

# **Reconnaissabilité pour les langages de graphes et de posets**

**Pascal Weil**

(LaBRI, CNRS, Université Bordeaux-1)

Bordeaux, novembre 2002

## Graphs and posets (finite): representation

$A$  a finite alphabet

Directed  $A$ -labeled graphs:  $G = (V, E, \lambda)$  with  $E \subseteq V \times V$  and  $\lambda: V \rightarrow A$

For our purpose, self-loops are irrelevant: assume that  $E$  is anti-reflexive (if necessary, the presence of a self-loop at vertex  $x$  can be encoded in the letter labeling  $x$ )

Dag: directed acyclic graph

Poset  $(V, <)$  : a transitive dag with  $<$  as the edge relation

Dag  $\rightarrow$  Poset: transitive closure

Minimal representation: the *Hasse diagram* of  $(V, <)$ , given by the **succ** relation; the only anti-transitive, anti-reflexive dag, whose transitive closure is  $<$

## Logics

$MS$  = Monadic second order logic, on the structure  $(V, E)$ : quantification is on (sets of) vertices,  $E$  is a binary predicate

$CMS$  = Counting monadic second order logic: this is  $MS$ , enriched with special quantifiers of the form  $\exists^{\text{mod } q} x \varphi(x)$ , interpreted to mean that the set of values of  $x$  such that  $\varphi(x)$  holds is  $0 \bmod q$

$CMS$  is strictly more expressive than  $MS$ :  $MS$  cannot express even size

## Algebraic framework

Necessary to discuss notions such as recognizability, equationality of a graph language  $L$

Endow the set of all finite  $A$ -labeled graphs with a structure of  $\mathcal{F}$ -algebra, for a well-chosen signature  $\mathcal{F}$

(Roughly)  $L$  is  $\mathcal{F}$ -recognizable if it is a union of classes in a finite index congruence of the  $\mathcal{F}$ -algebra of graphs (Mezei, Wright, 1967)

## Several signatures in the literature

*VR*-signature: for *graphs with ports*, decorated with finitely many flags from a set  $\mathcal{C}$ ; operations are  $\oplus$  (disjoint union, binary), **add** <sub>$c,d$</sub>  (unary), **mdf** <sub>$P$</sub>  (for  $P \subset_{fin} \mathcal{C} \times \mathcal{C}$ , unary)

*HR*-signature: for graphs with sources, decorated with flags from a set  $\mathcal{C}$ , each flag used at most once; operations are  $\parallel_{\mathcal{C},\mathcal{C}'}$  ( $\mathcal{C},\mathcal{C}'$  finite subsets of  $\mathcal{C}$ , parallel composition, binary), source renaming (wrt a given selfmap of  $\mathcal{C}$ , unary), source fusion (wrt a given equivalence relation on  $\mathcal{C}$ , unary)

*modular* signature: for graphs without flags; the operations are derived from the modular decomposition of graphs, precise definitions will be given

From results of Courcelle, Engelfriet, Olariu:

$$CMS\text{-def} \implies VR\text{-recog}$$

$$CMS_2\text{-def} \implies HR\text{-recog}$$

(quantification on vertex sets and on edge sets)

$$CMS_2\text{-def} \Leftarrow HR\text{-recog}$$

for languages of tree-width  $\leq k$  (Courcelle, Lapoire)

## Composition of graphs

$H = ([n], F)$  defines an  $n$ -ary operation on graphs

$$H = \quad H\langle G_1, \dots, G_4 \rangle =$$

Particular cases:

$$H_{\oplus} = \quad G_1 \oplus G_2 = G_1 \sqcup G_2$$

$$H_{\bullet} = \quad G_1 \bullet G_2 = G_1 \sqcup G_2 \text{ with all arrows from } V_1 \text{ to } V_2$$

$$H_{\otimes} = \quad G_1 \otimes G_2 = G_1 \sqcup G_2 \text{ with all arrows from } V_1 \text{ to } V_2 \text{ and back (does not preserve posets)}$$

If  $H = K\langle L_1, \dots, L_r \rangle$ , then  $H\langle G_1, \dots, G_n \rangle = K\langle L'_1, \dots, L'_r \rangle$  and  $L'_j = L_j\langle G_{i_{j,1}}, \dots, G_{i_{j,r_j}} \rangle$   
 associativity of substitution

Use only prime (indecomposable) graphs as operations

## Modular decomposition of graphs

$$G = (V, E)$$

Modules:  $X \subseteq V$   
 interacting  
 uniformly with  $V \setminus X$

Prime module: a module  $X$  s.t. for every module  $Y$ , either  $Y \subseteq X$  or  $X \subseteq Y$  or  $X \cap Y = \emptyset$

If  $X_1, \dots, X_n$  are the maximal proper modules, then  $G = H \langle X_1, \dots, X_n \rangle$  for

$$H = 1 \longrightarrow 2 \longrightarrow \dots \longrightarrow n \quad \text{or}$$

$$H = 1 \quad 2 \quad \dots \quad n \quad \text{or}$$

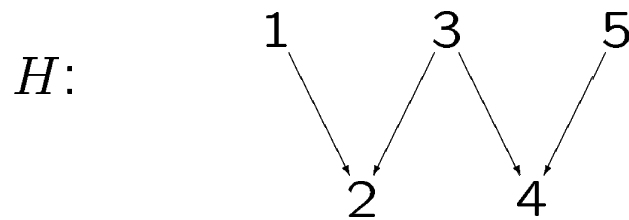
$$H = 1 \begin{array}{c} \longrightarrow \\ \longleftarrow \end{array} 2 \begin{array}{c} \longrightarrow \\ \longleftarrow \end{array} \dots \begin{array}{c} \longrightarrow \\ \longleftarrow \end{array} n \quad \text{or}$$

$$H = ([n], F) \text{ is a prime graph}$$



Thus each graph can be expressed from singletons using the composition operations associated with prime graphs,

This expression, the *modular decomposition*, is unique up to associativity of  $\bullet, \oplus, \otimes$  and commutativity of  $\oplus, \otimes$



$$H\langle G_1, G_2, G_3, G_4, G_5 \rangle = H\langle G_5, G_4, G_3, G_2, G_1 \rangle$$

The modular signature,  $\mathcal{F}_\infty$  is infinite

$A$ -labeled graphs = the  $\mathcal{F}_\infty$ -algebra generated by letter-labeled singletons =  $A$ -generated  $\mathcal{F}_\infty$ -algebra

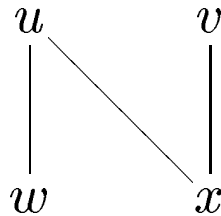
**Consider finite subsignatures  $\mathcal{F} \subseteq \mathcal{F}_\infty$**

$\mathcal{F} = \{\bullet\}$   $A$ -generated  $\mathcal{F}$ -algebra:  $A$ -labeled graphs with vertices  $\{1, \dots, n\}$  and edges of the form  $(i, j)$  with  $i < j$  = the free semigroup  $A^+$

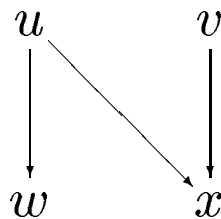
$\mathcal{F} = \{\oplus\}$   $A$ -generated  $\mathcal{F}$ -algebra: finite  $A$ -labeled discrete graphs (no edges) = the free commutative semigroup on  $A$

$\mathcal{F} = \{\otimes\}$   $A$ -generated  $\mathcal{F}$ -algebra = finite  $A$ -labeled cliques (also the free commutative semigroup on  $A$ )

$\mathcal{F} = \{\oplus, \otimes\}$  The  $\mathcal{F}$ -graphs are the *cographs* =  $P_4$ -free undirected graphs – where  $P_4$  is



$\mathcal{F} = \{\bullet, \oplus\}$  The  $\mathcal{F}$ -graphs are the *series-parallel posets* =  $N$ -free directed graphs – where  $N$  is



## Characterization of $\mathcal{F}$ -recognizability

(Courcelle 1997)

$\mathcal{F}$  finite subset of  $\mathcal{F}_\infty$ ,  $L$  a  $\mathcal{F}$ -graph language:

$L$   $CMS$ -definable  $\implies L$   $\mathcal{F}$ -recognizable

$L$   $\mathcal{F}_\infty$ -recognizable

a certain tree-like normal form  
language is  $CMS$ -definable

Courcelle shows more:

$\mathcal{F}$ -recognizable  $\iff MS_{lin}$ -definable

## Some special cases

$\mathcal{F} = \{\bullet\}$ : Büchi's theorem on the free semigroup; recognizability  $\iff MS$ -definability

$\mathcal{F} = \{\oplus\}$ : discrete graphs = free commutative semigroup; recognizability  $\iff CMS$ -definability

$\mathcal{F} = \{\otimes\}$ : same thing

$\mathcal{F} = \{\bullet, \oplus\}$ : *sp*-languages;  
recognizability  $\iff CMS$ -definability (Kuske '02)

For *sp*-languages of bounded width (= with a uniform bound on the size of an antichain):

*CMS*-def is equivalent to *MS*-def (Courcelle);

recognizability is characterized by a branching automaton model and by *series-rational expressions* (union,  $\bullet$ ,  $\oplus$ ,  $\bullet$ -iteration, i.e. Kleene star (Lodaya-Weil))

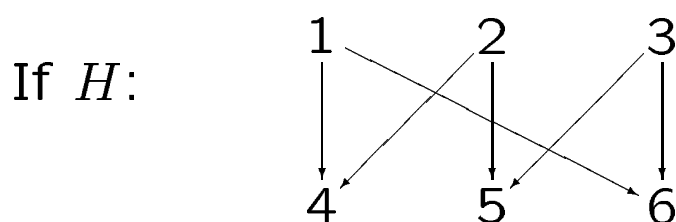
Weil: extension of this equivalence to larger signatures  $\mathcal{F} \subseteq \mathcal{F}_\infty$

In Kuske's and other proofs: use the fact that  $\bullet$ -nodes in a tree (term) decomposition of an  $sp$ -poset have a notion of first-child

## Weakly rigid operations

$H = ([n], F)$  a prime graph

If  $H = C_n$ , no child of an  $H$ -node is distinguished



$Aut(H) = \langle (123)(456), (12)(56) \rangle$  does not act transitively on  $[n]$ : there are two orbits; distinguish  $\{1, 2, 3\}$  as the **first** nodes, and  $\{4, 5, 6\}$  as the **last** ones

If  $H$  is a dag and not discrete, the  $H$ -product is weakly rigid

**Weakly rigid signature:** a finite subset of  $\mathcal{F}_\infty$ , consisting of at most one of  $\oplus$  and  $\otimes$ , and of weakly rigid operations  $H = ([n], F)$ ,  
 $+$

for each such  $H$ , the designation of a distinguished set of children (an orbit of  $[n]$  under  $\text{Aut}(H)$ )

Ex. Any finite signature consisting of dags  
(distinguished children: say, the maximal ones);

Any signature consisting of  $\oplus$  and prime graphs with non-uniform in- (resp. out-) degree

**Theorem:** For a weakly rigid signature  $\mathcal{F}$

$\mathcal{F}$ -recognizability  $\iff$  CMS-definability

## Hasse diagram representation of posets

the modular decomposition does not help

for each operation in  $\mathcal{F}_\infty$ , say

$$(G_1, \dots, G_n) \longmapsto f(G_1, \dots, G_n),$$

define  $f^h: (\mathbf{hasse}(G_1), \dots, \mathbf{hasse}(G_n)) \longmapsto \mathbf{hasse}(f(G_1, \dots, G_n))$

In order to preserve properties such as

$$CMS\text{-definability} \Rightarrow \mathcal{F}\text{-recognizability},$$

it is necessary to consider **enriched** graph structures:  $S = (V, E, m, M)$  where  $E$  is an anti-reflexive, anti-transitive, antisymmetric binary relation,  $m$  and  $M$  are unary, representing minimal and maximal elements



**Example:** if  $f$  is the binary  $\bullet$ -product,  $S \bullet^h S'$  is the disjoint union of  $S, S'$ , with edges from  $M$  to  $m'$ , new minimum predicate  $m$ , new maximal predicate  $M'$

The operations  $f^h$  can be expressed using disjoint unions and quantifier-free operations

If  $L$  is a poset language and  $L^h = \mathbf{hasse}(L)$ , if  $\mathcal{F} \subset_{fin} \mathcal{F}_\infty$  and  $\mathcal{F}^h = \{f^h \mid f \in \mathcal{F}\}$ , we have

$L$  is CMS-definable  $\iff L^h$  is CMS-definable

$L$  is  $\mathcal{F}$ -recognizable  $\iff L^h$  is  $\mathcal{F}^h$ -recognizable

## Quelques perspectives

Liens entre  $\mathcal{F}$ -,  $HR$ - et  $VR$ -reconnaissabilité

$FO$ -définissabilité pour les  $\mathcal{F}$ -graphes ( $sp$ -posets)

*po-traces*