

# The tight relationships of Recognizability and Monadic Second-Order Logic

based on article  
Recognizability and Hypergraph operations  
using local informations

by Achim Blumensath and Bruno Courcelle

LaBRI, Bordeaux 1 University  
GAMES Research Training Network

## Motivations and overview

F : signature of operations on graphs, hypergraphs, *i.e.*,  
relational structures

- F-algebra
- Recognizable sets  $REC(F)$  (finite congruences, no automata)
- Equational sets  $EQ(F)$  (« context-free » ; equation systems, no rewriting rules)

## Robustness :

$F \equiv G$  means  $REC(F) = REC(G)$  and  $EQ(F) = EQ(G)$ .

- **Equivalences** between small and large signatures
- **Extension to particular operations** like **fusion** and other **non quantifier-free** definable operations
- « Sources » in graphs and hypergraphs **are dispensable**

## Relationships with MS logic :

- MS-definable  $\subseteq$  REC
- MS-transductions preserve EQ
- **New result : inverse MS-transductions preserve REC.**

# 1. Background : The algebra of relational structures

$\Sigma_\infty$  = fixed countable set of relation symbols with arity

$STR(\Sigma)$  = finite relational  $\Sigma$ -structures, for  $\Sigma$  finite  $\subseteq \Sigma_\infty$

The fundamental signature  $QF$  :

Binary operation(s) :

disjoint union  $\oplus$  :  $STR(\Sigma) \times STR(\Gamma) \rightarrow STR(\Sigma \cup \Gamma)$

Unary operations :  $STR(\Sigma) \rightarrow STR(\Gamma)$

defined by quantifier-free formulas :

- domain restriction (ex. delete all isolated vertices)
- redefinition of relations (ex. edge-complement, addition of new edges based on vertex labels)

Nullary symbol **1** : designating the structure in  $STR(\emptyset)$  with a single element.

$STR$  the many-sorted  $QF$ -algebra with domains  $STR(\Sigma)$  ;  
each finite  $\Sigma$  is a sort.

$T(QF)_\Sigma$  = finite terms over  $QF$  (of sort  $\Sigma$ ).

Value mapping  $val$  :  $T(QF)_\Sigma \rightarrow STR(\Sigma)$ .

Reminder :  $S$  countable set of sorts,  $F$  an  $S$ -signature  
(means that each  $f$  in  $F$  has a type

$$s_1 s_2 \dots s_k \rightarrow s, \text{ with } s, s_i \in S)$$

$\mathbf{M} = \langle (M_s)_{s \in S}, (f_M)_{f \in F} \rangle$  an  $F$ -algebra,  $M_s \cap M_t = \emptyset$ , if  $s \neq t$

$$\text{where } f_M : M_{s_1} \times M_{s_2} \times \dots \times M_{s_k} \rightarrow M_s$$

## Equational sets

Context-Free (Graph) Grammars in an algebraic setting.

*Reminder* : For words the set of context-free rules :

$$S \rightarrow a S T ; \quad S \rightarrow b ; \quad T \rightarrow c T T T ; \quad T \rightarrow a$$

is equivalent to the system of two set equations:

$$S = a S T \quad \cup \quad \{ b \}$$

$$T = c T T T \quad \cup \quad \{ a \}$$

where  $S$  is the language generated by  $S$  (idem for  $T$  and  $T$ ).

For graphs we consider similarly systems of equations like:

$$S = f( k( S ), T ) \quad \cup \quad \{ b \}$$

$$T = f( T , f( g(T), m( T ))) \cup \{ a \}$$

where  $f$  is a binary operation,  $g, k, m$  are unary operations on graphs,  $a, b$  are basic graphs.

An *equational set* is a component of the least (unique) solution of such an equation system.

There are two sets of graph operations, two classes of context-free sets of graphs, the **HR-context-free sets** (for Hyperedge-Replacement) and the **VR-context-free sets** (for Vertex-Replacement).

## Equational sets in STR

$EQ(\Sigma)$  = the equational subsets of  $STR(\Sigma)$

$EQ = \bigcup EQ(\Sigma)$ , all  $\Sigma$

The operations are those of  $QF$

### Examples :

Context-free languages,

Trees, Graphs of tree-width  $< k$ ,

Cographs, Graphs of clique-width  $< k$

### Alternative characterization of equational sets :

$$EQ(\Sigma) = \mathit{val}(\mathit{REC}(T(QF))_{\Sigma})$$

Closure properties : union, homomorphisms,

intersection with a recognizable set

### Remark :

$L \in EQ(\Sigma) \Leftrightarrow L = \mathit{val}(K)$  for some

$K \in \mathit{REC}(T(QF(\Gamma))_{\Sigma})$ ,  $\Sigma \subseteq \Gamma$

*Question* : what is the minimal  $\Gamma$  ?

## Recognizable sets

$L \subseteq \text{STR}(\Sigma)$  is *recognizable* iff

$L$  is saturated for a finite congruence  $\cong$  over  $\text{STR}$ .

$\cong$  an equivalence relation,

$S \cong S' \Rightarrow S$  and  $S'$  belong to the same  $\text{STR}(\Sigma)$ ,

$\cong$  a congruence for the operations of  $QF$ ,

finitely many classes of each sort  $\Sigma$ .

**Notation :**  $\text{REC}(\Sigma)$  = the recognizable subsets of  $\text{STR}(\Sigma)$

$$\text{REC} = \bigcup \text{REC}(\Sigma), \text{ all } \Sigma$$

**Examples :** Regular languages and “tree languages”  
(i.e., sets of finite terms),

Trees, planar graphs, connected graphs,

(More to come)

**Closure properties :**

Boolean operations, inverse homomorphisms.

**Remark :**

Can  $L \in \text{REC}(\Sigma)$  be defined in terms of a congruence on  $\text{STR}(\Gamma)$  for a « small »  $\Gamma$  ?

**Question :** what is the minimal  $\Gamma$  ?

## Monadic Second-Order (MS) Logic

- = First-order logic on power-set structures
- = First-order logic extended with (quantified) variables denoting subsets of the domains.

### MS properties :

transitive closure,  
properties of paths, connectivity,  
planarity (via Kuratowski, uses connectivity),  
k-colorability.

### Notation :

$$\begin{aligned} \text{MS-def}(\Sigma) &\subseteq P(\text{STR}(\Sigma)) \\ &= \text{sets characterized as } \{ S \mid S \models \varphi \} \\ &\text{for an MS formula } \varphi \end{aligned}$$

### Main result : $\text{MS-def}(\Sigma) \subseteq \text{REC}(\Sigma)$

This result holds for CMS :

**Counting** Monadic Second-Order logic, using  
**modulo-counting set predicates.**

## MS transductions

Multivalued mappings  $\tau : \text{STR}(\Sigma) \rightarrow \text{STR}(\Gamma)$

$$S \longmapsto T = \tau(S)$$

where  $T$  is :

a) defined by MS formulas

b) inside the structure:  $S \oplus S \oplus \dots \oplus S$   
(fixed number of disjoint "marked" copies of  $S$ )

c) in terms of "parameters" i.e. subsets  
 $X_1, \dots, X_p$  of the domain of  $S$

*Proposition* : The composition of two

MS transductions is an MS transduction.

*Remark* : For each tuple of parameters  $X_1, \dots, X_p$  satisfying an MS property,  $T$  is uniquely defined.  $\tau$  is multivalued by the different choices of parameters.

**Quantifier-height** of  $\tau$  : the maximum quantifier-height  
of the formulas which specify  $\tau$





The fundamental property of MS transductions :

$$\begin{array}{ccc} S & \xrightarrow{\quad} & \tau(S) \\ & & \leftarrow \tau^\#(\psi) \\ & & \psi \end{array}$$

Every MS formula  $\psi$  has an effectively computable *backwards translation*  $\tau^\#(\psi)$ , an MS formula, such that :

$$S \models \tau^\#(\psi) \quad \text{iff} \quad \tau(S) \models \psi$$

The verification of  $\psi$  in the **object structure**  $\tau(S)$  **reduces** to the **verification of**  $\tau^\#(\psi)$  in the given structure  $S$ .

$S$  describes  $\tau(S)$  ; the MS properties of  $\tau(S)$  are described by MS properties of  $S$ .

**Main results :**

- 1) The value mapping  $T(QF(\Sigma \cup \Gamma))_\Sigma \rightarrow STR(\Sigma)$  is an MS transduction.
- 2)  $L \subseteq STR(\Sigma)$  is equational iff  $L = \tau(\text{Trees})$ , for an MS transduction  $\tau$ .
- 3) The image of an equational set under an MS transduction is equational

**Theorem :** The *inverse image* of a recognizable set under an MS transduction is recognizable.

## Relationships between algebraic and logical notions

Algebraic notions	Algebraic characterizations	Logical characterizations	Closure properties
EQ			union, $\cap$ Rec
	equation systems	MS-trans(Trees)	homo
	$Val(REC(Terms))$		MS-trans
REC			Boolean opns
	congruences	MS-def $\subset$ REC	homo <sup>-1</sup>
			MS-trans <sup>-1</sup>

Signatures for graphs and hypergraphs :

$HR$  : hypergraphs with “sources”

$VR$  : graphs with vertex labels

$VR+QF_{graphs}$  : with quantifier-free operations  
(ex. edge complement)

$QF$  : hypergraphs, i.e., relational structures

$QF+$  Fusion : hypergraphs.

## 2. Local information and annotations

Let  $S$  in  $STR(\Sigma)$ ,  $a_1, \dots, a_n$  in  $D_S$  (the *domain* of  $S$ ).

$$\begin{aligned} tp_k(S, a_1, \dots, a_n) &= \{ \varphi / S \mid \varphi(a_1, \dots, a_n), \\ &\quad \varphi \in MS(\Sigma, x_1, \dots, x_n) \text{ of quantifier-height } \leq k \} \\ &\in P(MS_k(\Sigma, x_1, \dots, x_n)) \\ &= \text{(k,n)-Types} \end{aligned}$$

$tp_k(S, a_1, \dots, a_n)$  is a **local information** relative to  $a_1, \dots, a_n$  in  $S$ .

$tp_k(S)$  is a **global information** relative to  $S$ .

**Annotation** :  $S \mid \rightarrow M_{k,m}(S) = (D_S, (T_p)_{p \in (k,n)\text{-Types}, n \leq m})$ .

$$T_p(a_1, \dots, a_n) : \Leftrightarrow tp_k(S, a_1, \dots, a_n) = p.$$

$$M_{k,m}(S) = S + \text{local information}$$

**Facts :**

- 1)  $M_{k,m}$  is an MS transduction of quantifier-height  $k$ .
- 2)  $M_{0,m}$  is a quantifier-free transduction.
- 3) The inverse of  $M_{k,m}$  is a quantifier-free transduction ; it is functional because  $M_{k,m}$  is injective.

## Homomorphic properties of $M_{k,m}$

**Proposition :**

$$1) M_{k,m}(S \oplus T) = \oplus^{\$} [tp_k(S), tp_k(T)] (M_{k,m}(S), M_{k,m}(T))$$

2) For  $g$  quantifier-free :

$$M_{k,m}(g(S)) = g^{\$} (M_{k,m}(S))$$

where  $g^{\$}$ ,  $\oplus^{\$} [p,p']$  are compositions of QF operations.

*Proof sketch :*

1) By Feferman-Vaught-Shelah :

$$S \oplus T \models \varphi(a_1, \dots, a_n, b_1, \dots, b_q)$$

iff for some  $i$  :

$$S \models \psi_i(a_1, \dots, a_n) \text{ and } T \models \theta_i(b_1, \dots, b_q)$$

where  $\psi_i$  and  $\theta_i$  are MS formulas of quantifier-height  $\leq$  that of  $\varphi$ , and that only depend on  $\varphi$ . Hence

$$M_{k,m}(S \oplus T) = f(M_{k,m}(S) \oplus M_{k,m}(T))$$

where  $f$  redefines the types in  $S$  and in  $T$  (in terms of  $tp_k(T)$  and  $tp_k(S)$ ) and creates appropriate relations for the types concerning both  $S$  and  $T$  (using  $add_{U,V,W}$  of Section 4).

2) By the fact that

$$tp_k(g(S), a_1, \dots, a_n) = g^{\textcircled{a}}(tp_k(S, a_1, \dots, a_n))$$

for some mapping  $g^{\textcircled{a}} : (k,n)\text{-Types} \rightarrow (k,n)\text{-Types}$  (by using backwards translation relative to  $g$ ).

We let  $g^{\$}$  replace  $q$  by  $p$  whenever  $g^{\textcircled{a}}(q) = p$

## Factorization :

Through  $M_{k,m}$  annotations, MS transductions of Quantifier-height  $\leq k$  reduce to quantifier-free ones :

If  $g : \text{STR}(\Sigma) \rightarrow \text{STR}(\Gamma)$  is an MS transduction of Quantifier-height  $\leq k$ , and  $m = \text{MaximumArity}(\Gamma)$  :

$$g(S) = f [tp_k(S)] (M_{k,m}(S))$$

where  $f [p] : \text{STR}(\Sigma_{k,m}) \rightarrow \text{STR}(\Gamma)$  is quantifier-free.

Easy construction, using renamings of the type relations and deletions of elements.

### 3. Inverse MS transductions preserve REC

This is already known for MS definable sets and the recognizable sets of graphs of bounded tree-width.

Proof sketch :

Every MS transduction is the composition of MS transductions of 3 types :

- Copy<sub>k</sub>
- Parameterless noncopying
- Guessing of unary relations

1) Copy<sub>k</sub> :

$$S \mapsto S \oplus S \oplus \dots \oplus S \text{ (k times)}$$

augmented with binary relations  $Y_{i,j}$  for  $1 \leq i < j \leq k$

defined as

$\{(x,y) \mid x \text{ is the } i\text{-copy, } y \text{ is the } j\text{-copy of some } u \text{ in } D_S\}$

Facts : a)  $\text{Copy}_k(S \oplus T) = \text{Copy}_k(S) \oplus \text{Copy}_k(T)$

b) For  $f$  quantifier-free, there is a quantifier-free  $g$  :

$$\text{Copy}_k(f(S)) = g(\text{Copy}_k(S))$$

Hence Copy<sub>k</sub> is “almost” a homomorphism, and REC is preserved under inverse homomorphisms.

## 2) Forgetting unary relations :

We let  $\text{fgt}_{\Pi} : \text{STR}(\Sigma \cup \Pi) \rightarrow \text{STR}(\Sigma)$  just « forget »  $\Pi$ ,  
 $\Pi$  is a finite set of unary relations.

Its inverse consists in « guessing » the relations in  $\Pi$ , this can be done by means of **parameters** of an MS transduction.

Lemma : If  $L \in \text{REC}$ , then  $\text{fgt}_{\Pi}(L) \in \text{REC}$ .

*Proof* : From a congruence  $\equiv$  for  $L$ , we define the equivalence :

$$S \cong T \text{ iff } \{ [U]_{\equiv} / \text{fgt}_{\Pi}(U) = S \} = \{ [U]_{\equiv} / \text{fgt}_{\Pi}(U) = T \}$$

It is finite, saturates  $\text{fgt}_{\Pi}(L)$ .

It is a congruence : we use the fact that

$$\begin{aligned} S \oplus T = \text{fgt}_{\Pi}(W) \text{ iff} \\ \text{there exist } S', T' \text{ such that } W = S' \oplus T', \\ S = \text{fgt}_{\Pi}(S'), \text{ and } T = \text{fgt}_{\Pi}(T'), \end{aligned}$$

and a similar observation for unary operations.

### 3) Parameterless noncopying MS transductions.

We let  $L \in \text{REC}(\Gamma)$

$\tau$  be a parameterless noncopying MS transduction :

$\text{STR}(\Sigma) \rightarrow \text{STR}(\Gamma)$  of quantifier-height  $k$

We prove that  $\tau^{-1}(L) \in \text{REC}(\Sigma)$

From a congruence  $\equiv$  for  $L$ , we define, on each set  $\text{STR}(\Delta)$ , the equivalence :

$S \cong T$  iff

$\text{tp}_k(S) = \text{tp}_k(T)$  and for every parameterless noncopying MS transduction  $\mu : \text{STR}(\Delta) \rightarrow \text{STR}(\Gamma)$  of quantifier-height at most  $k$ , we have  $\mu(S) \equiv \mu(T)$ .

( $\text{tp}_k(S)$  is the MS theory of  $S$  of quantifier height at most  $k$ .)

It is finite, saturates  $\tau^{-1}(L)$  (because  $\tau$  is one of the considered transductions  $\mu$ ).



It is a congruence. We consider  $f$  unary, quantifier-free. Let  $S \cong T$  :

$$\text{tp}_k(f(S)) = f^{\textcircled{a}}(\text{tp}_k(S)) = f^{\textcircled{a}}(\text{tp}_k(T)) = \text{tp}_k(f(T)) \text{ (cf. page 12).}$$

For every  $\mu$ ,  $\mu \circ f$  is a parameterless noncopying

MS transduction of quantifier-height  $k$  :

$$\mu(f(S)) = \mu \circ f(S) \equiv \mu \circ f(T) = \mu(f(T)). \text{ Qed}$$

Remains  $\oplus$  : Let  $S \cong T$ ,  $S' \cong T'$ , we want  $S \oplus S' \cong T \oplus T'$ .

We have  $\text{tp}_k(S \oplus S') = \text{tp}_k(T \oplus T')$  (by Feferman et al.)

For every  $\mu$  :

$$\begin{aligned} \mu(S \oplus S') &= f[\text{tp}_k(S \oplus S')](M_{k,m}(S \oplus S')) \quad (m = \text{MaxArity}(\Gamma), \text{ cf.} \\ &\quad \text{page 13}) \\ &= f[\text{tp}_k(S \oplus S')](\oplus^{\$}[\text{tp}_k(S), \text{tp}_k(S')](M_{k,m}(S), M_{k,m}(S'))) \end{aligned}$$

and similarly for  $\mu(T \oplus T')$  (cf. page 12)

Since  $M_{k,m}$  is an MS transduction of quantifier-height  $\leq k$  :

$$M_{k,m}(S) \equiv M_{k,m}(T), \quad M_{k,m}(S') \equiv M_{k,m}(T')$$

$$\begin{aligned} \text{Also : } \quad \oplus^{\$}[\text{tp}_k(T), \text{tp}_k(T')] &= \oplus^{\$}[\text{tp}_k(T), \text{tp}_k(T')] \\ f[\text{tp}_k(S \oplus S')] &= f[\text{tp}_k(T \oplus T')] \end{aligned}$$

Since  $\equiv$  is a QF-congruence, and  $f[p]$ ,  $\oplus^{\$}$  are in QF :

$$\mu(S \oplus S') \equiv \mu(T \oplus T'), \text{ hence } S \oplus S' \cong T \oplus T'$$

## 4. A « small » signature equivalent to $QF$

Vertex  $k$ -labelled graphs are in  $STR(\{\text{edge}, P_1, \dots, P_k\})$

Equational and recognizable sets are defined w.r.t. **all finite sets of relation symbols of unbounded arity.**

Can't we restrict to  $\{\text{edge}, P_1, \dots, P_k, \dots\}$ ? **Yes**

More generally, what about subsets of  $STR(\Sigma)$ ?

### The $VR$ signature for vertex labelled graphs :

Labels are  $1, \dots, k, \dots$

Operations are : disjoint union  $\oplus$

Edge addition  $\text{add}_{i,j}$  which adds to a graph the edges  $x \rightarrow y$  for  $x$  labelled  $i$  and  $y$  labelled  $j$

Label renaming  $\text{ren}_{i \rightarrow j}$  which replaces label  $i$  by  $j$ .

Constants : vertex labelled  $i$  ; the same with a loop.

Theorem (BC, JEngelfriet, PWeil) :

$$1) \text{EQ}(VR) = \text{EQ}(QF) \cap \text{Graphs} = \text{MS-trans}(\text{Trees})$$

$$2) \text{REC}(VR) = \text{REC}(QF) \cap \text{Graphs}$$

Proof sketch :

$$\text{MS-trans}(\text{Trees}) \subseteq \text{val}(\text{REC}(T(\text{VR}))) = \text{EQ}(\text{VR})$$

Let  $L = \tau(\text{Trees}) =$  a set of m-graphs.

1) Technical step :

One has :  $L = \sigma(K)$  where

-  $K \in \text{REC}(T(\{\diamond, *\}))$  ( $\diamond$  binary,  $*$  constant)

-  $\sigma$  is a noncopying, nondeleting, MS transduction without parameters. The vertices of  $\sigma(t)$  are the occurrences of  $*$  in  $t$ .

2) For  $t$  in  $K$  we let (cf. page 11)

$$\text{TH}(t) = M_{k,2}(t) \cap \text{Leaves}(t)$$

(= the leaves of  $t$  with labels and edges encoding the types of vertices and pairs of vertices for MS formulas of quantifier-height  $\leq k$ )

There exists a tree-transducer  $\theta$  that maps  $t$  in  $K$  to a VR-term  $\theta(t)$  such that :

$$\text{val}(\theta(t)) = \text{TH}(t).$$

3) Using the proposition page 13 :

$$\begin{aligned} \sigma(t) &= f(M_{k,2}(t)) \\ &= f(M_{k,2}(t) \cap \text{Leaves}(t)) = f(\text{TH}(t)) \\ &= f(\text{val}(\theta(t))) \end{aligned}$$

where  $f$  is a composition of VR-operations.

$$\text{Hence } L = \text{val}(f(\theta(K)))$$

$$\text{where } f(\theta(K)) \in \text{REC}(T(\text{VR}))$$

$$\text{Hence } L \in \text{EQ}(\text{VR})$$

For  $REC(VR) \subseteq REC(QF) \cap \text{Graphes}$  (proof from BC-PWeil).

Let  $L$  be VR-recognizable with congruence  $\approx$ . It is also w.r.t.  $QF$ . We define congruences  $\equiv$  on the sets  $STR(\Gamma)$  for all  $\Gamma$  by :

$$S \equiv T \text{ iff } tp_0(S) = tp_0(T) \text{ and } h(S) \approx h(T) \text{ for every quantifier-free transduction } h : STR(\Gamma) \rightarrow k\text{-Graphs.}$$

Congruence :

1) Let  $g$  be quantifier-free :  $STR(\Delta) \rightarrow STR(\Gamma)$ .

Let  $S \equiv T$ . For every  $h$  quantifier-free,  $hog$  is quantifier-free:

$$h(g(S)) = (hog)(S) \approx (hog)(T) = h(g(T)),$$

hence  $g(S) \equiv g(T)$ .

2) Next we consider  $\oplus$ .

Let  $S \equiv S'$ ,  $T \equiv T'$ . To be proved :  $S \oplus T \equiv S' \oplus T'$ .

*Lemma* : For every quantifier-free  $h : STR(\Gamma) \rightarrow k\text{-Graphs}$

$$h(S \oplus T) = t(g(S) \oplus g'(T))$$

for some composition  $t$  of VR operations, and some quantifier-free  $g, g' : STR(\Gamma) \rightarrow k\text{-Graphs}$

*Lemma* : For every quantifier-free  $h : \text{STR}(\Gamma) \rightarrow \text{k-Graphs}$

$$h(S \oplus T) = t(g(S) \oplus g'(T))$$

for some composition  $t$  of VR operations, and some quantifier-free  $g, g' : \text{STR}(\Gamma) \rightarrow \text{k-Graphs}$

*Proof sketch* : Cf p. 12 and 16. For  $\mu$  of quantifier-height 0, and  $m = 2$ :

$$\begin{aligned} \mu(S \oplus T) &= \\ &= f [tp_0(S \oplus T)] (\oplus^{\$} [tp_0(S), tp_0(T)](M_{0,2}(S), M_{0,2}(T))) \\ &= t(M_{0,2}(S) \oplus M_{0,2}(T)) \end{aligned}$$

where  $t$  is a composition of VR-operations. Qed

Now  $S \equiv S', T \equiv T'$  implies :

$$tp_0(S) = tp_0(S'), tp_0(T) = tp_0(T'), tp_0(S \oplus T) = tp_0(S' \oplus T')$$

$$M_{0,2}(S) \approx M_{0,2}(S'), M_{0,2}(T) \approx M_{0,2}(T')$$

because  $M_{0,2}$  is quantifier-free

$$\mu(S \oplus T) \approx \mu(S' \oplus T') \quad \text{because } \approx \text{ is a congruence for } \oplus \text{ and for } t .$$

Qed



## Improvement for small rank and large arity

**Notation** :  $S \in \text{STR}_n(\Gamma)$  iff each hyperedge has at most  $n$  distinct vertices, i.e., *rank* at most  $n$ . Clearly  $n \leq m$ .

One takes  $\Delta'$  finite sets of relation symbols of arity  $\leq n-1$ . We obtain, with  $\Delta'$  in place of  $\Delta$  :

**Theorem** : Same as above for  $L \subseteq \text{STR}_n(\Gamma)$ .

**Consequences** :

- 1) For 3-ary hypergraphs, all 3-ary hyperedges of which have loops,  $\text{rank} \leq 2$  hence unary “auxiliary” relations (vertex labels  $P_1, \dots, P_k$ ) suffice.
- 2) For general 3-ary hypergraphs, “auxiliary” binary relations are necessary.

**Example** :  $R$ , 3-ary relation

Let  $A_n$  with domain  $\{1, \dots, n\}$ ,  $R_{A_n} = \{ (i, j, k) / i < j < k \}$ .

1) The set  $L$  of  $A_n$ 's is the image of the set of terms  $\{ f^n(c) / n > 0 \}$  under an MS-transduction, hence is equational.

2) It cannot be generated by the operations of  $QF$  based on  $R$  and finitely many unary “auxiliary” relations  $P_1, \dots, P_m$ .

## 5. “Larger” signatures using local information

The **fusion** operation on vertex labelled hypergraphs :

$\text{fuse}_p(S)$  = the quotient of  $S$  obtained by fusing all vertices labelled  $p$

$$Fuse = \{ \text{fuse}_p \}$$

**Theorem** (BC, JMakowsky) :

The mapping  $val : T(QF \cup Fuse) \rightarrow STR$   
is an MS transduction.

$$EQ(QF) = EQ(QF \cup Fuse)$$

**Questions :**

What about REC ?

Can one use a quantifier-free formula  $\varphi(x)$  instead of  $p$   
(which corresponds to  $P_p(x)$ ) ?



## How to enlarge a signature into an equivalent one ?

$S \mid \rightarrow \text{Ann}(S) = S + \text{local information}$

$F$  a signature equivalent to  $QF$

$G$  a set of new operations

### Conditions :

1) For every  $f \in F \cup G$  :

$$\text{Ann}(f(S, \dots, T)) = \text{Ann}_f^{\text{Ann}}(\text{Ann}(S), \dots, \text{Ann}(T))$$

where  $\text{Ann}_f^{\text{Ann}}$  is a composition of operations in  $QF$

2)  $\text{Ann}^{-1}$  is an MS transduction

**Proposition** : Under these conditions :

$$\text{EQ}(F \cup G) = \text{EQ}(QF)$$

$$\text{REC}(F \cup G) = \text{REC}(QF)$$

**Proof** : The mapping  $val : T(F \cup G) \rightarrow \text{STR}$  is the composition of 3 MS transductions :

$t \mid \rightarrow \text{Ann}^{\text{Ann}} t$  (by Condition 1)

$val$  (from  $T(QF)$  to annotated structures),

$\text{Ann}^{-1}$  (by Condition 2).

Hence  $\text{EQ}(F \cup G) \subseteq \text{EQ}(QF) = \text{EQ}(F) \subseteq \text{EQ}(F \cup G)$ .

For  $\text{REC}(F \cup G) = \text{REC}(QF)$ , the proof uses the closure of REC under inverse MS transductions ( $\text{Ann}^{-1}$ ) and under inverse homomorphisms (cf. Condition 1).

## Application to Fusion

For  $\varphi(x)$  a quantifier-free formula,  $\text{fuse}_\varphi(S)$  is  $S$  where all elements satisfying  $\varphi$  are fused.

Example, for a 2-graph :  $\varphi(x)$  may be  $\text{edg}(x,x) \wedge \neg P_2(x)$

We let *Fuse* denote the set of all these operations

What kind of local information satisfies Condition 1) ?

Notation:

$\theta, \mu, \nu, \dots$  denote (0,1)-types (cf. page 11)  
 = conjunctions of atomic and negated atomic formulas in  $\text{tp}_0(S, a)$  for some  $S, a$

With  $S$  in  $\text{STR}(\Sigma)$ , we associate  $L(S)$  in  $\text{STR}(\Sigma_L)$  such that :

$L(S) = S +$  new relations, with typical case

$(a, b, c) \in R[\bullet\bullet\theta\mu\bullet\nu]_{L(S)}$  iff  
 $(a, b, u, v, c, w) \in R_S$  for some  $u, v, w$  that satisfy  
 respectively  $\theta, \mu, \nu$

$R[\bullet\bullet\bullet\bullet\bullet]$  is identical to  $R$ .

$\Sigma_L$  is  $\Sigma \cup$  new symbols like  $R[\bullet\bullet\theta\mu\bullet\nu]$  with at least one  $\bullet$

Global information associated with S :

$R[\theta_1 \dots \theta_m]$  abbreviates

$$\exists y_1, \dots, y_m (R(y_1, \dots, y_m) \wedge \theta_1(y_1) \wedge \dots \wedge \theta_m(y_m))$$

$LTP(S)$  = the set of (0,1)-types and of formulas  $R[\theta_1 \dots \theta_m]$  satisfied in S

$$\subseteq tp_n(S) \quad ((n,0)\text{-type, } n = \text{MaxArity}(\Sigma), \text{ cf. page 11})$$

Proposition : For the annotation  $L$  :

a)  $L^{-1}$  is a quantifier-free transduction, hence Condition 2) holds.

b) For a subsignature  $QF_{-}$  equivalent to  $QF$  :

b1) LTP is inductively computable w.r.t.  $Fuse \cup QF_{-}$

b2) Condition 1) holds for  $Fuse \cup QF_{-}$

*Proof* : a) is clear.

Consider b)

i) Disjoint union  $\oplus$  :

$$tp_0(S \oplus T, a) = tp_0(S, a) \text{ for } a \text{ in } S$$

$$LTP(S \oplus T) = LTP(S) \cup LTP(T)$$

$$L(S \oplus T) = L(S) \oplus L(T)$$

For unary operations  $f$  we want (cf. page 25) :

$$L(f(S)) = L_f [LTP(S)] (L(S))$$

where  $L_f$  is a composition of  $QF$  operations, depending on  $LTP(S)$ .

ii) Deletion of elements by  $f = \text{del}_\mu$  ( $\mu : (0,1)$ -type) :

$\text{tp}_0(f(S), a) = \text{tp}_0(S, a) \neq \mu$ , ( $a$  is deleted if  $\mu(a)$  holds in  $S$ )

$\text{LTP}(f(S)) = \text{LTP}(S) - \text{everything containing } \mu$

$\overset{L}{f}$  deletes elements by  $\text{del}_\mu$  and

“forgets” the relations  $R[\dots\mu\dots]$

iii) Case of  $\text{fuse}_\mu$  ( $\mu : (0,1)$ -type)

Let  $a$  in  $S$  yield  $b$  in  $\text{fuse}_\mu(S)$  (it is a quotient structure)

if  $\text{tp}_0(S, a) = \mu : \text{tp}_0(\text{fuse}_\mu(S), b) = \mu^* =$

$\{ R(x, x, \dots, x) / R[\mu\mu\dots\mu] \in \text{LTP}(S) \} \cup$

$\{ \neg R(x, x, \dots, x) / R[\mu\mu\dots\mu] \notin \text{LTP}(S) \} ;$

if  $\text{tp}_0(S, a) \neq \mu : \text{tp}_0(\text{fuse}_\mu(S), b) = \text{tp}_0(S, a)$

$\text{LTP}(\text{fuse}_\mu(S)) = \{ \mu^* \} \cup \{ \theta \in \text{LTP}(S) / \theta \neq \mu \}$

$\cup \{ R[\dots\mu^* \dots \theta \dots] / R[\dots\mu \dots \theta \dots] \in \text{LTP}(S) \}$

We now consider  $L(\text{fuse}_\mu(S)) ; f = \text{fuse}_\mu$

$$\begin{array}{ccc} S & \xrightarrow{\quad\quad\quad} & f(S) \\ \downarrow & & \downarrow \\ L(S) & \rightarrow T = \text{del}_\mu(L(S)) \rightarrow & L(f(S)) \end{array}$$

$L(f(S)) = \text{Link}(T \oplus s)$  where :

$s$  is a singleton satisfying  $\mu^*$  (this depends on  $\text{LTP}(S)$ )

**Link** : first **creates**  $(a, b, s, c, d)$  in  $R[\bullet\bullet\bullet\bullet\theta\bullet]_{L(f(S))}$

from  $(a, b, c, d)$  in  $R[\bullet\bullet\mu\bullet\theta\bullet]_{L(S)}$

then **renames**  $R[\bullet\bullet\mu\bullet\theta\bullet]$  into  $R[\bullet\bullet\mu^*\bullet\theta\bullet]$

This gives :  $\overset{L}{f} [\text{LTP}(S)] (U) = \text{Link}(\text{del}_\mu(U) \oplus s)$

iv ) **Only for certain** non-deleting quantifier-free operations f we have :

LTP(f(S)) computable from LTP(S), and

$$L(f(S)) = L_f [LTP(S)] (L(S))$$

These include :

Forget or rename (some relations)

Add<sub>U,V,W</sub> (cf. page 22)

Add<sub>U,V,W,h</sub> (cf. page 22)

and more generally those such that :

each relation  $R(x_1, \dots, x_n)$  is defined by a **disjunction** of conditions, each of the form  $A \wedge B \wedge C$  where

A is a set of **defining equalities**,

B is a conjunction of atomic and negated atomic formulas with a **single variable**,

C is a conjunction of atomic formulas where **no variable occurs more than once**.

*Example :*

For defining  $R(s,u,v,w,x,y,z)$

$$x=u \wedge y=v \wedge z=v \tag{A}$$

$$\wedge \text{edg}(u,u) \wedge \neg \text{edg}(w,w) \wedge \neg P(v) \tag{B}$$

$$\wedge \text{edg}(u,v) \wedge \text{edg}(w,s) \tag{C}$$

(x,y,z are defined from s,u,v,w)

*Proofs* : For such  $f$  :

$$tp_0(f(S),a) = f^{\textcircled{a}}(tp_0(S,a)) \quad (\text{cf. page 12})$$

$LTP(f(S))$  = the set of  $f^{\textcircled{a}}(\theta)$  for  $\theta$  in  $LTP(S)$ , and of formulas  $R[\theta_1 \dots \theta_m]$  built from  $LTP(S)$  ;  
in the example of p. 29, we let in  $LTP(f(S))$  formulas

$$R[f^{\textcircled{a}}(\theta_s)f^{\textcircled{a}}(\theta_u)f^{\textcircled{a}}(\theta_v)f^{\textcircled{a}}(\theta_w)f^{\textcircled{a}}(\theta_u)f^{\textcircled{a}}(\theta_v)f^{\textcircled{a}}(\theta_v)]$$

such that :

$$\theta_u \Rightarrow \text{edg}(u,u),$$

$$\theta_w \Rightarrow \neg \text{edg}(w,w),$$

$$\theta_v \Rightarrow \neg P(v) \quad (\text{cf. B})$$

$$\text{edg}[\theta_u \theta_v] \text{ and } \text{edg}[\theta_w \theta_s] \in LTP(S) \quad (\text{cf. C})$$

For defining  ${}^L f$  we only consider the example of  $f = \text{add}_{U,V,W}$  where  $U, V$  are of arity 2 and  $W$  of arity 4;  
Hence  $W(x,y,z,u)$  is defined in  $f(S)$  by the formula :

$$W(x,y,z,u) \vee ( U(x,y) \wedge V(z,u) )$$

Then  ${}^L f$  adds  $(x,z)$  in  $W[\bullet f^{\textcircled{a}}(\theta) \bullet f^{\textcircled{a}}(\mu)]$

whenever  $x \in U[\bullet \theta]$  and  $z \in V[\bullet \mu]$  in  ${}^L(S)$ .

This can be done by the operation

$$\text{add}_{U[\bullet \theta], V[\bullet \mu], W[\bullet f^{\textcircled{a}}(\theta) \bullet f^{\textcircled{a}}(\mu)]}$$

The signature  $QF_{\mu}$  consisting of  $\oplus$ ,  $\text{del}_{\mu}$ , and these particular quantifier-free operations is equivalent to  $QF$  by the theorem page 22.

From Conditions 1 and 2 we obtain :

**Theorem :**

$$1) \text{EQ}(QF \cup \text{Fuse}) = \text{EQ}(QF_{\mu} \cup \text{Fuse}) = \text{EQ}(QF)$$

$$2) \text{REC}(QF \cup \text{Fuse}) \subseteq \text{REC}(QF_{\mu} \cup \text{Fuse}) = \text{REC}(QF)$$

The proof is done for operations  $\text{fuse}_{\varphi}$  where  $\varphi(x)$  is a (0,1)-type (i.e., some  $\text{tp}_0$ ). The operations  $\text{fuse}_{\varphi}$  can be expressed as combinations of these special fuse operations and some quantifier-free operations. The theorem follows.

Example showing why the lemma page 29 does not extend to  $QF$  instead of  $QF_{\_}$

Let  $R$  of arity 3 and  $f : STR(\{R\}) \rightarrow STR(\{R\})$

redefine  $R$  by :

$R(x,y,z)$  in  $f(S)$  iff  $R(x,y,y)$  and  $R(y,z,z)$  in  $S$

We let  $S$  with domain  $\{1,2,3\}$  and

$R_S = \{(1,2,2), (2,3,3)\}$

$f(S)$  has domain  $\{1,2,3\}$  and  $R_{f(S)} = \{(1,2,3)\}$

We let  $T$  with domain  $\{1,2,3,4\}$  and

$R_T = \{(1,2,2), (4,3,3)\}$

$f(T)$  has domain  $\{1,2,3,4\}$  and  $R_{f(T)} = \emptyset$

$LTP(S) = LTP(T)$

Now if  $L(f(S)) = L_f [LTP(S)] (L(S))$  (1)

we also have

$L(f(T)) = L_f [LTP(S)] (L(T))$  (2)

Observe that :

$1 \in R[\bullet\theta\theta]$  in  $L(S)$ , in  $L(T)$ , in  $L(f(S))$ ,

but  $1 \notin R[\bullet\theta\theta]$  in  $L(f(T))$ , where  $\theta(x)$  is  $\neg R(x,x,x)$  ;

we should have  $1 \in R[\bullet\theta\theta]$  in  $L(f(T))$  from (1) and (2).



## 6. The fusion operation for graphs ; “gluing” operations based on local information

All relations have arity 1 or 2. Structures are vertex- and edge-labelled graphs.

New binary “gluing” operations forming a set *Glue*

$$G \otimes_g H = G \oplus H + \text{edges between } x \text{ and } y \\ \text{whenever } G \models \varphi(x), H \models \theta(y)$$

Formulas  $\varphi, \theta$  are first-order and represent “local information” relative to  $x$  and  $y$

Question : When do we have, for *Glue* :

$$\text{EQ}(VR) = \text{EQ}(QF_{\text{graphs}}) = \text{EQ}(QF_{\text{graphs}} \cup \text{Glue}) \\ \text{REC}(VR) = \text{REC}(QF_{\text{graphs}}) = \text{REC}(QF_{\text{graphs}} \cup \text{Glue}) ?$$

Complete annotation :

$\gamma, \gamma', \dots$  denote (0,2)-types (cf. page 11)  
= conjunctions of all the atomic and negated atomic formulas in  $\text{tp}_0(G, a, b)$  for some  $G, a, b \neq a$

For such  $\gamma = \text{tp}_0(G, a, b)$ ,  $\gamma_1$  denotes  $\text{tp}_0(G, a)$ ,  
 $\gamma_2$  denotes  $\text{tp}_0(G, b)$

$\Lambda$  = the set of formulas  $\lambda(x)$  of the form  $\exists y.\gamma(x,y)$

The complete 1-type of  $a$  in  $G$  is :

$ctp(G,a)$  = the conjunction of the valid formulas  $\lambda(x)$   
and of the negations of the others  
=  $\delta(x)$

$CTP(G)$  = the set of formulas  $\exists x.\delta(x)$  valid in  $G$

Complete annotation :

For every graph  $G$  :

$C(G)$  =  $G$  + each vertex  $a$  labelled by  $\delta$  with  $G \models \delta(a)$   
+ edges  $a \rightarrow b$  labelled by  $tp_0(G, a,b)$ , for all  $a \neq b$

( $C(G)$  is a complete labelled graph).

The new gluing operations

$G \otimes_g H = G \oplus H + R\text{-edges } x \rightarrow y$   
whenever  $G \models \delta(x)$ ,  $H \models \delta'(y)$

Formally  $g$  is a disjunction of triples  $(R(x,y), \delta(x), \delta'(y))$

## Verification of Condition 1 (p. 25)

i) Gluing  $\otimes_g$  (and disjoint union since  $\otimes_\emptyset = \oplus$ ) :

Facts : 1)  $tp_0(G \otimes_g H, a) = tp_0(G, a)$ ,  $tp_0(G \otimes_g H, a,b) = tp_0(G, a,b)$   
if  $a,b \in G$

2) an R-edge  $a \rightarrow b$  is defined from  $ctp(G,a)$  and  $ctp(H, b)$

3)  $ctp(G \otimes_g H, a)$  is defined from  $ctp(G,a)$  and  $CTP(H)$ .

Hence :

$$CTP(G \otimes_g H) = h(CTP(G), CTP(H))$$

$$C(G \otimes_g H) = k[CTP(G),CTP(H)](C(G), C(H))$$

where :

$h$  is a fixed function and

$k[p,p']$  is a composition of VR-operations :

disjoint union of  $C(G)$  and  $C(H)$ ,

addition of edges between them,

relabellings for updating local information.

ii) For unary operations  $f$  we want (cf. page 25) :

$$C(f(G)) = {}^C f [CTP(G)] (C(G))$$

where  ${}^C f$  is a composition of  $QF$  operations, depending on  $CTP(G)$ .

a) Deletion of elements by  $f = \text{del}_\mu$  ( $\mu$  is a (0,1)-type) :

Facts :

$\text{tp}_0(f(G), a) = \text{tp}_0(G, a) \neq \mu$ , ( $a$  is deleted if  $\mu(a)$  holds in  $G$ )

$\text{tp}_0(f(G), a, b) = \text{tp}_0(G, a, b)$  if  $\text{tp}_0(G, a) \neq \mu$ ,  $\text{tp}_0(G, b) \neq \mu$

$\text{ctp}(f(S), a)$  contains no formula  $\exists y. \gamma(x, y)$  such that  $\gamma_1$  or  $\gamma_2$  is equal to  $\mu$

Hence :

$CTP(f(G)) = CTP(G) - \text{everything "containing" } \mu$

${}^C f$  deletes the elements labelled by  $\delta$  containing  $\exists y. \gamma(x, y)$  such that  $\gamma_1 = \mu$ , and updates the labels  $\delta$  by removing formulas  $\exists y. \gamma(x, y)$  such that  $\gamma_2 = \mu$

b)  $f$  is nondeleting and quantifier-free

${}^C f$  updates the labels  $\delta$  by using the updating function  $f^\circledast$  on 0-types, cf. p. 12.

Fusion operation  $\text{fuse}_\mu$  is a composition of  $QF$  and *Glue* operations :

$$\text{fuse}_\mu(G) = \text{del}_\mu(G \otimes_g s)$$

where  $s$  is a singleton graph, possibly with loops and vertex labels, cf. page 28, and  $g$  define edges between  $G$  and  $s$

Hence we have the equivalence of the signatures :

$$VR, QF_{\text{graphs}}, QF_{\text{graphs}} \cup \textit{Fuse}, QF_{\text{graphs}} \cup \textit{Glue}$$

## 7. « Sources » are dispensable

Constant symbols  $a, b, c, \dots$  in relational structures denote unique elements called « sources » in the graph grammar terminology.

The typical use is for parallel composition of (hyper)graphs :

$$G // H =$$

$G \oplus H$  with fusion of sources with same names.

Quantifier-free operations may use constants in redefinitions like :

$$\text{edg}(x,y) : \Leftrightarrow \text{edg}(x,a) \wedge \text{edg}(y,b)$$

but are no longer quantifier-free if written :

$$\text{edg}(x,y) : \Leftrightarrow \exists u,v (\text{edg}(x,u) \wedge P_a(u) \wedge \text{edg}(y,v) \wedge P_b(v))$$

by the obvious coding of constants by unary predicates

We get a signature  $QF-C$

A coding of  $S$  in  $STR(\Sigma, C)$  into  $\underline{S}$  in  $STR(\underline{\Sigma})$

We delete  $C_S$ , the elements denoted by constants in  $C$

We replace a tuple  $(x, a_S, y, b_S, b_S)$  in a relation  $R_S$  where  $x, y$  are not in  $C_S$  by the pair  $(x, y)$  in a new relation  $R[\clubsuit a \clubsuit b b]$ .

Hence  $\underline{\Sigma}$  is the set of relations  $R[w]$  where :

$w$  is a word in  $(C \cup \{\clubsuit\})^*$  of length the arity of  $R$

containing at least one  $\clubsuit$  and

the arity of  $R[w]$  is the number of occurrences of  $\clubsuit$ .

**Fact :**  $S$  can be reconstructed in a unique way from  $\underline{S}$  and  $S[C_S]$ , by an operation in  $QF-C$  (one for each  $S[C_S]$ ).

We let  $L/\vartheta = \{ S \in L \mid S[C_S] = \vartheta \}$ .

**Theorem :** Let  $L \subseteq STR(\Sigma, C)$ :

1)  $L \in EQ(QF-C)$  iff  $\underline{L/\vartheta} \in EQ(\underline{\Sigma})$  for each  $\vartheta$ .

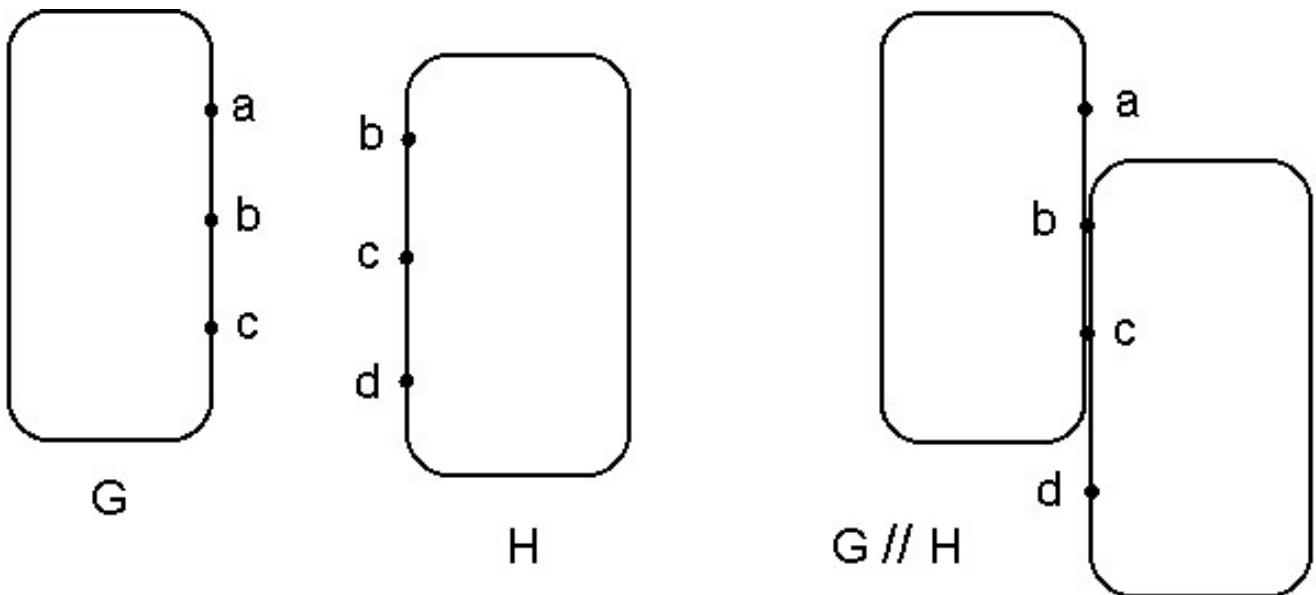
2)  $L \in REC(QF-C)$  iff  $\underline{L/\vartheta} \in REC(\underline{\Sigma})$  for each  $\vartheta$ .

## Reminder : Parallel Composition //

(Hyper)graphs have distinguished vertices called *sources*, designated by labels from a set of size  $k$  :  $\{a, b, c, \dots, h\}$ .

$G // H$  is the disjoint union of  $G$  and  $H$  and sources with same label are **fused**.

(If  $G$  and  $H$  are not disjoint, one first makes a copy of  $H$  disjoint from  $G$ .)





## 8. Open questions

1. Are the signatures  $QF$  and  $QF \cup Fuse$  equivalent, (without restriction to binary structures) ?
2. Which quantifier-free operations on relational structures preserve recognizability ?
3. Can one define on relational structures a complexity measure generalizing clique-width for graphs and “nicely” related with a signature equivalent to  $QF$  ?

*This document and the base article are available from*

*<http://www.labri.fr/~courcell/ActSci.html>*