begin{matrix} p \\ p' \end{matrix} \in SM_{m+m',n+n'}(\Sigma)$.

Let $\sim = (\sim_{m,n})$ be the doubly ranked equivalence on $SM(\Sigma)$, compatible with horizontal and vertical concatenation, generated by the relations

$$\begin{matrix} p_1\ p_1' \\ p_2\ p_2' \end{matrix} \sim \begin{matrix} p_1 & p_1' \\ p_2 & p_2' \end{matrix}$$

for all p_i, p_i' of suitable ranks. The quotient

$$SM(\Sigma)/\sim = (SM_{m,n}(\Sigma)/\sim_{m,n})$$

is denoted $smag(\Sigma)$ and obviously is a semi-magmoid. The elements of $smag_{m,n}(\Sigma)$ are called (m, n)-*patterns* or *patterns of rank* (m, n). Subsets of $smag(\Sigma)$ are called *pattern languages*. Our patterns are analogous with the unsorted abstract dags of [18,19,5]. For another formalization see also [15].

Convention. *In order to avoid confusion in the pattern calculus instead of* $\begin{matrix} p \\ p' \end{matrix}$ *we write* $\begin{pmatrix} p \\ p' \end{pmatrix}$. *The associativity law takes the form*

$$\begin{pmatrix} p_1 \\ \begin{pmatrix} p_2 \\ p_3 \end{pmatrix} \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} p_1 \\ p_2 \end{pmatrix} \\ p_3 \end{pmatrix}.$$

This common pattern will be denoted

$$\begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix}.$$

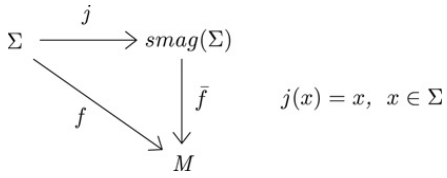
Author's personal copy

The distributivity law takes the form

$$\begin{pmatrix} p_1\, p'_1 \\ p_2\, p'_2 \end{pmatrix} = \begin{pmatrix} p_1 \\ p_2 \end{pmatrix} \begin{pmatrix} p'_1 \\ p'_2 \end{pmatrix}.$$

Actually $smag(\Sigma)$ is the *free* semi-magmoid generated by Σ as confirmed by the next result.

Proposition 1. *For every semi-magmoid $M = (M_{m,n})$ and every doubly ranked function $f : \Sigma \to M$, there exists a unique morphism of semi-magmoids $\bar{f} : smag(\Sigma) \to M$ making the following triangle commutative.*



Actually, $\bar{f}$ is given by the clauses,

- $\bar{f}(x) = f(x)$, *for all* $x \in \Sigma$,
- $\bar{f}(p\, q) = \bar{f}(p) \circ \bar{f}(q)$, $\bar{f}\begin{pmatrix} p \\ p' \end{pmatrix} = \bar{f}(p) \,\square\, \bar{f}(p')$,

for all $p, q, p' \in smag(\Sigma)$ of suitable rank.

The construction of the free magmoid follows naturally. Let $(e_n)_{n\geq 0}$ be a sequence of symbols not in Σ and denote by $mag(\Sigma)$ the free semi-magmoid $smag(\Sigma \cup \{e_n \mid n \geq 0\})$ divided by the congruence generated by the relations

$$e_m\, p \approx p \approx p\, e_n, \quad \begin{pmatrix} e_0 \\ p \end{pmatrix} \approx p \approx \begin{pmatrix} p \\ e_0 \end{pmatrix}, \quad e_m\, e_n \approx e_{m+n}$$

for all $m, n \geq 0$ and all patterns p of suitable rank. Then $mag(\Sigma)$ clearly constitutes a magmoid which has a universal property analogous to the one stated in Proposition 1, i.e., $mag(\Sigma)$ is the *free* magmoid generated by Σ (cf. [6]).

Next we introduce the magmoid of hypergraphs which will be of constant use throughout this paper. Given a finite alphabet X, we denote by X^* the set of all words over X and for every word $w \in X^*$, $|w|$ denotes its length. Formally, a *concrete (m, n)-graph* over a doubly ranked alphabet $\Sigma = (\Sigma_{m,n})$ is a tuple

$$G = (V, E, s, t, l, begin, end),$$

where

- V is the finite set of nodes,
- E is the finite set of hyperedges,
- $s : E \to V^*$ is the source function,
- $t : E \to V^*$ is the target function,
- $l : E \to \Sigma$ is the labelling function such that $rank(l(e)) = (|s(e)|, |t(e)|)$ for every $e \in E$,
- $begin \in V^*$ with $|begin| = m$ is the sequence of begin nodes and
- $end \in V^*$ with $|end| = n$ is the sequence of end nodes.

Notice that according to this definition vertices can be duplicated in the begin and end sequences of the graph and also at the sources and targets of an edge. For an edge e of a hypergraph G we simply write $rank(e)$ to denote $rank(l(e))$.

The specific sets V and E chosen to define a concrete graph G are actually irrelevant. We shall not distinguish between two isomorphic graphs. Hence we have the following definition of an abstract graph. Two concrete (m, n)-graphs $G = (V, E, s, t, l, begin, end)$ and $G' = (V', E', s', t', l', begin', end')$ over Σ are isomorphic iff there exist two bijections $h_V : V \to V'$ and $h_E : E \to E'$ commuting with source, target, labelling, *begin* and *end* in the usual way.

An *abstract (m, n)-graph* is defined to be the equivalence class of a concrete (m, n)-graph with respect to isomorphism. We denote by $GR_{m,n}(\Sigma)$ the set of all abstract (m, n)-graphs over Σ. Since we shall mainly be

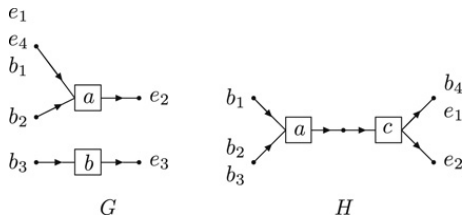
Author's personal copy

interested in abstract graphs we simply call them graphs except when it is necessary to emphasize that they are defined up to an isomorphism. Any graph $G \in GR_{m,n}(\Sigma)$ having no edges is called a *discrete* (m, n)*-graph*.

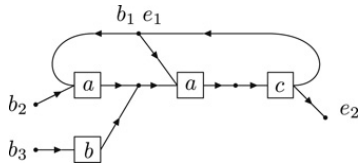
If G is an (m, n)-graph represented by $(V, E, s, t, l, begin, end)$ and H is an (n, k)-graph represented by $(V', E', s', t', l', begin', end')$ then their *product* $G \circ H$ is the (m, k)-graph represented by the concrete graph obtained by taking the disjoint union of G and H and then identifying the ith end node of G with the ith begin node of H, for every $i \in \{1, ..., n\}$; also, $begin(G \circ H) = begin(G)$ and $end(G \circ H) = end(H)$.

The *sum* $G \square H$ of arbitrary graphs G and H is their disjoint union with their sequences of begin nodes concatenated and similarly for their end nodes.

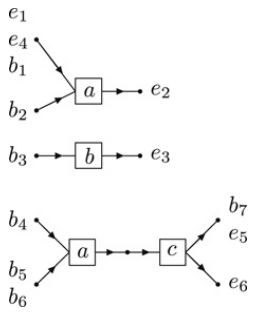
For instance let $\Sigma = \{a, b, c, d\}$, with $rank(a) = (2, 1)$, $rank(b) = (1, 1)$ and $rank(c) = (1, 2)$. In the following pictures, edges are represented by boxes, nodes by dots, and the sources and targets of an edge by directed lines that enter and leave the corresponding box, respectively. The order of the sources and targets of an edge is the vertical order of the directed lines as drawn in the pictures. We display two graphs $G \in GR_{3,4}(\Sigma)$ and $H \in GR_{4,2}(\Sigma)$, where the ith begin node is indicated by b_i, and the ith end node by e_i.



Then their product $G \circ H$ is the $(3, 2)$-graph



and, their sum $G \square H$ is the $(7, 6)$-graph



For every $n \in \mathbb{N}$ we denote by E_n the discrete graph of rank (n, n) with nodes $x_1, \ldots, x_n$ and $begin(E_n) = end(E_n) = x_1 \cdots x_n$; we write E for E_1. Note that E_0 is the empty graph.

It is straightforward to verify that $GR(\Sigma) = (GR_{m,n}(\Sigma))$ with the operations defined above is a magmoid, whose units are the graphs $E_n, n \geq 0$, see Lemma 6 of [12]. Subsets of $GR(\Sigma)$ are referred to as graph languages. The discrete graphs of $GR(\Sigma)$ form obviously a sub-magmoid $DISC$ of $GR(\Sigma)$ and the function sending each graph $G \in GR(\Sigma)$ to its underlying discrete graph is indeed an epimorphism of magmoids

$$disc_\Sigma : GR(\Sigma) \to DISC.$$

Engelfriet and Vereijken proved that, $GR(\Sigma)$ is finitely generated, that is, any graph can be built from a specific finite set of elementary graphs (cf. [12]). More precisely, let us denote by $I_{p,q}$ the discrete (p, q)-graph having a single node x and whose begin and end sequences are $x \cdots x$ (p times) and $x \cdots x$ (q times) respectively. Note that $I_{1,1}$ is equal with E. Let also Π be the discrete $(2, 2)$-graph having two nodes x and y and whose begin and end sequences are xy and yx, respectively. Finally, for every $\sigma \in \Sigma_{m,n}$, we denote again by σ the (m, n)-graph having only one edge

Author's personal copy

and $m+n$ nodes $x_1, \ldots, x_m, y_1, \ldots, y_n$. The edge is labelled by σ, and the begin (resp. end) sequence of the graph is the sequence of sources (resp. targets) of the edge, viz. $x_1 \cdots x_m$ (resp. $y_1 \cdots y_n$).

Now let us introduce the alphabet D, formed by the following five symbols

$$i_{21} : 2 \to 1 \quad i_{01} : 0 \to 1 \quad i_{12} : 1 \to 2 \quad i_{10} : 1 \to 0 \quad \pi : 2 \to 2,$$

where $x : m \to n$ indicates that symbol x has first rank m and second rank n, and denote by $mag(\Sigma \cup D)$ the free magmoid generated by the doubly ranked alphabet $\Sigma \cup D$. We denote by

$$val_\Sigma : mag(\Sigma \cup D) \to GR(\Sigma)$$

the unique magmoid morphism extending the function described by the assignments

$$i_{21} \mapsto I_{2,1}, \quad i_{01} \mapsto I_{0,1}, \quad i_{12} \mapsto I_{1,2}, \quad i_{10} \mapsto I_{1,0}, \quad \pi \mapsto \Pi,$$
$$\sigma \mapsto \sigma, \text{ for all } \sigma \in \Sigma, \qquad e_n \mapsto E_n, \text{ for all } n \in \mathbb{N}.$$

Theorem 1 (*cf. [12]*). *The magmoid $GR(\Sigma)$ is generated by the set*

$$\Sigma \cup \{I_{12}, I_{10}, I_{21}, I_{01}, \Pi\}.$$

The previous theorem implies that the morphism val_Σ is a surjection. However, val_Σ is not an injection and in fact, for any given hypergraph, there are infinitely many patterns representing it.

This ambiguity was recently settled by constructing a *finite* set of equations with the property that two patterns represent the same hypergraph if and only if one can be transformed into the other through these equations (cf. [6]).

More precisely, we denote by $\pi_{n,1}$ the pattern inductively defined by

$$\pi_{1,0} = e, \quad \pi_{n,1} = \begin{pmatrix} \pi_{n-1,1} \\ e \end{pmatrix} \begin{pmatrix} e_{n-1} \\ \pi \end{pmatrix}.$$

Notice that for $n = 1$, $\pi_{1,1} = \pi$.

Given a finite doubly ranked alphabet Σ, the set of equations $\mathcal{E}$:

$$\pi\pi = e_2, \qquad \begin{pmatrix} e \\ \pi \end{pmatrix} \begin{pmatrix} \pi \\ e \end{pmatrix} \begin{pmatrix} e \\ \pi \end{pmatrix} = \begin{pmatrix} \pi \\ e \end{pmatrix} \begin{pmatrix} e \\ \pi \end{pmatrix} \begin{pmatrix} \pi \\ e \end{pmatrix},$$
$$\begin{pmatrix} e \\ i_{21} \end{pmatrix} i_{21} = \begin{pmatrix} i_{21} \\ e \end{pmatrix} i_{21}, \qquad \begin{pmatrix} e \\ i_{01} \end{pmatrix} i_{21} = e,$$
$$\pi i_{21} = i_{21}, \qquad \begin{pmatrix} e \\ i_{01} \end{pmatrix} \pi = \begin{pmatrix} i_{01} \\ e \end{pmatrix},$$
$$\begin{pmatrix} \pi \\ e \end{pmatrix} \begin{pmatrix} e \\ \pi \end{pmatrix} \begin{pmatrix} i_{21} \\ e \end{pmatrix} = \begin{pmatrix} e \\ i_{21} \end{pmatrix} \pi,$$
$$i_{12} \begin{pmatrix} e \\ i_{12} \end{pmatrix} = i_{12} \begin{pmatrix} i_{12} \\ e \end{pmatrix}, \qquad i_{12} \begin{pmatrix} e \\ i_{10} \end{pmatrix} = e,$$
$$i_{12}\pi = i_{12}, \qquad \pi \begin{pmatrix} e \\ i_{10} \end{pmatrix} = \begin{pmatrix} i_{10} \\ e \end{pmatrix},$$
$$\begin{pmatrix} i_{12} \\ e \end{pmatrix} \begin{pmatrix} e \\ \pi \end{pmatrix} \begin{pmatrix} \pi \\ e \end{pmatrix} = \pi \begin{pmatrix} e \\ i_{12} \end{pmatrix},$$
$$i_{12}\, i_{21} = e, \qquad \begin{pmatrix} i_{12} \\ e \end{pmatrix} \begin{pmatrix} e \\ i_{21} \end{pmatrix} = i_{21}\, i_{12},$$
$$\pi_{m,1} \begin{pmatrix} \sigma \\ e \end{pmatrix} = \begin{pmatrix} e \\ \sigma \end{pmatrix} \pi_{n,1}, \quad \text{where } \sigma \in \Sigma_{m,n}, \quad m, n \geq 0,$$

has the following property: for all patterns p and q,

$$val_\Sigma(p) = val_\Sigma(q) \quad \text{if and only if } p \underset{\mathcal{E}}{=} q.$$

Author's personal copy

To better understand the internal structure of a graph, a normal form representation is necessary. As usual $\mathcal{S}_m$ stands for the group of all permutations of the set $\{1, 2, \ldots, m\}$. Given a permutation $\alpha \in \mathcal{S}_m$

$$\alpha = \begin{pmatrix} 1 & 2 & \ldots & m \\ \alpha(1) & \alpha(2) & \ldots & \alpha(m) \end{pmatrix}$$

the discrete graph having $\{x_1, x_2, \ldots, x_m\}$ as set of nodes, $x_{\alpha(1)}x_{\alpha(2)} \cdots x_{\alpha(m)}$ as begin sequence and $x_1x_2 \cdots x_m$ as end sequence, is denoted Π_α and is called the *permutation graph associated with* α. Observe that Π is the graph associated with the permutation

$$\alpha = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$$

whereas the pattern $\pi_{n,1}$, introduced previously, represents the graph associated with the permutation

$$\begin{pmatrix} 1 & 2 & \ldots & n+1 \\ 2 & \ldots & n+1 & 1 \end{pmatrix}$$

interchanging the last n numbers with the first one.

It is easy to see that for all $\alpha, \beta \in \mathcal{S}_m$ and $\alpha' \in \mathcal{S}_{m'}$,

$$\Pi_\alpha \circ \Pi_\beta = \Pi_{\alpha\circ\beta} \quad \text{and} \quad \Pi_\alpha \,\square\, \Pi_{\alpha'} = \Pi_{\alpha\square\alpha'}$$

so that the sets

$$PERM_m = \{\Pi_\alpha \mid \alpha \in \mathcal{S}_m\}$$

with $m \geq 0$, form a sub-magmoid $PERM = (PERM_{m,n})$ of DISC, by setting $PERM_{m,m} = PERM_m$ and $PERM_{m,n} = \emptyset$, if $m \neq n$.

For every $\sigma \in \Sigma_{m,n}$, $BF(\sigma)$ is the same as the graph σ, except that $begin(BF(\sigma)) = begin(\sigma)end(\sigma) = x_1 \cdots x_m y_1 \cdots y_n$ and $end(BF(\sigma)) = \varepsilon$ (where ε denotes the empty sequence). The $(m+n, 0)$-graph $BF(\sigma)$ is equal with (cf. [6])

$$BF(\sigma) = (\sigma \,\square\, E_n) \circ \Pi_{\alpha_n} \circ I_{20}^n,$$

where

$$I_{20}^n = (I_{21} \circ I_{10}) \,\square \cdots \square\, (I_{21} \circ I_{10}), \quad n \text{ times},$$

and $\alpha_n \in \mathcal{S}_{2n}$ is defined by $\alpha_n(i) = 2i - 1$, $\alpha_n(n+i) = 2i$, for $1 \leq i \leq n$.

Now we present the canonical decomposition of a graph (cf. [6]).

Theorem 2. (i) *Every discrete graph $\Delta \in DISC$, $\Delta \neq E_0$, can be written in the following normal form.*

$$\Pi_\alpha \circ (I_{p_1,q_1} \,\square \cdots \square\, I_{p_n,q_n}) \circ \Pi_\beta.$$

(ii) *Every graph $G \in GR(\Sigma)$, $G \neq 0$, can be written in the following normal form.*

$$\Pi_\alpha \circ (I_{p_1,q_1} \,\square \cdots \square\, I_{p_n,q_n}) \circ \Pi_\beta \circ (E_s \,\square\, BF(\sigma_1) \,\square \cdots \square\, BF(\sigma_k)),$$

where Π_α, Π_β are permutation graphs.

The next definition will be used later on. We call *size* of a pattern $p \in smag(\Sigma)$ the number of symbols of Σ occurring in p. The *size* of a graph $G \in GR(\Sigma)$ is then

$$size(G) = min\{size(p) \mid p \in val_\Sigma^{-1}(G)\}.$$

Author's personal copy

3. Graphoids

As we have seen in the previous section, the equations $\mathcal{E}$ are satisfied in $GR(\Sigma)$ by replacing π by Π and $i_{\kappa\lambda}$ by $I_{\kappa,\lambda}$. Magmoids with such a property are called *graphoids*. More precisely, a graphoid $\mathbf{M} = (M, D)$ consists of a semi-magmoid M and a set $D = \{\eta_0, \eta, s, d_{01}, d_{21}, d_{10}, d_{12}\}$, where $\eta_0 \in M_{0,0}$, $\eta \in M_{1,1}$, $s \in M_{2,2}$, $d_{01} \in M_{0,1}$, $d_{21} \in M_{2,1}$, $d_{10} \in M_{1,0}$, $d_{12} \in M_{1,2}$ such that (M, η_0, η) is a magmoid, i.e.,

(1) $\eta_m \circ f = f = f \circ \eta_n, \quad \eta_0 \,\square\, f = f = f \,\square\, \eta_0,$

where $\eta_n = \eta \,\square \cdots \square\, \eta$ (n-times, $n \geq 0$), $f \in M_{m,n}$, $(m, n \geq 0)$, and additionally the following equations hold

(2) $s \circ s = \eta_2, \quad (s \,\square\, \eta) \circ (\eta \,\square\, s) \circ (s \,\square\, \eta) = (\eta \,\square\, s) \circ (s \,\square\, \eta) \circ (\eta \,\square\, s)$

(3) $(\eta \,\square\, d_{21}) \circ d_{21} = (d_{21} \,\square\, \eta) \circ d_{21}, \quad (\eta \,\square\, d_{01}) \circ d_{21} = \eta,$

$s \circ d_{21} = d_{21}, \quad (\eta \,\square\, d_{01}) \circ s = (d_{01} \,\square\, \eta)\,,$

$(s \,\square\, \eta) \circ (\eta \,\square\, s) \circ (d_{21} \,\square\, \eta) = (\eta \,\square\, d_{21}) \circ s,$

(4) $d_{12} \circ (\eta \,\square\, d_{12}) = d_{12} \circ (d_{12} \,\square\, \eta)\,, \quad d_{12} \circ (\eta \,\square\, d_{10}) = \eta,$

$d_{12} \circ s = d_{12}, \quad s \circ (\eta \,\square\, d_{10}) = (d_{10} \,\square\, \eta)\,,$

$(d_{12} \,\square\, \eta) \circ (\eta \,\square\, s) \circ (s \,\square\, \eta) = s \circ (\eta \,\square\, d_{12})\,,$

(5) $d_{12} \circ d_{21} = \eta, \quad (d_{12} \,\square\, \eta) \circ (\eta \,\square\, d_{21}) = d_{21} \circ d_{12}.$

(6) $s_{m,1} \circ (f \,\square\, \eta) = (\eta \,\square\, f) \circ s_{n,1}, \quad$ for all $f \in M_{m,n},$

where $s_{m,1}$ is defined inductively by s analogously with $\pi_{m,1}$ (see Section 2).

We point out that the last equation holds in $GR(\Sigma)$ since it holds for all the letters of the doubly ranked alphabet Σ (cf. [6]). Thus the pair $GR(\Sigma) = (GR(\Sigma), D)$, where $D = \{E_0, E, \Pi, I_{0,1}, I_{2,1}, I_{1,0}, I_{1,2}\}$ is a graphoid.

The semi-magmoid $Rel(Q)$ of relations over a set Q, can be structured into a graphoid by considering the following relations.

- $\eta_0^Q \in Rel_{0,0}(Q)$, with $\eta_0^Q = \{(\varepsilon, \varepsilon)\}$,
- $\eta^Q \in Rel_{1,1}(Q)$, with $\eta^Q = \{(q, q) \mid q \in Q\}$,
- $s^Q \in Rel_{2,2}(Q)$, with $s^Q = \{(q_1q_2, q_2q_1) \mid q_1, q_2 \in Q\}$,
- $d_{01}^Q \in Rel_{0,1}(Q)$, with $d_{01}^Q = \{(\varepsilon, q) \mid q \in Q\}$,
- $d_{21}^Q \in Rel_{2,1}(Q)$, with $d_{21}^Q = \{(qq, q) \mid q \in Q\}$,
- $d_{10}^Q \in Rel_{1,0}(Q)$, with $d_{10}^Q = \{(q, \varepsilon) \mid q \in Q\}$,
- $d_{12}^Q \in Rel_{1,2}(Q)$, with $d_{12}^Q = \{(q, qq) \mid q \in Q\}$,

where ε is the empty word of Q^*. The pair $\mathbf{Rel}(Q) = (Rel(Q), D_Q)$, with

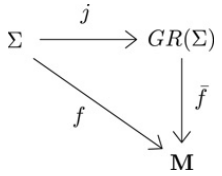
$$D_Q = \{\eta_0^Q, \eta^Q, s^Q, d_{01}^Q, d_{21}^Q, d_{10}^Q, d_{12}^Q\}$$

obviously constitutes a graphoid which will be used in the construction of graph automata.

Given graphoids (M, D) and (M', D'), a semi-magmoid morphism $H : M \to M'$ preserving D-sets, i.e., $H(\eta_0) = \eta_0'$, $H(\eta) = \eta'$, $H(s) = s'$ and $H(d_{\kappa\lambda}) = d'_{\kappa\lambda}$, is called a morphism of graphoids.

We have already discussed how the set $GR(\Sigma)$ can be structured into a graphoid; in fact it is the free graphoid generated by Σ.

Theorem 3. *The doubly ranked function $j : \Sigma \to GR(\Sigma)$, with $j(\sigma) = \sigma$, for all $\sigma \in \Sigma$, has the following universal property: for any graphoid $\mathbf{M} = (M, D)$, $D = \{\eta_0, \eta, s, d_{10}, d_{12}, d_{01}, d_{21}\}$ and any doubly ranked function $f : \Sigma \to M$, there exists a unique morphism of graphoids $\bar{f} : GR(\Sigma) \to \mathbf{M}$ making commutative the following triangle.*



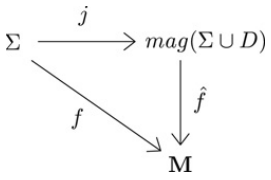
Author's personal copy

The morphism $\bar{f}$ is defined by the clauses

- $\bar{f}(\sigma) = f(\sigma), \sigma \in \Sigma$,
- $\bar{f}(E_0) = \eta_0$, $\bar{f}(E) = \eta$, $\bar{f}(\Pi) = s$, $\bar{f}(I_{ij}) = d_{ij}$,
- $\bar{f}(G_1 \circ G_2) = \bar{f}(G_1) \circ \bar{f}(G_2)$,
- $\bar{f}(G_1 \square G_2) = \bar{f}(G_1) \square \bar{f}(G_2)$,

for all graphs G_1, G_2 of suitable rank.

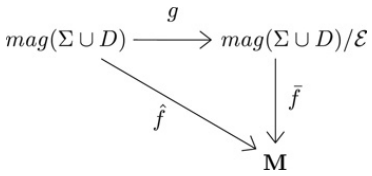
Proof. By virtue of Proposition 1 there exists a unique magmoid morphism $\hat{f}$ making commutative the triangle:



Since all equations $\mathcal{E} = (1) \cup (2) \cup (3) \cup (4) \cup (5) \cup (6)$ are valid in $\mathbf{M}$, the kernel of $\hat{f}$ includes $\mathcal{E}$, $Ker(\hat{f}) \supseteq \mathcal{E}$, and thus $\hat{f}$ induces a unique graphoid morphism

$$\bar{f} : mag(\Sigma \cup D)/\mathcal{E} \longrightarrow \mathbf{M}$$

rendering commutative the triangle



where g is the canonical projection sending every element of $mag(\Sigma \cup D)$ to its class with respect to the congruence generated by $\mathcal{E}$. The result comes by combining the above two diagrams and the main result of [6] stating that the magmoids $mag(\Sigma \cup D)/\mathcal{E}$ and $GR(\Sigma)$ are isomorphic. □

A *graph homomorphism* $H : GR(\Sigma) \to GR(\Sigma')$ is just a morphism of graphoids. Hence, by virtue of the previous theorem it is completely determined by its values $H(\sigma)$, $\sigma \in \Sigma$. A graph homomorphism $H : GR(\Sigma) \to GR(\Sigma')$ is called a projection whenever $H(\Sigma) \subseteq \Sigma'$.

4. Pattern and graph automata

Pattern automata have originated in [10], see also [21], and their basic properties are examined in [18,19] and [5] under the name of pdag automata. A *pattern automaton* is a structure

$$\mathcal{A} = (\Sigma, Q, \theta_{\mathcal{A}}, I_{\mathcal{A}}, T_{\mathcal{A}}),$$

where Σ, Q are the finite sets of input alphabet and states respectively, $I_{\mathcal{A}}, T_{\mathcal{A}} \subseteq Q^*$ are rational languages for the initial and final configurations and

$$\theta_{\mathcal{A}} : \Sigma \to Rel(Q)$$

is the move function. Its *behavior* is the pattern language

$$|\mathcal{A}| = \{p \mid p \in smag_{m,n}(\Sigma),\ \bar{\theta}_{\mathcal{A}}(p) \cap (I_{\mathcal{A}}^{(m)} \times T_{\mathcal{A}}^{(n)}) \neq \emptyset,\ m, n \in \mathbb{N}\},$$

where $I_{\mathcal{A}}^{(m)} = I_{\mathcal{A}} \cap Q^m$ and $T_{\mathcal{A}}^{(n)} = T_{\mathcal{A}} \cap Q^n$ and $\bar{\theta}_{\mathcal{A}} : smag(\Sigma) \to Rel(Q)$ is the unique semi-magmoid morphism extending $\theta_{\mathcal{A}}$.

A *graph automaton* over Σ is a structure $\mathcal{A} = (\Sigma, Q, \delta_{\mathcal{A}}, I_{\mathcal{A}}, T_{\mathcal{A}})$, where:

- Q is the finite set of states;

Author's personal copy

- $\delta_{\mathcal{A}} : \Sigma \to \mathbf{Rel}(Q)$ is the doubly ranked transition function;
- $I_{\mathcal{A}}, T_{\mathcal{A}}$ are the initial and final rational subsets of Q^*.

According to Theorem 3 the function

$$\delta_{\mathcal{A}} : \Sigma \to \mathbf{Rel}(Q)$$

is uniquely extended into a morphism of graphoids

$$\bar{\delta}_{\mathcal{A}} : GR(\Sigma) \to \mathbf{Rel}(Q).$$

The behavior of $\mathcal{A}$ is given by

$$|\mathcal{A}| = \{F \mid F \in GR_{m,n}(\Sigma),\ \bar{\delta}_{\mathcal{A}}(F) \cap (I_{\mathcal{A}}^{(m)} \times T_{\mathcal{A}}^{(n)}) \neq \emptyset,\ m, n \in \mathbb{N}\},$$

where $I_{\mathcal{A}}^{(m)} = I_{\mathcal{A}} \cap Q^m$ and $T_{\mathcal{A}}^{(n)} = T_{\mathcal{A}} \cap Q^n$. Graph automata are finite machines due to the fact that the set of equations (1)–(6) is finite.

Notice that the difference between the two notions of automata above lies in the way that their moves are defined: $\theta_{\mathcal{A}}$ ranges over the semi-magmoid $Rel(Q)$ whereas $\delta_{\mathcal{A}}$ ranges over the graphoid $\mathbf{Rel}(Q)$.

A graph language is called recognizable whenever it is obtained as the behavior of a graph automaton. The class of all such languages over the doubly ranked alphabet Σ is denoted by $Rec(\Sigma)$.

Example 2. We recall that $I_{p,q}$ denotes the discrete (p, q)-graph having a single node x and whose begin and end sequences are $x \cdots x$ (p times) and $x \cdots x$ (q times) respectively. Moreover, for all $n \geq 2$ it holds

$$I_{1,n} = I_{1,2} \circ (I_{1,n-1} \square E) = I_{1,2} \circ (E \square I_{1,n-1})$$

and similarly for $I_{n,1}$.

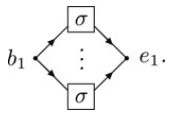
Now let $\sigma \in \Sigma_{1,1}$, then the graph automaton

$$\mathcal{A}_p = (\Sigma, Q, \delta_{\mathcal{A}_p}, I_{\mathcal{A}_p}, T_{\mathcal{A}_p})$$

with $Q = \{q_1, q_2\}$, $I = \{q_1\}$, $T = \{q_2\}$ and $\delta(\sigma) = \{(q_1, q_2)\}$, clearly computes the graph language L_p, containing the graphs:

$$G_n = I_{1,n} \circ (\sigma \square \cdots \square \sigma) \circ I_{n,1},$$

with n occurrences of σ, depicted by



Proposition 2. *The membership problem is decidable in polynomial time for recognizable graph languages.*

Proof. Consider an automaton $\mathcal{A} = (\Sigma, Q, \delta_{\mathcal{A}}, I_{\mathcal{A}}, T_{\mathcal{A}})$ and a graph $F \in GR_{m,n}(\Sigma)$. By inspecting the proof of Theorem 7 of [12], we see that a pattern $f \in smag_{m,n}(\Sigma \cup D)$, can be constructed in polynomial time so that $val_{\Sigma \cup D}(f) = F$. Thus, $\bar{\delta}_{\mathcal{A}}(F)$ can be calculated in polynomial time via f. Moreover, because of the rationality of $I_{\mathcal{A}}, T_{\mathcal{A}}$, the nonemptiness conditions

$$I_{\mathcal{A}}^{(m)} = I_{\mathcal{A}} \cap Q^m \neq \emptyset, \qquad T_{\mathcal{A}}^{(n)} = T_{\mathcal{A}} \cap Q^n \neq \emptyset$$

can be checked in polynomial time as well. Hence, the set

$$\mathbf{C} = \{C \mid C \subseteq Q^m \times Q^n, C \cap (I_{\mathcal{A}}^{(m)} \times T_{\mathcal{A}}^{(n)}) \neq \emptyset\}$$

is constructible in polynomial time and so is the condition

$$F \in |\mathcal{A}|, \quad \text{if and only if, } \bar{\delta}_{\mathcal{A}}(F) \in \mathbf{C}. \quad \square$$

Author's personal copy

We say that the pattern automaton $\mathcal{A} = (\Sigma \cup D, Q, \theta_{\mathcal{A}}, I_{\mathcal{A}}, T_{\mathcal{A}})$ has *graph structure* if

$$\theta_{\mathcal{A}}(e_0) = \eta_0^Q, \qquad \theta_{\mathcal{A}}(e) = \eta^Q, \qquad \theta_{\mathcal{A}}(\pi) = s^Q, \qquad \theta_{\mathcal{A}}(i_{\kappa,\lambda}) = d_{\kappa,\lambda}^Q.$$

Consider the seven element alphabet $D = \{e_0, e, \pi, i_{01}, i_{21}, i_{10}, i_{12}\}$ and the mapping

$$val_\Sigma : smag(\Sigma \cup D) \to GR(\Sigma).$$

Proposition 3. *A graph language $L \subseteq GR(\Sigma)$ is recognizable if and only if the language $val_\Sigma^{-1}(L) \subseteq smag(\Sigma \cup D)$ can be recognized by a pattern automaton with graph structure.*

Proof. Assume that L is recognized by the graph automaton

$$\mathcal{A} = (\Sigma, Q, \delta_{\mathcal{A}}, I_{\mathcal{A}}, T_{\mathcal{A}})$$

then $val_\Sigma^{-1}(L)$ is recognized by the pattern automaton with graph structure

$$\mathcal{A}' = (\Sigma \cup D, Q, \theta_{\mathcal{A}'}, I_{\mathcal{A}}, T_{\mathcal{A}})$$

with $\theta_{\mathcal{A}'} = \delta_{\mathcal{A}} \circ val_\Sigma$.

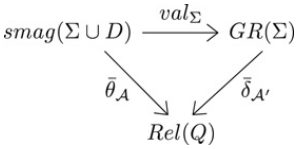
Conversely, let

$$\mathcal{A} = (\Sigma \cup D, Q, \theta_{\mathcal{A}}, I_{\mathcal{A}}, T_{\mathcal{A}})$$

be a graph structure pattern automaton recognizing $val_\Sigma^{-1}(L)$. Then the function $\delta_{\mathcal{A}'} : \Sigma \to Rel(Q)$ with

$$\delta_{\mathcal{A}'}(\sigma) = \theta_{\mathcal{A}}(\sigma), \quad \text{for all } \sigma \in \Sigma$$

is such that the following triangle commutes.



Hence, the graph automaton $\mathcal{A}' = (\Sigma, Q, \delta_{\mathcal{A}'}, I_{\mathcal{A}}, T_{\mathcal{A}})$ recognizes the graph language

$$val_\Sigma(val_\Sigma^{-1}(L)) = L. \quad \square$$

5. Closure properties

In this section we examine the main closure properties of recognizable graph languages.

Proposition 4. *The class $Rec(\Sigma)$ is closed under intersection.*

Proof. Given $R_1 \in Rel(Q_1)$, $R_2 \in Rel(Q_2)$, we define

$$R_1 \boxtimes R_2 = \{((q_1, q_1') \cdots (q_m, q_m'), (h_1, h_1') \cdots (h_n, h_n')) \mid (q_1 \cdots q_m, h_1 \cdots h_n) \in R_1 \text{ and } (q_1' \cdots q_m', h_1' \cdots h_n') \in R_2\}.$$

It holds

$$(R_1 \boxtimes R_2)\Diamond(S_1 \boxtimes S_2) = (R_1 \Diamond S_1) \boxtimes (R_2 \Diamond S_2), \qquad \Diamond = \circ, \Box.$$

We construct the automaton

$$\mathcal{A}_1 \boxtimes \mathcal{A}_2 = (\Sigma, Q_1 \times Q_2, \delta_{\mathcal{A}_1 \boxtimes \mathcal{A}_2}, I_{\mathcal{A}_1 \boxtimes \mathcal{A}_2}, T_{\mathcal{A}_1 \boxtimes \mathcal{A}_2})$$

with

- $\delta_{\mathcal{A}_1 \boxtimes \mathcal{A}_2}(\sigma) = \delta_{\mathcal{A}_1}(\sigma) \boxtimes \delta_{\mathcal{A}_2}(\sigma), \sigma \in \Sigma,$
- $I_{\mathcal{A}_1 \boxtimes \mathcal{A}_2} = pr_{Q_1}^{-1}(I_{\mathcal{A}_1}) \cap pr_{Q_2}^{-1}(I_{\mathcal{A}_2}), T_{\mathcal{A}_1 \boxtimes \mathcal{A}_2} = pr_{Q_1}^{-1}(T_{\mathcal{A}_1}) \cap pr_{Q_2}^{-1}(T_{\mathcal{A}_2})$

Author's personal copy

and

$$pr_{Q_1} : (Q_1 \times Q_2)^* \to Q_1^*, \qquad pr_{Q_2} : (Q_1 \times Q_2)^* \to Q_2^*$$

are the canonical projections. By induction we can show that

$$\bar{\delta}_{\mathcal{A}_1 \boxtimes \mathcal{A}_2}(G) = \bar{\delta}_{\mathcal{A}_1}(G) \boxtimes \bar{\delta}_{\mathcal{A}_2}(G)$$

for all $G \in GR(\Sigma)$. We have

$$I^m_{\mathcal{A}_1 \boxtimes \mathcal{A}_2} = I_{\mathcal{A}_1 \boxtimes \mathcal{A}_2} \cap (Q_1 \times Q_2)^m = I^{(m)}_{\mathcal{A}_1} \times I^{(m)}_{\mathcal{A}_2}$$

$$T^n_{\mathcal{A}_1 \boxtimes \mathcal{A}_2} = T_{\mathcal{A}_1 \boxtimes \mathcal{A}_2} \cap (Q_1 \times Q_2)^n = T^{(n)}_{\mathcal{A}_1} \times T^{(n)}_{\mathcal{A}_2}$$

and so

$$\bar{\delta}_{\mathcal{A}_1 \boxtimes \mathcal{A}_2}(G) \cap (I^{(m)}_{\mathcal{A}_1 \boxtimes \mathcal{A}_2} \times T^{(n)}_{\mathcal{A}_1 \boxtimes \mathcal{A}_2}) \neq 0 \quad \text{iff}$$

$$(\bar{\delta}_{\mathcal{A}_1}(G) \boxtimes \bar{\delta}_{\mathcal{A}_2}(G)) \cap ((I^{(m)}_{\mathcal{A}_1} \times T^{(n)}_{\mathcal{A}_1}) \times (I^{(m)}_{\mathcal{A}_2} \times T^{(n)}_{\mathcal{A}_2})) \neq 0 \quad \text{iff}$$

$$\bar{\delta}_{\mathcal{A}_1}(G) \cap (I^{(m)}_{\mathcal{A}_1} \times T^{(n)}_{\mathcal{A}_1}) \neq 0 \quad \text{and} \quad \bar{\delta}_{\mathcal{A}_2}(G) \cap (I^{(m)}_{\mathcal{A}_2} \times T^{(n)}_{\mathcal{A}_2}) \neq 0.$$

In other words $|\mathcal{A}_1 \boxtimes \mathcal{A}_2| = |\mathcal{A}_1| \cap |\mathcal{A}_2|$. □

Let us set $GR^+(\Sigma) = (GR_{m,n}(\Sigma))_{m,n>0}$.

Proposition 5. *If the languages $L_1, L_2 \subseteq GR^+(\Sigma)$ are recognizable, then so is their union $L_1 \cup L_2$.*

Proof. Let

$$\mathcal{A}_1 = (\Sigma, Q_1, \delta_{\mathcal{A}_1}, I_{\mathcal{A}_1}, T_{\mathcal{A}_1}) \quad \text{and} \quad \mathcal{A}_2 = (\Sigma, Q_2, \delta_{\mathcal{A}_2}, I_{\mathcal{A}_2}, T_{\mathcal{A}_2})$$

be graph automata such that $|\mathcal{A}_1| = L_1$, $|\mathcal{A}_2| = L_2$ and $Q_1 \cap Q_2 = \emptyset$. Consider the automaton

$$\mathcal{A} = (\Sigma, Q_1 \cup Q_2, \delta_{\mathcal{A}}, I_{\mathcal{A}_1} \cup I_{\mathcal{A}_2}, T_{\mathcal{A}_1} \cup T_{\mathcal{A}_2}),$$

where $\delta_{\mathcal{A}}$ is defined by

$$\delta_{\mathcal{A}}(\sigma) = \delta_{\mathcal{A}_1}(\sigma) \cup \delta_{\mathcal{A}_2}(\sigma) \quad \text{for all } \sigma \in \Sigma.$$

Using induction on the size of a graph G we show that

$$(u, v) \in \bar{\delta}_{\mathcal{A}}(G) \quad \text{and} \quad u \in Q_1^* \ \ (\text{resp. } u \in Q_2^*)$$

implies $v \in Q_1^*$ (resp. $v \in Q_2^*$).

Indeed, this implication holds for all elementary graphs $I_{12}, I_{10}, I_{21}, I_{01}, \Pi, \sigma \in \Sigma$. Furthermore, any other graph G can be written either $G = G_1 \circ G_2$ or $G = G_1 \square G_2$ with $size(G_i) < size(G)$, $i = 1, 2$. In the first case if

$$(u, v) \in \bar{\delta}_{\mathcal{A}}(G) = \bar{\delta}_{\mathcal{A}}(G_1) \circ \bar{\delta}_{\mathcal{A}}(G_2), \quad u \in Q_1^*$$

then there exists a state word $w \in (Q_1 \cup Q_2)^*$ such that

$$(u, w) \in \bar{\delta}_{\mathcal{A}}(G_1), \quad (w, v) \in \bar{\delta}_{\mathcal{A}}(G_2).$$

Applying successively the induction hypothesis for the graphs G_1 and G_2 we get $w \in Q_1^*$ and $v \in Q_1^*$ as well. In the second case we have

$$(u, v) \in \bar{\delta}_{\mathcal{A}}(G) = \bar{\delta}_{\mathcal{A}}(G_1 \square G_2), \quad u \in Q_1^*$$

implies that there are factorizations

$$u = u_1 u_2, \ \text{ with } u_1 u_2 \in Q_1^* \quad \text{and} \quad v = v_1 v_2 \ \text{ with } v_1 v_2 \in (Q_1 \cup Q_2)^*$$

Author's personal copy

such that

$$(u_1, v_1) \in \bar{\delta}_{\mathcal{A}}(G_1), \quad (u_2, v_2) \in \bar{\delta}_{\mathcal{A}}(G_2).$$

Again by the induction principle we get $v_1, v_2 \in Q_1^*$ and so $v = v_1 v_2 \in Q_1^*$. Consequently, $G \in |\mathcal{A}|$ means that the relation $\bar{\delta}_{\mathcal{A}}(G)$ contains either a pair $(u, v) \in I_{\mathcal{A}_1} \times T_{\mathcal{A}_1}$ or a pair $(u, v) \in I_{\mathcal{A}_2} \times T_{\mathcal{A}_2}$, and so $G \in |\mathcal{A}_1|$ or $G \in |\mathcal{A}_2|$ as wanted. □

Remark. The assumption $L_1, L_2 \subseteq GR^+(\Sigma)$ means that none of the two languages contains a graph of type either $(0, 0)$, $(0, k)$, $(k', 0)$, k, k' arbitrary. In the opposite case the automaton $\mathcal{A}$ of the above proof would also recognize graphs of the form $G_1 \square G_2$, where $G_1 \in L_1$ and $G_2 \in L_2 \cap GR_{0,0}(\Sigma)$ or vise versa. A similar phenomenon appears when $L_1 \cap GR_{0,k}(\Sigma) \neq \emptyset$ and $L_2 \cap GR_{k',0}(\Sigma) \neq \emptyset$.

Proposition 6. *The class $Rec(\Sigma)$ is not closed under complement provided that $\Sigma_{m,n} \neq \emptyset$ for some $m, n > 0$.*

Proof. First we observe that the graph automaton $\mathcal{A} = (\Sigma, \{0\}, \delta_{\mathcal{A}}, I_{\mathcal{A}}, T_{\mathcal{A}})$, with $\delta_{\mathcal{A}} = \emptyset$ and $I_{\mathcal{A}} = T_{\mathcal{A}} = 0^*$ recognizes the graph language *DISC* of all discrete graphs. Hence, *DISC* is recognizable.

Now assume that the behavior of the graph automaton

$$\mathcal{B} = (\Sigma, Q, \delta_{\mathcal{B}}, I_{\mathcal{B}}, T_{\mathcal{B}})$$

is the complement $DISC^c$ of *DISC*. Let $\sigma \in \Sigma_{1,1}$, then the graph (loop)

$$F = I_{1,2} \circ (\sigma \square E) \circ I_{2,1} \in DISC^c.$$

Thus, there exist $q_1 \in I_{\mathcal{B}}$ and $q_2 \in T_{\mathcal{B}}$ such that $(q_1, q_2) \in \bar{\delta}_{\mathcal{B}}(F)$. It holds:

$$\delta_{\mathcal{B}}(I_{12}) = \{(q, qq) \mid q \in Q\}, \qquad \delta_{\mathcal{B}}(E) = \{(q, q) \mid q \in Q\} \text{ and}$$
$$\delta_{\mathcal{B}}(I_{21}) = \{(qq, q) \mid q \in Q\}.$$

Hence $q_1 = q_2 = q$, with $q \in I_{\mathcal{B}}, T_{\mathcal{B}}$, and thus $E \in |\mathcal{B}|$, which is a contradiction.

It is easy to see that the above argument can be applied for every $\sigma \in \Sigma_{m,n}$, $m, n > 0$. □

Notice that the complement of a recognizable graph language lies always in the larger class of syntactically recognizable graph languages (see Section 6).

Proposition 7. *If the languages $L_1, L_2 \subseteq GR(\Sigma)$ are recognizable then so is $L_1 \square L_2$.*

Proof. Let

$$\mathcal{A}_i = (\Sigma, Q_i, \delta_{\mathcal{A}_i}, I_{\mathcal{A}_i}, T_{\mathcal{A}_i})$$

be graph automata such that $|\mathcal{A}_i| = L_i$, $(i = 1, 2)$ and $Q_1 \cap Q_2 = \emptyset$. We construct the graph automaton

$$\mathcal{A} = (\Sigma, Q_1 \cup Q_2, \delta_{\mathcal{A}}, I_{\mathcal{A}_1} I_{\mathcal{A}_2}, T_{\mathcal{A}_1} T_{\mathcal{A}_2})$$

with $\delta_{\mathcal{A}}(\sigma) = \delta_{\mathcal{A}_1}(\sigma) \cup \delta_{\mathcal{A}_2}(\sigma)$, for all $\sigma \in \Sigma$.

We are going to show that for all graphs $G \in GR_{m,n}(\Sigma)$ and words $u_1, v_1 \in Q_1^*$ and $u_2, v_2 \in Q_2^*$, if $(u_1 u_2, v_1 v_2) \in \bar{\delta}_{\mathcal{A}}(G)$ then

$$G = G_1 \square G_2 \quad \text{and} \quad (u_1, v_1) \in \bar{\delta}_{\mathcal{A}_1}(G_1), \quad (u_2, v_2) \in \bar{\delta}_{\mathcal{A}_2}(G_2).$$

Clearly this implication holds true for any discrete graph D.

For the general case we need Theorem 2(ii) (see also [6]) i.e., every graph G can be factorized in the form

$$G = \Delta \circ (E_s \square BF(\sigma_1) \square \cdots \square BF(\sigma_k)),$$

where Δ is a discrete graph and $BF(\sigma)$ $(\sigma \in \Sigma_{m,n})$ is the following graph of rank $(m + n, 0)$:

$$BF(\sigma) = (\sigma \square E_n) \circ \Pi_{\alpha_n} \circ I_{20}^n. \qquad \text{(b)}$$

Author's personal copy

Recall that I_{20}^n is the discrete graph introduced at the end of Section 2 and that $\alpha_n \in \mathcal{S}_{2n}$ is defined by $\alpha_n(i) = 2i - 1$, $\alpha_n(n+i) = 2i$, for $1 \leq i \leq n$.

Now if $w \in (Q_1 \cup Q_2)^*$ is such that

$$(u_1u_2, w) \in \bar{\delta}_{\mathcal{A}}(D)$$

then

$$(w, v_1v_2) \in \bar{\delta}_{\mathcal{A}}(E_s \square BF(\sigma_1) \square \cdots \square BF(\sigma_k))$$

and we get that $w = v_1v_2w'$ with $w' \in (Q_1 \cup Q_2)^*$ and

$$(w', \varepsilon) \in \bar{\delta}_{\mathcal{A}}(BF(\sigma_1) \square \cdots \square BF(\sigma_k)).$$

Taking into account Eq. (b) as well as the fact that $\delta_{\mathcal{A}}(\sigma) = \delta_{\mathcal{A}_1}(\sigma) \cup \delta_{\mathcal{A}_2}(\sigma)$ we obtain that

$$\bar{\delta}_{\mathcal{A}}(BF(\sigma_i)) = \{(t, \varepsilon) \mid t \in Q_1^{m_i+n_i} \text{ or } t \in Q_2^{m_i+n_i}\},$$

where $rank(\sigma_i) = (m_i, n_i)$, $(i = 1, 2)$.

Now we recall the following equality from [6] (Lemma 6):

$$\Pi_{m,n} \circ (A \square B) = (B \square A) \circ \Pi_{k,l},$$

where $\Pi_{m,n}$, is the permutation graph associated with the permutation $\alpha_{m,n}$ which interchanges the last m numbers with the first n one's:

$$\alpha_{m,n} = \begin{pmatrix} 1 & \cdots & n & n+1 & \cdots & m+n \\ m+1 & \cdots & m+n & 1 & \cdots & m \end{pmatrix}$$

and $A \in GR_{m,k}(\Sigma)$ and $B \in GR_{n,l}(\Sigma)$. We point out that this equation for $l = 0$ takes the form:

$$\Pi_{m,n} \circ (A \square B) = (B \square A) \circ \Pi_{k,0} \Leftrightarrow \Pi_{m,n} \circ (A \square B) = (B \square A).$$

Since $\Pi_{k,0} = E_k$. Applying repeatedly this equation we can write

$$\begin{aligned} G = D \circ (E_s \square BF(\sigma_1) \square \cdots \square BF(\sigma_k)) &= D \circ \Pi_\gamma \\ &\circ [E_{s_1} \square BF(\sigma_{i_1}) \square \cdots \square BF(\sigma_{i_r}) \square E_{s_2} \square BF(\sigma_{j_1}) \square \cdots \square BF(\sigma_{j_s})], \end{aligned}$$

where Π_γ is a permutation graph and it holds $(u_1u_2, v_1w_1v_2w_2) \in \bar{\delta}_{\mathcal{A}}(D \circ \Pi_\gamma)$ with $w_i \in Q_i^*$ $(i = 1, 2)$.

Since $u_1, v_1w_1 \in Q_1^*$ and $u_2, v_2w_2 \in Q_2^*$ and $D \circ \Pi_\gamma$ is a discrete graph, we get that

$$D \circ \Pi_\gamma = D_1 \square D_2 \quad \text{and} \quad (u_1, v_1w_1) \in \bar{\delta}_{\mathcal{A}}(D_1), (u_2, v_2w_2) \in \bar{\delta}_{\mathcal{A}}(D_2)$$

with D_1, D_2 discrete graphs. It turns out that

$$\begin{aligned} G &= (D_1 \square D_2) \circ [E_{s_1} \square BF(\sigma_{i_1}) \square \cdots \square BF(\sigma_{i_r}) \square E_{s_2} \square BF(\sigma_{j_1}) \square \cdots \square BF(\sigma_{j_s})] \\ &= [D_1 \circ (E_{s_1} \square BF(\sigma_{i_1}) \square \cdots \square BF(\sigma_{i_r}))] \square [D_2 \circ (E_{s_2} \square BF(\sigma_{j_1}) \square \cdots \square BF(\sigma_{j_s}))] = G_1 \square G_2 \end{aligned}$$

with $(u_1, v_1) \in \bar{\delta}_{\mathcal{A}}(G_1)$, $(u_2, v_2) \in \bar{\delta}_{\mathcal{A}}(G_2)$, $u_1, v_1 \in Q_1^*$, $u_2, v_2 \in Q_2^*$ or equivalently $(u_1, v_1) \in \bar{\delta}_{\mathcal{A}_1}(G_1)$, $(u_2, v_2) \in \bar{\delta}_{\mathcal{A}_2}(G_2)$ as desired.

Now let $G \in |\mathcal{A}|$. Then there are words $u \in I_{\mathcal{A}_1}I_{\mathcal{A}_2}$ and $v \in T_{\mathcal{A}_1}T_{\mathcal{A}_2}$ so that $(u, v) \in \bar{\delta}_{\mathcal{A}}(G)$. Hence $u = u_1u_2$, $v = v_1v_2$ with $u_1 \in I_{\mathcal{A}_1}$, $v_1 \in T_{\mathcal{A}_1}$, $u_2 \in I_{\mathcal{A}_2}$, $v_2 \in T_{\mathcal{A}_2}$. According to our previous argument, $G = G_1 \square G_2$ and

$$(u_1, v_1) \in \bar{\delta}_{\mathcal{A}_1}(G_1), \quad (u_2, v_2) \in \bar{\delta}_{\mathcal{A}_2}(G_2)$$

and so $G_1 \in |\mathcal{A}_1|$, $G_2 \in |\mathcal{A}_2|$, thus $|\mathcal{A}| \subseteq |\mathcal{A}_1| \square |\mathcal{A}_2|$. The opposite inclusion is obvious. □

Proposition 8. *The class $Rec(\Sigma)$ is not closed under $\circ$-product.*

Author's personal copy

Proof. We recall the language L_p of Example 2 and let $\mathcal{A} = (\Sigma, Q, \delta_{\mathcal{A}}, I_{\mathcal{A}}, T_{\mathcal{A}})$ be a graph automaton computing the language $L_p \circ L_p$. The graph $\sigma \circ \sigma \in L_p \circ L_p$ and thus there exist $q_1 \in I_{\mathcal{A}}$ and $q_2 \in T_{\mathcal{A}}$ such that $(q_1, q_2) \in \bar{\delta}_{\mathcal{A}}(\sigma \circ \sigma)$. Now let H be the graph

$$I_{1,2} \circ ((\sigma \circ \sigma) \square (\sigma \circ \sigma)) \circ I_{2,1}.$$

Then it holds $(q_1, q_2) \in \bar{\delta}_{\mathcal{A}}(H)$ and thus $H \in |\mathcal{A}|$, which is a contradiction since $H \notin L_p \circ L_p$. □

Proposition 9. *The class $Rec(\Sigma)$ is closed under inverse graph homomorphisms and projections.*

Proof. Let

$$\mathcal{A} = (\Sigma, Q, \delta_{\mathcal{A}}, I_{\mathcal{A}}, T_{\mathcal{A}})$$

be a graph automaton with $|\mathcal{A}| = L$. Given a graph homomorphism $H : GR(\Gamma) \to GR(\Sigma)$ the graph automaton

$$H^{-1}(\mathcal{A}) = (\Gamma, Q, \delta_{H^{-1}(\mathcal{A})}, I_{\mathcal{A}}, T_{\mathcal{A}})$$

with $\delta_{H^{-1}(\mathcal{A})}(\gamma) = \bar{\delta}_{\mathcal{A}}(H(\gamma))$, $\gamma \in \Gamma$, recognizes the graph language $H^{-1}(|\mathcal{A}|) = H^{-1}(L)$.

Finally assume that $H : GR(\Sigma) \to GR(\Gamma)$ is a projection. Then for the graph automaton

$$H(\mathcal{A}) = (\Gamma, Q, \delta_{H(\mathcal{A})}, I_{\mathcal{A}}, T_{\mathcal{A}}), \quad \text{with } \delta_{H(\mathcal{A})}(\gamma) = \bigcup_{H(\sigma)=\gamma} \delta_{\mathcal{A}}(\sigma),$$

it holds $|H(\mathcal{A})| = H(|\mathcal{A}|) = H(L)$. □

Example 3. If $L \subseteq DISC$ is recognizable then the set of all graphs $G \in GR(\Sigma)$ whose underlying discrete graph belongs to L, that is the set $disc_{\Sigma}^{-1}(L)$, is recognizable as well.

6. Comparison with syntactic recognizability

In this section we connect automata recognizability with syntactic recognizability as it is presented in [7].

Let $\xi_{m,n}$ be a new symbol with rank (m, n) and denote by $FR_{m,n}^{\alpha,\beta}(\Sigma)$ the subset of $GR_{\alpha,\beta}(\Sigma \cup \{\xi_{m,n}\})$ with just one occurrence of $\xi_{m,n}$. The elements of $FR_{m,n}^{\alpha,\beta}(\Sigma)$ are called frames with exterior rank (α, β) and interior rank (m, n). The set $FR_{m,n}^{\alpha,\beta}(\Sigma)$ acts on $GR_{m,n}(\Sigma)$ via substitution at $\xi_{m,n}$: for $F \in FR_{m,n}^{\alpha,\beta}(\Sigma)$ and $G \in GR_{m,n}(\Sigma)$,

$$F \cdot G = F[G/\xi_{m,n}].$$

An equivalence $\sim = (\sim_{m,n})$ on the magmoid $M = (M_{m,n})$ is a congruence whenever, for all $m, n \geqslant 0$, $f, g \in M_{m,n}$ and all $\omega \in FR_{m,n}^{\alpha,\beta}(M)$

$$f \sim_{m,n} g \quad \text{implies } \omega[f] \sim_{\alpha,\beta} \omega[g].$$

Let L be a subset of the magmoid M and $f \in M_{m,n}$, we set

$$C_L(f) = \{\omega \mid \omega \in FR_{m,n}^{\alpha,\beta}(M),\ \omega[f] \in L\}.$$

The equivalence $\sim_L$ on M defined by

$$f \sim_{L,m,n} g, \quad \text{whenever } C_L(f) = C_L(g)$$

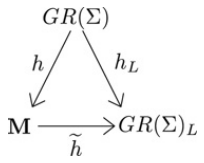
is a congruence.

Given a magmoid M and a set $L \subseteq M$, $\sim_L$ is called the *syntactic congruence* of L and the quotient magmoid $M_L = M/\sim_L$ is the *syntactic magmoid* of L. Since the epimorphic image of a graphoid is again a graphoid, the syntactic magmoid of a graph language $L \subseteq GR(\Sigma)$ is a graphoid.

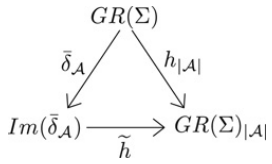
The next universal property characterizes $GR(\Sigma)_L$ up to isomorphism.

Author's personal copy

Theorem 4 (*cf. [7]*). *Let $L \subseteq GR(\Sigma)$ and $h : GR(\Sigma) \to \mathbf{M}$ an epimorphism of graphoids such that $h^{-1}(h(L)) = L$. Then there is an epimorphism of graphoids $\widetilde{h} : \mathbf{M} \to GR(\Sigma)_L$, making commutative the next triangle.*



In a graph automaton $\mathcal{A} = (\Sigma, Q, \delta_{\mathcal{A}}, I_{\mathcal{A}}, T_{\mathcal{A}})$ the *transition relation* corresponding to the graph $G \in GR_{m,n}(\Sigma)$ is $\bar{\delta}_{\mathcal{A}}(G) \subseteq Q^m \times Q^n$. Applying the previous remark to the triangle



we get the following proposition.

Proposition 10. *If the graphs G_1, G_2 define the same transition relation, then they have the same contexts with respect to $|\mathcal{A}|$, i.e.,*

$$\bar{\delta}_{\mathcal{A}}(G_1) = \bar{\delta}_{\mathcal{A}}(G_2) \quad \textit{implies } C_{|\mathcal{A}|}(G_1) = C_{|\mathcal{A}|}(G_2).$$

Recognizability at the level of magmoids is defined by suitably adapting the corresponding notion for monoids. We say that a congruence $\sim = (\sim_{m,n})$ on the magmoid $GR(\Sigma)$ *saturates* $L \subseteq GR(\Sigma)$ whenever, for all $m, n \geq 0$, the subset $L_{m,n}$ is a union of $\sim_{m,n}$-classes. If for all $m, n \geq 0$, the congruence $\sim_{m,n}$ has finite index (i.e., finite number of equivalence classes) we say that $\sim$ has locally finite index.

A subset L of $GR(\Sigma)$ is called syntactically recognizable if there exists a locally finite magmoid $N = (N_{m,n})$ (i.e., $N_{m,n}$ finite for all $m, n \in \mathbb{N}$) and a morphism $h : GR(\Sigma) \to N$, so that $L = h^{-1}(P)$, for some $P \subseteq N$.

The class of all syntactically recognizable subsets of $GR(\Sigma)$ is denoted by $SRec(\Sigma)$.

Theorem 5 (*cf. [7]*). *Let $L \subseteq GR(\Sigma)$; the following conditions are equivalent:*

(i) *L is syntactically recognizable,*
(ii) *L is saturated by a congruence of a locally finite index,*
(iii) *$\sim_L$ has locally finite index,*
(iv) *the set $card\{C_L(f) \mid f \in GR(\Sigma)_{m,n}\}$ is finite for all $m, n \in \mathbb{N}$,*
(v) *the syntactic magmoid $GR(\Sigma)_L$ is locally finite,*
(vi) *there exist a locally finite, graphoid $\mathbf{N}$ and a morphism of graphoids $h : GR(\Sigma) \to \mathbf{N}$ so that $L = h^{-1}(P)$, for some $P \subseteq \mathbf{N}$.*

Corollary 1. *The class $SRec(\Sigma)$ is closed under Boolean operations and inverse graph homomorphisms.*

Proposition 11. *The graph language $CON(\Sigma)$ of all connected graphs over Σ is not recognizable.*

Proof. Suppose that $\mathcal{A} = \{\Sigma, Q, \delta_{\mathcal{A}}, I_{\mathcal{A}}, T_{\mathcal{A}}\}$ is a graph automaton accepting $CON(\Sigma)$. The graph $I_{21} \circ I_{12}$ is connected and thus $I_{21} \circ I_{12} \in |\mathcal{A}|$. Hence, there exist states $q_1, q_2, q_3, q_4 \in Q$, with $q_1q_2 \in I_{\mathcal{A}}$, $q_3q_4 \in T_{\mathcal{A}}$ such that

$$(q_1q_2, q_3q_4) \in \bar{\delta}_{\mathcal{A}}(I_{21} \circ I_{12}).$$

Moreover, there exists a state $q_2' \in Q$ such that $(q_1q_2, q_2') \in \delta_{\mathcal{A}}(I_{21})$ and since $\delta_{\mathcal{A}}(I_{21}) = \{(qq, q) \mid q \in Q\}$, we deduce that $q_1 = q_2 = q_2'$. Using similar arguments for I_{12} it follows that $q_1 = q_2 = q_3 = q_4$. Hence there exists an element $q \in Q$ such that $qq \in I_{\mathcal{A}}$ and $qq \in T_{\mathcal{A}}$ but

$$(qq, qq) \in \bar{\delta}_{\mathcal{A}}(E \,\square\, E)$$

which is a contradiction because the graph $E \,\square\, E$ is not connected. $\square$

Author's personal copy

Remark. An element a of a magmoid M is said to be connected if it cannot be factorized as $a = a_1 \square a_2$ for $a_1, a_2 \in M - \{e_0\}$. It is easily seen that $val_{\Sigma}^{-1}(CON(\Sigma))$ is just the set of all connected patterns of $smag(\Sigma \cup D)$ which is shown to be recognizable in [5]. However according to Proposition 11 and Proposition 3 this graph language cannot be recognized by a pattern automaton with graph structure.

Proposition 12. *The class of recognizable graph languages is properly included in the class of syntactically recognizable graph languages.*

$$Rec(\Sigma) \subsetneq SRec(\Sigma).$$

Proof. As we have seen the behavior of any graph automaton

$$\mathcal{A} = (\Sigma, Q, \delta_{\mathcal{A}}, I_{\mathcal{A}}, T_{\mathcal{A}})$$

is given by $|\mathcal{A}| = \bar{\delta}_{\mathcal{A}}^{-1}(\Theta)$, where $\Theta = (\Theta_{m,n})$ is the doubly ranked set defined by $\Theta_{m,n} = (I_{\mathcal{A}} \cap Q^m) \times (T_{\mathcal{A}} \cap Q^n)$. Thus, the inclusion follows by taking into account that $\mathbf{Rel}(Q)$ is a locally finite graphoid.

The proper inclusion assertion follows by Proposition 11. □

As we have seen in the previous section $Rec(\Sigma)$ fails to be closed under complement. However,

Proposition 13. *The complement of any recognizable graph language is a syntactically recognizable graph language.*

Proof. We only have to combine Corollary 1 and Proposition 12. □

7. Conclusion and future work

The introduction of graph automata seems to be one step towards the construction of a powerful and robust graph language recognizability theory. In this perspective, it is natural to examine the next subjects:

1. Graph automata and monadic second-order logic.
2. Infinite behavior of finite graph automata.
3. Weighted graph automata and formal series on graphs.
4. Graph transducers.

As another future research direction we point out that the definition of a graph automaton $\mathcal{A}$ requires that the set of relations $Rel(Q)$ over the state set Q of $\mathcal{A}$ should be a graphoid. Due to the importance of recognizing graph languages by means of automata the question that arises is whether we can give to Q an extra algebraic structure so that $Rel(Q)$ becomes a graphoid. In this way new recognizability classes of graph languages will appear.

Acknowledgement

We deeply thank the referees for their numerous comments and suggestions.

References

[1] M.A. Arbib, Y. Give'on, Algebra automata I: Parallel programming as a prolegomena to the categorical approach, Inform. and Control 12 (1968) 331–345.
[2] A. Arnold, M. Dauchet, Théorie des magmoides. I, RAIRO Inform. Théor. 12 (3) (1978) 235–257.
[3] A. Arnold, M. Dauchet, Théorie des magmoides. II, RAIRO Inform. Théor. 13 (2) (1979) 135–154.
[4] S.L. Bloom, Z. Ésik, Iteration theories, in: EATCS Monographs on Theoretical Computer Science, Springer-Verlag, 1991.
[5] F. Bossut, M. Dauchet, B. Warin, A Kleene theorem for a class of planar acyclic graphs, Inform. and Comput. 117 (1995) 251–265.
[6] S. Bozapalidis, A. Kalampakas, An axiomatization of graphs, Acta Inform. 41 (2004) 19–61.
[7] S. Bozapalidis, A. Kalampakas, Recognizability of graph and pattern languages, Acta Inform. 42 (2006) 553–581.
[8] V. Claus, Ein Vollständigkeitssatz für programme und schaltkreise, Acta Inform. 1 (1971) 64–78.
[9] H. Ehrig, G. Engels, H.-J. Kreowski, G. Rozenberg (Eds.), Handbook of Graph Grammars and Computing by Graph Transformation, vol. I, World Scientific, Singapore, 1997.
[10] S. Eilenberg, J.B. Wright, Automata in general algebras, Inform. and Control 11 (1967) 452–470.
[11] J. Engelfriet, Context-free graph grammars, in: G. Rozenberg, A. Salomaa (Eds.), Handbook of Formal Languages, vol. III: Beyond Words, Springer, 1997, pp. 125–213 (Chapter 3).

Author's personal copy

[12] J. Engelfriet, J.J. Vereijken, Context-free graph grammars and concatenation of graphs, Acta Inform. 34 (1997) 773–803.
[13] Z. Esik, A variety theorem for trees and theories, Publ. Math. 54 (1999) 711–762.
[14] F. Gécseg, M. Steinby, Tree Automata, Akademiai Kiado, Budapest, 1984.
[15] J. Gibbons, An initial-algebra approach to directed acyclic graphs, in: Mathematics of Program Construction (Kloster Irsee, 1995), in: LNCS, vol. 947, Springer, Berlin, 1995, pp. 282–303.
[16] G. Hotz, Eine algebraisierung des syntheseproblems von schaltkreisen, EIK 1 (1965) 185–205. 209–231.
[17] G. Hotz, Eindeutigkeit und mehrdeutigkeit formaler sprachen, EIK 2 (1966) 235–246.
[18] T. Kamimura, G. Slutzki, Parallel and two-way automata on directed ordered acyclic graphs, Inform. and Control 49 (1981) 10–51.
[19] T. Kamimura, G. Slutzki, Transductions of DAGS and trees, Math. Syst. Theory 15 (1982) 225–249.
[20] S. MacLane, Categories for the Working Mathematician, Springer Verlag, 1971.
[21] C.P. Schnorr, Transformational classes of grammars, Inform. and Control 14 (1969) 252–277.